A re-examination of the dry gets drier and wet gets wetter paradigm over global land: insight from terrestrial water storage changes

Jinghua Xiong¹, Shenglian Guo¹, Abhishek², Jie Chen¹, and Jiabo Yin¹

¹ State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, 430072, China
² School of Environment and Society, Tokyo Institute of Technology, Yokohama 226-8503, Japan

Correspondence to: S. Guo (slguo@whu.edu.cn)

Abstract. The “dry gets drier and wet gets wetter” (DDWW) paradigm has been widely used to summarise the expected trends of the global hydrologic cycle under climate change. However, the paradigm is challenged over land due to the choice of different metrics and datasets used and is still unexplored from the perspective of terrestrial water storage anomaly (TWSA). Considering the essential role of TWSA in wetting and drying of the land system, here we built upon a large ensemble of TWSA datasets, including satellite-based products, global hydrological models, land surface models, and global climate models to evaluate the DDWW hypothesis during the historical (1985-2014) and future (2071-2100) periods under various scenarios with a 0.05 significance level. We find that 28.1% of global land confirms the DDWW paradigm, while 23.3% of the area shows the opposite pattern during the historical period. In the future, the DDWW paradigm is still challenged with the percentage supporting the pattern lower than 20%, and both the DDWW-validated and DDWW-opposed proportion increase along with the intensification of emission scenarios. The different choices of data sources and varying significance levels (0.01-0.1) have subtle influences on the evaluation results of the DDWW paradigm. Our findings will provide insights and implications for global wetting and drying trends from the perspective of TWSA under climate change.

1 Introduction

The hydrological conditions of the land surface have experienced considerable changes due to climate change and anthropogenic interventions, exerting a tremendous impact on regional agriculture, ecological environment, and freshwater availability globally (Shugar et al., 2020; Gampe et al., 2021). Assessing the variations of terrestrial wetness and dryness is crucial in understanding the hydrological response and dealing with water-related issues in the context of global change (Moreno-Jimenez et al., 2019; Zhao et al., 2021). Under these circumstances, the ‘dry gets drier and wet gets wetter’ (DDWW) paradigm is one of the most widely used hypotheses to summarise the long-term trends in the global hydrological cycle (Yang et al., 2019). Initially, it was developed based on the deficit between precipitation and evapotranspiration \( P - E \), which is expected to increase due to the enhancement of atmospheric water vapour in humid regions (i.e., convergence zones) under a warming climate, and decrease over arid regions (i.e., divergence zones) (Durack et al., 2012). The DDWW paradigm has been used to represent the historical and future trends in various constituent components of the hydrologic
cycle on regional (Chou et al., 2009; Allan et al., 2010; Hu et al., 2019; Zeng et al., 2019) and global scales (Held and Soden, 2006; Donat et al., 2016). However, the rationale and validity of the DDWW mechanism are recently questioned at different levels through the growing accessibility of datasets, models, and indicators (Polson and Hegerl, 2017; Yang et al., 2019; Y. Li et al., 2021). Byrne and Gorman (2015) used simulations from 10 climate models to reveal an ocean-land contrast pattern in the response of $P - E$ to global warming in historical (1976-2005) and future (2071-2099) periods, highlighting the DDWW mechanism as more suitable over ocean than over land. Given the fact that historical evaluation of the DDWW paradigm is mainly based on oceanic records, Greve et al. (2014) adopted 2142 possible combinations of $P - E$ to assess the trends in wetting and drying over global land and discovered merely 10.8% of the area following the DDWW pattern during the period 1948-2005. Alternatively, Yang et al. (2019) integrated an ensemble of six hydro-climatic indicators for the global assessment of the DDWW paradigm between 1982 and 2012, suggesting the catchphrase only occurred over 20% of the global land. In short, there are great uncertainties still remaining in the assessments of global trends in dryness and wetness under climate change (Dai, 2011; Trenberth et al., 2014).

The uncertainties within previous studies are mainly sourced from different choices of metrics adopted and datasets used for evaluating the changes in dryness and wetness (Vicente-Serrano et al., 2010; Feng and Zhang, 2015; Huang et al., 2016). Specifically, the widely used metric $P - E$ over the ocean has been proven overwhelmingly positive over land based on both observations and simulations (Greve et al., 2014; Byrne and O’Gorman, 2015; Greve and Seneviratne, 2015). Moreover, some meteorological indices derived from precipitation and evapotranspiration, such as the standardised precipitation evapotranspiration index (SPEI), aridity index (AI), and standardised precipitation/evapotranspiration index (SPI/SETI), do not capture the integrated response of the land system due to the trade-off between the simplicity of meteorological factors and computational requirements of process-based variables (Huntington, 2006; Dai, 2011; Slette et al., 2020; Barnard et al., 2021). A few indexes like the standardised soil moisture index (SSI), standardised groundwater index (SGI), and standardised runoff index (SRI), however, focus on a single aspect of the water cycle and do not describe the integrated status of the terrestrial water storage (TWS) (AghaKouchak, 2014; Wu et al., 2018; Guo et al., 2021). In the coupled human-natural systems, where the synergistic impacts of natural and anthropogenic drivers are exceedingly difficult to disentangle, an integrated representation of the land systems is of paramount importance for policymakers (Abhishek et al., 2021; Rodell et al., 2018). TWS, consisting of water storage in surface water, soil moisture, groundwater, snow and ice, and canopies, can physically provide integrated information about the overall status of the land, whose changes are closely linked to the terrestrial wetting and drying tendency (Tapley et al., 2019; Pokhrel et al., 2021). Apart from the societal and economic importance, TWS plays a vital role in Earth system processes, including climate, weather, and biogeochemical cycles (Abhishek et al., 2021, Seyoum and Milewski, 2017). Therefore, understanding the past and future TWS trends dynamics is not only essential for human life but also crucial for assessing the water cycle, planning, policymaking, and other management strategies for water resources in a changing climate and for a continuously increasing population (Abhishek et al., 2021). There are several studies dealing with TWS or derived indicators to assess freshwater availability (Rodell et al., 2018), water storage dynamics (Abhishek et al., 2021, Scanlon et al., 2018), droughts and floods monitoring
(Long et al., 2014), among others. Divergent patterns of TWS changes have been reported over arid and humid regions under the combined effects of climate change (e.g., global warming), climatic variability (e.g., ENSO), and human activity (e.g., groundwater pumping) (Chang et al., 2020; An et al., 2021; Hu et al., 2021). However, there is no study to examine the global variability and validity of DDWW paradigm in the past and future in terms of TWS changes. Furthermore, divergent data sets produce different trends in TWS due to distinctive internal variability and external forcing (from satellites and meteorological stations), especially from precipitation and evapotranspiration (Chen et al., 2020). For example, Scanlon et al. (2018) conducted comprehensive comparisons between decadal trends in TWS from seven global models and three Gravity Recovery and Climate Experiment (GRACE) satellite solutions over major basins globally and showed a large underestimation of the increasing and decreasing trends of models primarily due to human water use and forcing climate variations.

Therefore, we conduct a systematic re-examination of the DDWW paradigm from the perspective of terrestrial water storage anomalies (TWSA) using a large ensemble of nine different TWS datasets from the GRACE reconstructions, global hydrological models, and land surface models to examine the DDWW paradigm over the global land between 1985 and 2014. Subsequently, an alternative ensemble of eight global climate models (GCMs) from the Coupled Model Intercomparison Project 6 (CMIP6) is used to further test the paradigm under various scenarios during the future (2071-2100) period.

2 Data and Methods

2.1 Data

We used an ensemble of nine data sets (hereinafter “DATASET”) to evaluate the DDWW paradigm during the historical period 1985-2014, which includes three of GRACE reconstructions, global hydrological models (GHMs), and global land surface models (LSMs) each (see Table 1). Since no dataset presents the absolutely ‘true’ value, the ensemble mean was estimated using a simple average method to avoid the uncertainty implicit in any individual dataset. All of the members and their mean have been resampled to 1° × 1° grid cell to compare the average value of three GRACE mass concentrations (mascon) solutions between June 2002 and December 2014. The missing months of GRACE data have been filled using a linear interpolation method. Alternatively, an ensemble of eight simulations from CMIP6 was used to examine the DDWW paradigm in the future period (2071-2100). All the ensemble members have been resampled to 1° × 1° scale using a bilinear interpolation approach for consistency and better comparison in the spatial domain. Similarly, the ensemble mean of CMIP6 models has been estimated using simple averaging. All the DATASET and CMIP6 members and their ensemble are represented as the long-term anomaly relative to the baseline between 1985 and 2014 to be consistent with reconstructed GRACE data.
### Table 1: Datasets used in this study

<table>
<thead>
<tr>
<th>Type</th>
<th>Dataset name</th>
<th>DOI/URL</th>
<th>Selected period</th>
<th>Temporal resolution</th>
<th>Spatial resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GRACE Reconstructions</strong></td>
<td>Li et al., 2021</td>
<td>Doi: 10.1029/2021GL093492</td>
<td>1985-2014</td>
<td>Monthly</td>
<td>0.5°</td>
</tr>
<tr>
<td><strong>GRACE</strong></td>
<td>CSR RL06 mascons-v02</td>
<td><a href="http://www2.csr.utexas.edu/grace/">http://www2.csr.utexas.edu/grace/</a></td>
<td>2002-2014</td>
<td>Monthly</td>
<td>0.25°</td>
</tr>
<tr>
<td></td>
<td>JPL RL06 mascons-v02</td>
<td><a href="http://podaac.jpl.nasa.gov/grace">http://podaac.jpl.nasa.gov/grace</a></td>
<td>2002-2014</td>
<td>Monthly</td>
<td>0.5°</td>
</tr>
<tr>
<td></td>
<td>GSFC RL06 mascons-v02</td>
<td><a href="https://earth.gsfc.nasa.gov/geo/data/">https://earth.gsfc.nasa.gov/geo/data/</a></td>
<td>2002-2014</td>
<td>Monthly</td>
<td>0.5°</td>
</tr>
<tr>
<td><strong>Global Hydrological Models</strong></td>
<td>WGHM-v2.2d</td>
<td><a href="https://gmd.copernicus.org/">https://gmd.copernicus.org/</a></td>
<td>1985-2014</td>
<td>Monthly</td>
<td>0.25°</td>
</tr>
<tr>
<td></td>
<td>PCR-GLOBWB</td>
<td><a href="https://globalhydrology.nl/research/">https://globalhydrology.nl/research/</a></td>
<td>1985-2014</td>
<td>Monthly</td>
<td>0.5°</td>
</tr>
<tr>
<td></td>
<td>CPC-v2</td>
<td><a href="https://www.psl.noaa.gov/data/">https://www.psl.noaa.gov/data/</a></td>
<td>1985-2014</td>
<td>Monthly</td>
<td>0.5°</td>
</tr>
<tr>
<td></td>
<td>MPI-ESM1-2-HR</td>
<td><a href="https://esgf-node.llnl.gov/projects/">https://esgf-node.llnl.gov/projects/</a></td>
<td>1985-2100</td>
<td>Monthly</td>
<td>0.9375°×0.9375°</td>
</tr>
</tbody>
</table>
2.1.1 GRACE and GRACE Reconstructions

A total of three GRACE reconstructions provided by Humphrey and Gudmundsson (2019) and Li et al. (2021) were selected in the DATASET for evaluation of the DDWW paradigm. The ensemble of GRACE reconstructions is generated based on state-of-the-art machine learning models using historical and near-real-time meteorological forcing. It is informative to note that the GRACE reconstructions from Humphrey and Gudmundsson (2019) were calibrated with GRACE mascon solutions from NASA JPL and NASA Goddard Space Flight Center (GSFC), respectively, and that supplied by Li et al. (2021) were trained by the GRACE mascon product from Center for Space Research (CSR). The GRACE reconstructions of JPL and GSFC mascon data (Humphrey and Gudmundsson, 2019) are forced with three different climatic datasets, including the multisource weighted-ensemble precipitation (MSWEP), the Global Soil Wetness Project Phase 3 (GSWP3), and the European Centre for Medium-Range Weather Forecasts reanalysis (ERA5). We have implemented an ensemble of each reconstructed solution to reduce the uncertainties caused by different forcing datasets. The CSR reconstruction is derived from four kinds of meteorological variables (e.g., precipitation and 2 m air temperature) and three kinds of hydrological variables (e.g., soil moisture and runoff) (Li et al., 2021a). The accuracy and applicability of three GRACE reconstructions have been fully evaluated over global land in several previous studies (Xu et al., 2021; Yi et al., 2021). Correspondingly, the three latest GRACE mascon solutions (RL06-v02) from JPL, GSFC, and CSR were prepared for comparison purposes. Compared to other GRACE products (e.g., conventional spherical harmonic solutions), mascon solutions do not need spatial (e.g., smoothing) or spectral (e.g., de-striping) filtering or other empirical scaling and therefore have higher signal-to-noise ratio, higher spatial resolutions, and eventually reduced errors (Save et al., 2016; Watkins et al., 2015).

2.1.2 Global Hydrological Models

We used three global hydrological models, including the Variable Infiltration Capacity macroscale model (VIC-v4.1.2), the WaterGAP hydrological model (WGHM-v2.2d), and PCRaster GLOBal Water Balance model (PCR-GLOBWB-v2.0) to estimate TWS for evaluation of the DDWW paradigm. The physically-based, semi-distributed, and grid-based VIC model is managed by the NASA Global Land Data Assimilation System Version 2.1 (GLDAS-v2.1) (Liang et al., 1994; Syed et al., 2008). Forced by the Global Data Assimilation System atmospheric analysis fields (Derber et al., 1991) and the Air Force Weather Agency’s AGRicultural METeorological modeling system radiation fields, the VIC model can effectively capture the terrestrial water cycle by simulating the water stored in the canopies, snow, and soil moisture within the depth of 3 soil layers (200 cm). The VIC model has been widely used to analyze terrestrial water storage changes at regional and global scales (Hao and Singh, 2015; Hao et al., 2018). The WGHM is a grid-based global hydrological model quantifying the human water use and continental water fluxes for all land areas excluding Antarctica (Müller Schmied et al., 2021). Unlike most global hydrological models, the WGHM forced by the ERA40 and ERA-Interim reanalysis is able to simulate groundwater storage by coupling with global water use models like the Groundwater-Surface Water Use, suggesting a
comparably better representation of TWS (Döll et al., 2014). Several frequently-used model outputs such as TWS, stream, and water use have been evaluated against global observations (Wan et al., 2021). The PCR-GLOBWB model is a grid-based global-scale hydrology and water resources model that fully integrates water use, such as water consumption, water withdrawal, and return flows (Sutanudjaja et al., 2018). Forced with the EC-Earth data, including atmospheric, oceanic, and land surface variables, the PCR-GLOBWB model can simulate the entire terrestrial system over global land at a daily time scale. The model performance has been fully evaluated using global discharge measurements and supported by many TWS studies globally (Scanlon et al., 2018; van der Wiel et al., 2019).

2.1.3 Land Surface Models

We used three land surface models consisting of the Noah (v3.6), Catchment (CLSM-vF2.5), and CPC (v2) models to calculate TWS globally for assessment of the DDWW paradigm. The Noah and CLSM models are managed by GLDAS (v-2.1) from the NASA GSFC institute. GLDAS is a composite of global hydrological and land surface models that model optimal fields of land surface by integrating multi-source observations such as in situ stations and satellites based on state-of-the-art data assimilation and land surface simulation techniques (Rodell et al., 2004). GLDAS has been widely used to compare with GRACE TWSA in data-sparse regions such as Africa and Qinghai-Tibetan Plateau (Ogou et al., 2021; Xing et al., 2021). The Noah-modelled TWS is considered as the sum of canopy water storage, snow water equivalent, and soil moisture of four layers with a total depth of 200 cm. Different from that, the CLSM simulates shallow groundwater and the vertical levels of soil moisture are not explicitly divided within the depth of 100 cm. Developed by the U.S. National Oceanic and Atmospheric Administration (NOAA), the CPC model provides global soil moisture conditions in 160 cm column soil forced with observations of different meteorological and hydrological fluxes (e.g., precipitation, temperature, and humidity) from CPC (Fan, 2004). The reasonably good ability of CPC simulations to capture the TWS dynamics has been examined over many areas of the globe in spite of its simplicity in the calculation of TWS (Jin et al., 2012; Agutu et al., 2020).

2.1.4 Global Climate Models

We used a suite of eight global climate models belonging to the ensemble “r1i1p1f1” of CMIP6 to evaluate the DDWW paradigm under climate change. The CMIP6 serves as a category of experiments of GCMs coupled to a dynamic ocean, a simple land surface, and thermodynamic sea ice (Eyring et al., 2016). We chose these eight models out of the 34 CMIP6 models because, as we write, they are the only models for which TWSA results are available in both the historical and future periods under multiple emission scenarios (see Table 1). The CMIP6 TWSA represents the sum of total soil moisture and snow equivalent water, which has been comprehensively validated, with embedded uncertainties, over global major river basins compared with the GRACE data (Freedman et al., 2014; Wu et al., 2021). The CMIP6 comparisons have become a diagnostic tool to better understand climate change in past, present, and future periods (Krishnan and Bhaskaran, 2020), which includes a total of five Shared Socio-economic Pathways (SSPs) representing global economic and demographic
changes under different greenhouse gas emissions. We selected three out of five SSP scenarios, including SSP126, SSP245, and SSP585, representing the green road, middle of the road, and the highway road, respectively (Iqbal et al., 2021). In specific, the monthly average TWS from CMIP6 is estimated as the sum of total soil moisture and snow water, which has been proven reliable to assess the TWS changes (Wu et al., 2021). To avoid the considerable uncertainties in TWS of different CMIP6 models, a trend-preserving method was employed to perform bias correction combined with historical GRACE data. The trend-preserving method initially developed by Hempel et al. (2013) modifies the monthly means of the simulated data to match the observed data using a constant offset between the simulations and observations and has been widely used in the Intersectoral Model Intercomparison Project (ISIMIP2b). The detailed procedure of the bias correction for CMIP6 TWSA has been described in detail in a recent study (Xiong et al., 2022).

2.2 Detection of Wetting and Drying

The non-dimensional TWS drought severity index (TWS-DSI) was adopted to reflect the long-term trends in terrestrial dryness and wetness at both $1^\circ \times 1^\circ$ grid cell and regional scales over global land (see Figure S1 and Table S1). TWS-DSI has been widely used in hydrology and climate fields due to its simple structure and effective ability in capturing drying and wetting conditions (Pokhrel et al., 2021). Monthly TWS-DSI was calculated for all ensemble members and their mean from DATASET and CMIP6 as follows (Zhao et al., 2017):

$$TWS-DSI_{ij} = \frac{TWS_{ij} - \mu_j}{\sigma_j}$$  \hspace{1cm} (1)

where $TWS_{ij}$ is the TWS value in year $i$ and month $j$; $\mu_j$ and $\sigma_j$ denote the mean and standard deviation of the annual TWS in month $j$, respectively. Long-term trends in TWS-DSI were estimated using the linear regression method and the significant trend values were evaluated using the t-test at a 5% significance level (Greve et al., 2014). The area having a significant trend of increasing/decreasing TWS-DSI is considered to be undergoing wetting/drying; otherwise, it is defined as an uncertain region. We calculated the regional/global mean trends of TWS-DSI using a spatially weighted method to account for the changing area of grid cells with latitudes.

2.3 Selected Regions

We performed the assessment of the DDWW paradigm over global land at both gridded $1^\circ \times 1^\circ$ cell and regional scales excluding Greenland and Antarctica. A total of 43 regions are selected based on the Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Adaptation (SREX) from the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6), which covers all the land area except for the Greenland and Antarctica (see Figure S1). Their basic information is summarised in Table S1.
3 Results and Discussion

3.1 Global Trends of Dryness and Wetness

Prior to the detection of the DDWW paradigm, we performed the evaluation of TWSA derived from the ensemble mean of the DATASET and CMIP6 archive. Figure S2 presents the global distribution of the normalized root mean square error (NRMSE) between GRACE TWSA and that from DATASET and CMIP6 data after bias correction during the period April 2002-December 2014, which is calculated as the ratio of RMSE to the change range of TWSA. The NRMSE between GRACE and DATASET is generally lower than 0.3 (for 95.7% of the area), of which 12.9% of the land area shows NRMSE below 0.1 and the percentage is 77.0% for the NRMSE lower than 0.2. Relatively larger NRMSE ranging from 0.3 to 0.4 occurs in east and central Asia, south Australia, north Africa, and north-eastern America, indicating the relatively poorer performance of GHMs and LSMs. Results over central America, east Europe and west Asia are mostly below 0.1. However, the NRMSE between GRACE and CMIP6 is generally lower than that of DATASET, most (98.7%) of which change from 0.1 to 0.3. Approximately 76.9% of the land area shows NRMSE<0.2 and 8.6% has NRMSE<0.1. Spatially, there are some extreme values as high as 0.3 to 0.4 located in north and east Asia, south Africa, and central America, probably arising from the uncertainties in the CMIP6 simulations even undergoing the bias correction. Figure S3 further presents the regional results over the 43 selected SREX regions. Most of the mid-latitude regions like the WCE, EEU, WSB, ESB, and RFE present relatively lower NRMSE (0-0.1) between GRACE and DATASET, suggesting better performance than that in the NZ, ECA, NEU, and NEN. For CMIP6 data, a similar pattern can be found with most mid-latitude regions having relatively lower NRMSE values from 0.1 and 0.2, while higher values (0.2-0.3) are located in NEN, ECA, SSA, and EAU. Although the bias correction has been performed to the CMIP6 TWSA, a comparatively large bias still exists owing to the uncertainty in parameters, hydrometeorological forcing, and internal variability of GCMs. Such biases can potentially influence the assessment of the DDWW paradigm in the future period (2071-2100) climate change.

A temporal comparison of global average TWSA derived from DATASET/CMIP6 and GRACE during the period 2002-2014 is shown in Figure S4. The GRACE TWSA changing from −20 to 20 mm shows obvious seasonal characteristics with relatively higher uncertainty in the dry season than that in the wet season. A similar change pattern is captured by the DATASET, with the change range covering the variations of GRACE data. The NRMSE between the ensemble mean of DATASET and GRACE data is 0.11, equaling to that (0.11) between the ensemble mean of CMIP6 and GRACE results. Moreover, the fluctuation range of DATASET data is generally greater than the CMIP6 before 2010 and slightly underestimate that and GRACE measurements after the year with increasing range, highlighting the considerable uncertainty sourced from different forcing variables and model parameterizations.

To provide insights into the aspect of terrestrial water storage changes for the evaluation of the DDWW paradigm, TWS-DSI is estimated to determine the terrestrial wetness and dryness. Figure 1 shows the global distribution of long-term trends in TWS-DSI over the historical period 1985-2014 and the future period 2071-2100 under SPSP126, SSP245, and SSP585 scenarios. During the historical period, a clear spatial divergence is observed globally and the average TWS-DSI has
a significant decreasing slope of $-0.07/a$ (p<0.05), similar to the results from SPI, SPEI, and AI (Wang et al., 2018; Yang et al., 2019). Spatially, severe drying exists in the Gulf of Alaska coast and the Canadian archipelago with significant slopes of TWS-DSI ranging from $-0.08/a$ to $-0.12/a$, which is caused by rapid ice-sheet and glacier ablation under a warming climate (Luthcke et al., 2013; Velicogna et al., 2014). Triggered by severe historical droughts over decades, the drying trends in central Canada, southern California, and Texas can be clearly discovered, with decreasing trend of TWS-DSI ranging from $-0.04/a$ to $-0.12/a$ (p<0.05) (Bouchard et al., 2013; Haacker et al., 2016), so as the eastern Brazil (Getirana, 2016). Moreover, overwhelming groundwater depletion due to unsustainable human water use such as irrigation is responsible for the declining dryness at significant slopes ranging from $-0.12/a$ to $-0.16/a$ in southeast and north Africa, south Europe, North China Plain, and northern India (Rodell et al., 2009; Feng et al., 2013; Remillien et al., 2014). Naturally, a moderate drying trend in southwestern Africa caused by precipitation decrease is detected by the reduction of TWS-DSI. On the contrary, increasing precipitation dominates the wetting trend in mid-latitude regions, including southern Russia and Canadian, west Africa, southeast Asia and Qinghai-Tibetan Plateau, with significant slopes ranging from 0.04/a to 0.12/a (Siebert et al., 2010; Ndehedehe et al., 2017). Alternatively, some regions, such as the Amazon River basin, south Africa and eastern Australia, presenting wetting trends are considered to experience progression from wet to dry period (Chen et al., 2010; Gaughan and Waylen, 2012).

In the future, most of the mid-latitude regions such as north China, south Mongolia, and central Europe are projected to be wet because of the growth in precipitation under SSP126 scenario (Milly et al., 2005; Seneviratne et al., 2006). Similar trends can be found in North China Plain and Caspian regions that underwent drying during the historical period, mainly due to groundwater abstraction and sporadic droughts. Some areas, including northern India and southwestern America are expected to continue drying under SSP126 scenario in the future owing to the increasing evapotranspiration in a warming climate (Allen et al., 2010; Vicente-Serrano et al., 2010). Alternatively, the obvious drying trend around Canada’s subarctic lakes are attributed to the high vulnerability to droughts when snow cover declines under increasing temperature (Bouchard et al., 2013). Many regions around the Aral Sea and north Russia are prone to experience wet-to-dry transition under climate change. It is worth noting that a higher emission scenario can be translated to a more intensive trend of either drying or wetting, the pattern is also revealed by a recent study (Pokhrel et al., 2021).

We further demonstrate the global distribution of the long-term trends in TWS-DSI over 43 selected SREX regions in Figure 2. During the historical period, 53.5% and 46.5% of land area present drying and wetting trend, respectively, of which 59.4% and 48.4% is significant (p<0.05). NAU has the highest percentage (73.7%) of land area with a significant increasing trend of TWS-DSI, which is mainly caused by precipitation increase (Rajah et al., 2014). While the ARP has the greatest percentage of 81.9% of pixels showing a significant drying trend jointly affected by the groundwater depletion and droughts over the Arabian Peninsula (Lelieveld et al., 2012). During the future period under climate change, the proportion of drying areas with a significant slope increases from SSP126 (23.6%) to SSP585 (30.1%) scenario. Similar growth is detected in the percentage with significant wetting trends, which reaches 15.7%, 17.4%, and 23.5% under SSP126, SSP245, and SSP585 scenarios, respectively. Some mid-latitude regions including WCE, EEU, WSB, WNA, and ECA present wetting trends that
benefit from precipitation increase and become wetter when a high emission scenario is expected to occur. SAS also illustrates a dry-to-wet transformation and a higher radiative forcing from SSP126 to SSP585. Alternatively, all the regions in North America and Russia except for WNA are expected to become dry, and so do some regions over south-eastern Africa, central Asia, and south of Australia. Generally, the percentage of the land area showing significant trends of both wetting and drying stably increase from the SSP126 to SSP585 scenarios, and the drying is always ~10% higher than the wetting.

Figure 1: Global distribution of the long-term trends (left panel) and classification (right panel) in TWS-DSI during (a, b) the historical (1985-2014) and future (2071-2100) period under (c, d) SSP126, (e, f) SSP245, and (g, h) SSP585 scenarios. Note: The stippling marks regions with a significant trend (p<0.05). “D” and “W” indicate regions with drying and wetting trends, respectively.
3.2 Assessment of the DDWW Paradigm

To evaluate the DDWW paradigm over global land, the effective Aridity index (AI) is used to classify a grid cell as arid, humid, and transitional region, following Yang et al. (2019). The AI is calculated as the ratio of annual precipitation to potential evapotranspiration provided by the Climatic Research Unit Time series (CRU TS-v4.05) during the same period as TWS-DSI (i.e., 1985-2014). The global distribution of multi-year average AI and the classifications during the period 1985-2014 is presented in Figure S5. It can be seen that most of the arid regions are located in southwestern America, north and south Africa, central Asia, Arabian regions, and Australia, accounting for 33.6% of the land. The percentage of humid areas that are mainly located in east America, the Amazon region, central Africa, south China, west Europe, and Russia reaches 58.1% of the land. An approximate 8.3% of the land area is defined as the transitional region. We compare AI and TWSA derived from DATASET and CMIP6 between 1985 and 2014 in Figure S6, with the latter presenting a similar spatial distribution to AI. Moreover, the CMIP6 data has a relatively higher amplitude than that of the DATASET, in line with the temporal results (see Figure S4).

Figure 3 illustrates the test results of DDWW paradigm at a 5% significance level ($p=0.05$) during the historical and future periods. Limited proportions (<10%) of area illustrating the “transition gets drier” (TD) and “transition gets wetter” (TW) patterns are reported in both past and future periods. Much of the land area over north Africa, Arabian regions, east Asia, and southwest America show the “dry gets drier” (DD) phenomenon. In contrast to that, a substantial portion of area over the arid regions of the north and south of Africa, Australia, and central Asia shows the “dry gets wetter” (DW) hypothesis. Moreover, the “wet gets wetter” (WW) paradigm is confirmed mainly in east Russia and north Amazon, with the “wet gets drier” (WD) pattern happening in central Africa, north-eastern Amazon, and north Asia. Under climate change, a
similar pattern under the SSP126 scenario is revealed in the historical results. Nevertheless, the SSP245 scenario presents a slightly different distribution from historical results, with many regions in north Asia and central Europe showing DW and WW situations instead of DD and WD. In addition to that, the south and northeast of China, together with the majority of Russia show the WD situation, and the DD paradigm is gradually dominating Australia. This difference is further confirmed based on the results under the SSP585 scenario. We further conducted the regional analysis as shown in Figure 4. Among 54.9% of land showing significant trends in drying and wetting, 50.5% confirms the DDWW paradigm, of which 30.1% and 20.5% are drying and wetting, respectively, during the historical period (1985-2014). ARP has the highest percentage of 96.3% of the area with significant trends showing the DD hypothesis, while RFE achieves the highest proportion of 92.1% presenting the WW theory. During the future period under climate change, multiple SSP scenarios highlight a generally consistent distribution with some minor differences. Some regions such as WCA, WNA, and MED show a DD paradigm while others such as WCE, SES, MDC, and NC have a WW paradigm. Apart from that, a lot of regions located in Russia, Mongolia, and Canada present WD conditions, with few areas like NEAF and ARP showing DW situation. It can be clearly observed that the proportion of regions showing WD pattern increases from SSP126 to SSP585 scenario, indicating the comprehensive trend of drying at region scale.

Global statistics of the regions with various patterns during the historical (1985-2014) and future periods (2071-2100) are shown in Figure 5. During the period 1985-2014, a percentage as high as 54.9% of area shows significant trends in wetting or drying ($p<0.05$). Further, 28.1% of the area shows the DDWW paradigm, in which 16.7% and 11.4% of area is drying and wetting, respectively. 23.3% of the area, however, shows the opposite pattern of DW (8.4%) and WD (14.9%), respectively. The confirmed percentage for the DDWW paradigm (27.1%) for the land mass (represented by TWSA) in our study is more than twice as high as that for the land surface (represented by precipitation, evaporation, and aridity) in a previous study (10.8%) (Greve et al., 2014). Feng and Zhang (2015) used soil moisture to conclude a proportion of 15.12% followed the DDWW pattern while a percentage of 7.7% of the land showed an opposite pattern between 1979 and 2013, relatively lower than our study. Yang et al. (2019) applied a combined measure employing six different drought indices to evaluate the DDWW paradigm and discovered the percentage following and opposing the DDWW paradigm is 29% and 20%, respectively, during the period 1982-2012, typically consistent with our study. Cheng et al. (2020) utilized the GRACE data during 2002-2017 and reported the area having the DDWW pattern reached 11.2% except for the 4.7% of cold regions over global land, which is comparatively lower than our study. Observed differences among various studies are attributed to the differences in datasets used, metrics employed for assessment, and the study period.

Under climate change, the proportion of areas supporting the DDWW paradigm is 18.2%, 17.4%, and 20.7% under SSP126, SSP245, and SSP585 scenarios, respectively. Alternatively, the area having the opposite DDWW pattern achieves 17.9%, 22.4%, and 28.5%, respectively. The percentage of areas with significant wetting and drying trends slightly increases over the enhancement of emission scenarios, consistent with the increase of DDWW-validated areas from SSP126 to SSP585 scenarios. It is worth noting that the internal variability of climate models might affect the potential agreement with the DDWW pattern (Kumar et al., 2015). Greve and Senevirtne (2015) used climate projections from CMIP5 to establish the
measure $P - E$ for assessment of the DDWW paradigm and discovered the hypothesis was validated over 19.5% of land area between 2080 and 2100 under the RCP8.5 scenario, which is close to our result (20.7%). Moreover, Li et al. (2021) further applied the P-E index to test the DDWW theory based on GCMs from the third phase of Paleoclimate Modelling Inter-comparison Project (PMIP3) simulations, concluding a similar proportion of 22.8% of the global land to our study that held the DDWW paradigm. This similarity reveals the similar atmospheric and terrestrial responses under future warming for both the studies.

Figure 3: Global assessment of the DDWW paradigm during the (a) historical (1985-2014) and future (2071-2100) period under (b) SSP126, (c) SSP245, and (d) SSP585 scenarios. Note: DD indicates the dry gets drier; DW indicates the dry gets wetter; WW indicates the wet gets wetter; WD indicates the wet gets drier; TD indicates the transition gets drier; TW indicates the transition gets wetter.

Figure 4: Global assessment of the DDWW paradigm in 43 selected IPCC SREX regions during the (a) historical (1985-2014) and future (2071-2100) period under (b) SSP126, (c) SSP245, and (d) SSP585 scenarios. Note: The fan shape represents the regional proportion of area with different patterns to the total area with significant ($p<0.05$) patterns and the bar plot means the global
percentage. “D” and “W” indicate regions with drying and wetting trends, respectively. DD indicates the dry gets drier; DW indicates the dry gets wetter; WW indicates the wet gets wetter; WD indicates the wet gets drier; TD indicates the transition gets drier; TW indicates the transition gets wetter. Please refer to Figure S1 for abbreviations of the IPCC SREX regions.

Figure 5: Global statistics of the regions with different patterns during the historical (1985-2014) and future (2071-2100) period under SSP126, SSP245, and SSP585 scenarios. Note: DD indicates the dry gets drier; DW indicates the dry gets wetter; WW indicates the wet gets wetter; WD indicates the wet gets drier; TD indicates the transition gets drier; TW indicates the transition gets wetter; Uncertain indicates the regions showing insignificant (p>0.05) trends in TWS-DSI.

3.3 Uncertainties and Implications

Each ensemble member of the DATASET has embedded uncertainties inherently originating from one or more forcing variables, simplified assumptions of complex processes in the models and their physical structure, retrieval algorithms, and systematic biases, which might have inevitably propagated to the results presented herein. For example, the original GRACE mascon observations contain the measurement error and signal leakage at the gridded scale, which persists in the reconstruction of TWSA when training via the statistical methods (Humphrey and Gudmundsson, 2019; Li et al., 2021a). Unlike observed GRACE and reconstructed GRACE-like data, simulations from the models (GHMs, LSMs, and GCMs) are inherently featured by incomplete TWSA representation (Table S2). They are generally based on simplified hydrological processes, resulting in the lack of certain TWSA components. For example, the widely used Noah and VIC models lack surface water and groundwater storage in TWSA (Scanlon et al., 2018). Similarly, GCMs can only simulate the snow water and soil moisture within a limited depth from 2 to 10 m below the land surface (Xiong et al., 2022). This inadequate representation of the land mass (and hence TWSA) in these global models can lead to regional bias in some aquifers with overexploitation of the particular TWSA components (e.g., groundwater depletion in North China Plain). Despite the noted
inevitable uncertainties, a satisfactory accuracy in simulating TWS over large basins worldwide provides confidence in our results (Freedman et al., 2014; Pokhrel et al., 2021; Wu et al., 2021).

Moreover, the eight CMIP6 GCMs are forced with the future projections of many meteorological variables such as precipitation and air temperature, which have been reported to show underestimation or overestimation over the global land (Eyring et al., 2016; Kim et al., 2020). Despite employing bias correction with GRACE data, uncertainty from the forcing and models can influence the accuracy of TWSA simulations (Xiong et al., 2022). Although it is challenging to explicitly attribute and quantify these uncertainties in the absence of a ‘true’ reference observation dataset, the ensemble averaging method has been used to integrate the multi-source TWSA data. The global distributions of NRMSE between GRACE observations and each ensemble member and their mean during April 2002-December 2014 show improved performance of the latter (Figure S7). Three GRACE reconstructions present relatively lower error than the GHMs and LSMs, especially in the high-latitude northern hemisphere, where snow, ice, and glaciers contribute more to TWS than other regions, which is not considered in most of the global models. The ensemble-mean solution illustrates reasonably good accuracy with the NRMSE generally below 0.2, highlighting the reduced uncertainty compared with the individual solution. The GRACE reconstructions compare better than other data because they are directly calibrated with the GRACE measurements during 2002-2017, while their performances need more validation beyond the GRACE era (i.e., prior to April 2002 and during July 2017-June 2018). Similar patterns are also evident from the probability density functions of NRMSE, of which there is an overall negative deviation in the ensemble-mean relative to other solutions except for the CSR reconstruction (see Figure S8). This outperformance of the ensemble dataset is ascertained by the increased correlation and decreased standard deviation as shown by the Taylor diagram (Figure S8). In addition, the comparison between GCM-modelled and GRACE-derived TWSA before and after bias correction in the past (April 2002-December 2014) is conducted (see Figure S9 and S10). The spatial distributions clearly show that the ensemble mean of eight GCMs outperforms each member globally, particularly in Australia, southern Africa, and North America. The outperformance become more obvious after bias correction. An overall decrease in NRMSE is also observed according to the probability density functions after performing bias correction, which is also detected from the Taylor diagram results (see Figure S11). The largely reduced bias after bias correction and ensemble averaging give us confidence for the future projection of TWSA.

To investigate the influence of different models and datasets on the robustness of the re-examination for the DDWW paradigm, we carried out an independent analysis at the individual member level (see Figure S12). For the historical period (1985-2014), a clear overestimation of the CSR reconstructions is detected, with 42.4% of the area agreeing with the DDWW pattern and 36.6% showing the opposite situation. Moreover, the modeled results from VIC and WGHM illustrate the underestimation of the area validating the DDWW paradigm, reaching 15.6% (WGHM) and 12.2% (VIC), respectively. Their proportion with the opposite DDWW paradigm is 10.2% (WGHM) and 17.8% (VIC), respectively. Therefore, it can be concluded that the differences among different members of DATASET limitedly affect the evaluation of the DDWW during the historical period. In the future (2071-2100), the GFDL-ESM4 model presents overestimation but the IPSL-CM6A and CanESM5 models have underestimation for different percentages compared with the ensemble mean. Specifically, the area
dominated by the DDWW paradigm changes from 8.9% (CanESM5) to 21.9% (GFDL-ESM4), while that showing the opposite pattern ranges from 7.8% (CanESM5) to 14.8% (GFDL-ESM4) under the SSP126 scenario. For the SSP245 scenario, the DDWW-validated regions account for 7.4% (CanESM5) to 21.5% (GFDL-ESM4), the opposite pattern occurs over a range from 9.7% (CanESM5) to 16.0% (GFDL-ESM4) of land. The proportion supporting the DDWW paradigm varies from 10.4% (CanESM5) to 24.0% (GFDL-ESM4), while that presenting the opposite pattern ranges from 8.4% (CanESM5) to 22.3% (GFDL-ESM4) under the SSP585 scenario. Overall, the comparatively large difference among various models might source from unforced internal climate variability of distinctive CMIP6 members and different emission scenarios (Kumar et al., 2015).

Our choice of the significance level (i.e., 0.05) may also affect the rationale of the DDWW examination results, thus different significance levels are alternatively tested (see Figure S13). At a significance level of 0.01, 22.2% of land area agrees well with the DDWW theory, while the 17.1% of area illustrates the opposite pattern during the period 1985-2014. As for the 0.1 significance level, the DDWW-validated regions account for 30.6% of the total area, with 25.4% of land agreeing with the opposite hypothesis. In the future period, a similar pattern is discovered that both DDWW-confirmed and DDWW-opposed regions are increasing on account of the enhancement of projected strength of radiative forcing, with the reduction of the area showing insignificant trends in wetting and drying. However, the magnitudes of results at the 0.01 significance level are generally lower than that at the 0.1 significance level due to the different thresholds of detected trends in drying and wetting. Considering the similar tendency with marginal effects of the varying choices of the p-value (e.g., 8.4% change in DDWW area from 0.01 to 0.1 level), our adopted significance level (i.e., 0.05) can reasonably explain the global trends of dryness/wetness.

Despite the multisource uncertainties, our study provides important implications for the long-term trends in dryness/wetness over global land in the past and future from the perspective of TWSA. Compared with other widely used indexes that are purely derived from the hydrometeorological variables (e.g., SPI, SPEI, and PDSI) or incorporate a single component of the TWSA (e.g., SSI, SGI, and SRI), our developed TWS-DSI describes the overall status of the land system, which is jointly influenced by different components including soil moisture, river runoff, and groundwater that play different roles in the hydrological cycle (Tapley et al., 2019). Furthermore, the projected changes in global TWSA and associated TWS-DSI improve our understanding of the large-scale hydrological response to climate change, particularly in regions with strong human interventions, such as the south and east of Asia. Despite the magnitude bias from satellite products, simulations of LSMs and GHMs, and GCMs projections, the ensemble averaging method has presented an effective and efficient ability to alleviate the multi-source uncertainty, which can be further applied over data-sparse areas globally with limited in-situ observations like Africa and central Asia. In addition, the regional aggregation of the analysis based on the IPCC AR6 SREX references regions provides an improved understanding of the TWSA dynamics in the past and valuable inferences for policymakers and stakeholders for better water resources management in a changing environment (Iturbide et al., 2020).
4 Conclusion

In this study, the historical TWSA series over global land during 1985-2014 was calculated from the ensemble mean of nine model outputs, including three each from GHMs (VIC, WGHM, PCR-GLOBWB), LSMs (Noah, CLSM, CPC), and GRACE reconstructions (CSR, JPL, GSFC). Future TWSA projections from 2070 to 2100 under SSP126, SSP245, and SSP585 scenarios were derived from the average of eight selected CMIP6 GCMs after bias-correction using GRACE observations. Subsequently, TWS-DSI was estimated to detect the long-term trends in dryness/wetness in the past and future periods. Further, the DDWW paradigm has been re-examined with a significance level of 0.05 from the perspective of terrestrial water storage change. The uncertainty sourced from different choices of models, methods, and confidence levels has been discussed systematically. The new findings are summarised as follows.

1. During the historical period, 32.9% and 22.1% of land area present significant ($p<0.05$) drying and wetting trends, respectively. During the future period under climate change, the proportion of drying areas (always ~10% higher than wetting) with a significant slope increases from SSP126 (23.6%) to SSP585 (30.1%) scenario. A similar change is detected in the percentage with significant wetting trends, which reaches 15.7%, 17.4%, and 23.4% under SSP126, SSP245, and SSP585 scenarios, respectively.

2. A total of 28.1% of the global land area shows the DDWW paradigm valid, in which 16.7% and 11.4% of the area is drying and wetting, respectively during the period 1985-2014. 23.3% of the area, however, shows the opposite pattern like “dry gets wetter” (DW, 8.4%) or “wet gets drier” (WD, 14.9%), respectively. The proportion of areas supporting the DDWW paradigm is 18.2%, 17.4%, and 20.7% under SSP126, SSP245, and SSP585 scenarios, respectively. Alternatively, the area opposing the DDWW paradigm achieves 17.9%, 22.4%, and 28.5%, respectively.

3. The ensemble mean of TWSA generally compares better with GRACE observations during 2002-2014 than the individual solution, especially for the eight bias-corrected CMIP6 GCMs. Independent experiments based on the individual TWSA dataset suggest that the divergent choices of data source might lead to reasonable overestimations (CSR mascon) and underestimations (WGHM and VIC) for both the DDWW-agreed and DDWW-opposed patterns. Moreover, the use of distinctive GCMs suggests slightly overrated (GFDL-ESM4) and underrated (CanESM5) percentages of DDWW-pro and DDWW-con area in the future under multiple emission scenarios.

4. Sensitivity analysis on different choices of significance levels from 0.01 to 0.1 indicate similar patterns, in which 22.2% (17.1%) of the land area supports (opposes) the DDWW theory historically under the 0.01 level, and the DDWW-validated regions account for the 30.6% of total area with 25.4% of land agreeing with the opposite hypothesis under the 0.1 level. Such consistency is also evidenced by the projected TWS-DSI in the future under various scenarios.

New insights from the TWSA perspective highlight that the widely-used DDWW paradigm is still challenging in both historical and future periods under climate change. In addition, our developed ensemble-mean method can effectively and efficiently alleviate the uncertainty sourced from different data sources, implying an alternative way to assess the TWSA variations over major basins globally. The regional aggregation of our study based on IPCC SREX reference regions can...
provide important inferences for decision-makers and stakeholders for the sustainable management and efficient utilization of water resources under global change.

**Data Availability**


**Supplement**

The supplement related to this article is available online.

**Author contributions**

Jinghua Xiong conceived and designed the experiments. Jinghua Xiong performed the experiments. Jinghua Xiong and Abhishek analyzed the data. Jinghua Xiong, Shenglian Guo, Abhishek, Jie Chen, and Jiabo Yin wrote and edited the paper.

**Competing interests**

The authors declare that they have no conflict of interest.

**Acknowledgments**

This study was financially supported by the National Key Research and Development Program of China (2021YFC3200303), the National Natural Science Foundation of China [U20A20317]. The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.
References


