Reply to Reviewers' comments (Reviewer#1)

Legend

Reviewers' comments

Authors' responses

Direct quotes from the revised manuscript

We thank the reviewer for his/her time in reading our manuscript and detailed comments on our manuscript. Point-by-point replies to the comments or suggestions made can be found below. Overall, we have made the following major changes to the manuscript:

- Performed additional analysis using P-E-R and analyzed and compared the results with TWSA-DSI.
- Instead of showing only the ensemble mean of various model and observation-based results, we have now shown the results from individual datasets during the historical period (1985-2014).
- Added extensive discussion about the various mechanisms and governing processes for the observed patterns and the similarities and disparities from the previous studies.

Reviewer #1: This is a resubmitted manuscript and this is my second review. This study re-examine the "dry gets drier and wet gets wetter" (DDWW) paradigm using the terrestrial water storage anomaly (TWSA) derived from GRACE observational products, land surface models, and GCMs. The results showed the global patterns of dryness/wetness trends in both history (1985-2014) and future (2071-2100).

In this version, the authors have improved the text, added discussion, and provided more uncertainty analyses. I am happy with the authors' efforts. However, there are substantial issues which need to be addressed. The authors should set out to solve scientific problem rather than analyzing data. At present, I did not feel the new knowledge and new (and convincing) methods provided by this paper. At least, the authors have not fully express the innovation and significance of this study.

Response: Thank you very much for the second review of our manuscript and for encouraging feedback. We provide the global evaluation of the topical DDWW paradigm from the TWSA perspective in the past and future, which has never been performed before. Furthermore, we have added mechanism analysis of the TWSA patterns in the revised version, including comparison with the conventionally used wetness/dryness metric P-E-R and performing regional studies with significantly decreasing TWS-DSI over the Qinghai-Tibetan Plateau. New knowledge shows the contradictory DDWW patterns in terrestrial land mass over global land for the past and future based on a total of 18 datasets, including LSMs, GHMs, GCMs, and observation-based products. We have highlighted the additional insights and have clarified the innovation and significance of this study in the new version.

Concept: The authors should recall the original meaning of "dry gets drier and wet gets wetter" paradigm from existing studies, because the title/the authors intend to perform a “re-examine” work. I think the authors acknowledge that the DDWW rule is used to explain the changing trend of surface dryness/wetness or climate condition, while this study explains the DDWW rule from a TWS perspective which includes groundwater/glacier changes. The GRACE observation contains the signal of changes in groundwater/glacier. As the climate warms, ice/glaciers are degrading with an increase in runoff/soil moisture (moisten the land surface). Meanwhile, as the mass decreases (water flows away), what GRACE observes is a decrease trend in gravity (drying). There are processes in the opposite direction. As such, the TWSA trends can be opposite with the previous studies focusing on the land-surface conditions (soil moisture/runoff and ET) (Wang et al., 2021: Long-term relative decline in evapotranspiration with increasing runoff on fractional land surfaces; Yang et al., 2019: Combined use of multiple drought indices for global assessment of dry gets drier and wet gets wetter paradigm). Rather than a new perspective, I would also think of this study as a simulation or an application of GRACE, land surface models, and climate models.

Response: We thank the reviewer for this important suggestion about the title.

Title Change: Since we provide the first global evaluation of the DDWW paradigm from the TWSA perspective in both the past and future, we have updated the title to - ‘Global evaluation of the dry gets drier and wet gets wetter paradigm from terrestrial water storage changes perspective’ to better reflect the approach and contents of the manuscript.

Importance of and need for TWSA perspective: We agree that conventionally DDWW paradigm has been studied either by directly using the two competing variables, i.e., precipitation and evapotranspiration (Held and Soden, 2006), or the derived indices such as P/ET (Greve et al., 2014), SPEI (Yang et al., 2019), and PDSI (Hu et al., 2019). However, there has been increasing attention to the DDWW paradigm from different perspectives (e.g., soil moisture (Feng and Zhang, 2015) and runoff (Yang et al., 2019)) in the last decade. Inconsistent usage of the term “wetter” and “drier” across disciplines (Roth et al., 2021) and different physical meanings of these variables further limit their implications in the context of total land water storage. Since the terrestrial/land water storage (i.e., TWS) is a crucial variable for the community working on, e.g., ecosystem functioning (Humphrey et al., 2018), sea-level budget studies (Frederikse et al., 2018), terrestrial water balance, hydroclimatic extremes, and freshwater availability (Rodell et al., 2018), it merits indispensable consideration. Given the different meaning of TWSA with previous metrics (e.g., P and ET), the evaluation of the DDWW paradigm from the TWSA perspective and inter-comparison and subsequent analysis of governing processes/mechanisms as carried out in our study can potentially provide new evidence. Please also see our responses to the major comments on the ‘methods’ below and response to specific comment#4 for a detailed explanation of the similarity, differences, and significance of TWSA compared to P/E (or derived indices).

As rightly indicated by the reviewer, unlike many evaluations focusing on the land surface water balance (e.g., P-E), our developed TWS-DSI contains the signals of groundwater/glacier that were impossible to be considered previously. We regret and are surprised that the reviewer has an impression of the revised manuscript (with, in our understanding, all previously raised concerns resolved) different than that of the original manuscript (‘The topic is interesting and this study potentially provides a new perspective.’). However, we have changed the title and have tried to thoroughly incorporate all the suggestions in this version.

References:

Feng, H., Zhang, M., 2015. Global land moisture trends: drier in dry and wetter in wet over land. Sci. Rep. 5, 18018. https://doi.org/10.1038/srep18018

Frederikse, T., Jevrejeva, S., Riva, R. E. M., and Dangendorf, S.: A Consistent Sea-Level Reconstruction and Its Budget on Basin and Global Scales over 1958–2014, J. Climate, 31, 1267–1280, https://doi.org/10.1175/jcli-d-17-0502.1, 2018.

Held, I. M., Soden, B. J. 2006. Robust responses of the hydrological cycle to global warming. Journal of climate, 19(21), 5686-5699.

Humphrey, V., Zscheischler, J., Ciais, P. et al. 2018. Sensitivity of atmospheric CO2 growth rate to observed changes in terrestrial water storage. Nature 560, 628–631. https://doi.org/10.1038/s41586-018-0424-4

Hu, Z.Y., Chen, X., Chen, D.L. Li, J.F., Wang, S., Zhou, Q., Yin, G., Guo, M. 2019. “Dry gets drier, wet gets wetter”: a case study over the arid regions of central Asia. Int J Climatol 39(2):1072–1091

Rodell, M., Famiglietti, J.S., Wiese, D.N. et al. 2018. Emerging trends in global freshwater availability. Nature 557, 651–659. https://doi.org/10.1038/s41586-018-0123-1

Roth, N., Jaramillo, F., Wang-Erlandsson, L., Zamora, D., Palomino-Ángel, S., Cousins, S. A. 2021. A call for consistency with the terms ‘wetter’and ‘drier’in climate change studies. Environmental Evidence, 10(1), 1-7.

Yang, T., Ding, J., Liu, D., Wang, X., Wang, T., 2019. Combined Use of Multiple Drought Indices for Global Assessment of Dry Gets Drier and Wet Gets Wetter Paradigm. J. Clim. 32, 737–748. https://doi.org/10.1175/JCLI-D-18-0261.1

Method: In the discussion, the authors need to justify why it is necessary to assess the changes in dryness/wetness from a perspective of terrestrial water storage change? What are the advantages of the methodology used in this method? This study has many redundant operations (e.g., the use of GRACE to correct GCMs). I feel if the study directly using P-ET is more convincing than using partial outputs (soil moisture, snow water...). While changes in TWSA do not equal to changes in surface dryness/wetness, there should have "bridges" to connect the integrated TWS and various land-surface processes (runoff, soil moisture and ET) (Trautmann et al, 2022: The importance of vegetation in understanding terrestrial water storage variations). It is a pity that this study did not find such “bridges” as it leaned toward analyzing data. The use of TWS retrieved by the GRACE to correct GCM simulations is not convincing. Not only are there many uncertainties in the GRACE retrieval product, but also what GRACE observes is completely different from what GCMs simulate (Table S2). Since these models express different objects, how can these outputs ensemble? What will happen if the study do not use GRACE to correct the GCM simulations as most climatologists do? There are still have a prediction result from GCM, right? What are the differences? One way is to show that the corrected results are more reliable than the previous one, which may involve using in-situ observed data. Moreover, the authors criticize the use of P-ET as an indicator to identify dry/wet changes, but I think P-ET is closer to changes in TWS because various hydrological models and GCMs appear to do not account for surface water storage. I suggest the authors provide a technical route.

Response: We thank the reviewer for the comment. Please find the detailed explanation of all the concerns below.

<u>Significance of TWSA and performance disparity:</u>

Let us briefly discuss the significance of TWS using two examples from a process perspective, i.e., ecosystem functioning (Humphrey et al., 2018) and freshwater availability (Rodell et al., 2018). The interannual fluctuations of TWSA significantly influence the terrestrial carbon sink and are essential for the global water and carbon cycles-two major cycles of the earth system sciences (Humphrey et al., 2018). Its long-term trends are also indicative of the global water's landscape influenced by climate variability, climate change, and human activities and offer important inferences for global water and food security (Rodell et al., 2018). Although the amount of water stored in land is governed by the precipitation (and evapotranspiration and runoff) influxes (outfluxes), the change in the storage is governed by the synergistic impact of climatic and human-induced changes, which are imperative for a wide range of the applications. Hence, we infer that water storage is more relevant than P or ET or a combination thereof. Please also see our response to the previous comment for more details on the need and importance of the TWSA perspective.

References:

Humphrey, V., Zscheischler, J., Ciais, P. et al. 2018. Sensitivity of atmospheric CO2 growth rate to observed changes in terrestrial water storage. Nature, 560, 628–631. https://doi.org/10.1038/s41586-018-0424-4

Rodell, M., Famiglietti, J.S., Wiese, D.N. et al. 2018. Emerging trends in global freshwater availability. Nature, 557, 651–659. https://doi.org/10.1038/s41586-018-0123-1

<ins>Advantages of the methods used</ins>:

Our study establishes the normalized TWS-DSI index based on different TWSA datasets, which accounts for the regional hydro climatological variability and is suitable for comparing dryness/wetness status for different locations and periods (Zhao et al., 2017). In addition, the modified Mann-Kendall test used for the long-term trend estimations could avoid the autocorrelation of the time series (Hamed and Rao, 1998). The future projections from CMIP6 GCMs are bias-corrected using the trend-preserving method (Hempel et al., 2013). We fully agree the GCM simulations have considerable uncertainties, which might be further strengthened over regions with significant variations in vegetation, surface water, and groundwater due to the constrained representations of TWS in GCMs. To show the difference, we selected two typical regions (i.e., Amazon and Mekong River basins) with abundant surface and groundwater resources (Pham et al., 2019), of which the Mekong River basin experienced severe human interventions such as groundwater pumping, dams constructions, and city extension while the Amazon River basin is considered as one of the largest natural river basins with low urbanization and human activities (Xiong et al., 2022). It is discovered that the GCM simulations without bias correction show obvious underestimations over two regions with large uncertainty, however, which have significantly reduced after bias correction along with a lower spread range (Figure R1). The amplitudes of the GCM series are adjusted to nearly the same as GRACE data, with the long-term trends unaffected.

Moreover, given the favorable consistency between CMIP6 GCMs TWS and both GRACE observations and in-situ measurements (Wu et al., 2021), our bias-correction based on GRACE data is expected to decrease their differences derived from the missing key TWSA components of the models by comparing with GRACE observations (Figures R2). Additionally, independent evaluation of GCM TWSA with and without bias-correction against the water balance estimates of TWSA changes from the observational products of CRU P, GLEAM ET, and GRUN R also presents a satisfactory correlation temporally and spatially (Figures R3-R4). The regions with good accuracy, like Alaska, western parts of the Tibetan Plateau, and northern Russia, decrease after bias correction. These differences over the high-latitude regions might be explained by the simplified treatment of permafrost in GCMs due to the prevailing uncertainties in, e.g., changes in thermophysical properties of the soil during freezing and thawing cycles (Burke et al., 2020). On the contrary, the areas with relatively poorer accuracy before bias correction, such as North Africa and northern South America, slightly improve.

Notwithstanding the observed differences in some regions, our trend-preserving method used for bias correction would not influence the long-term trend estimations of both TWSA and TWS-DSI, and, therefore not impact our evaluation of the DDWW paradigm (Hempel et al., 2013). We take the ensemble mean of eight selected GCMs since all of them can simulate soil moisture and snow water. As suggested by the reviewer, we do not process the historical datasets similarly due to the different objects of different ensemble members (e.g., Table R1 and Table R2). It is noteworthy that the trend-preserving method would not affect the long-term trends of the GCM TWSA, and, therefore not influence our DDWW evaluation results in any way.

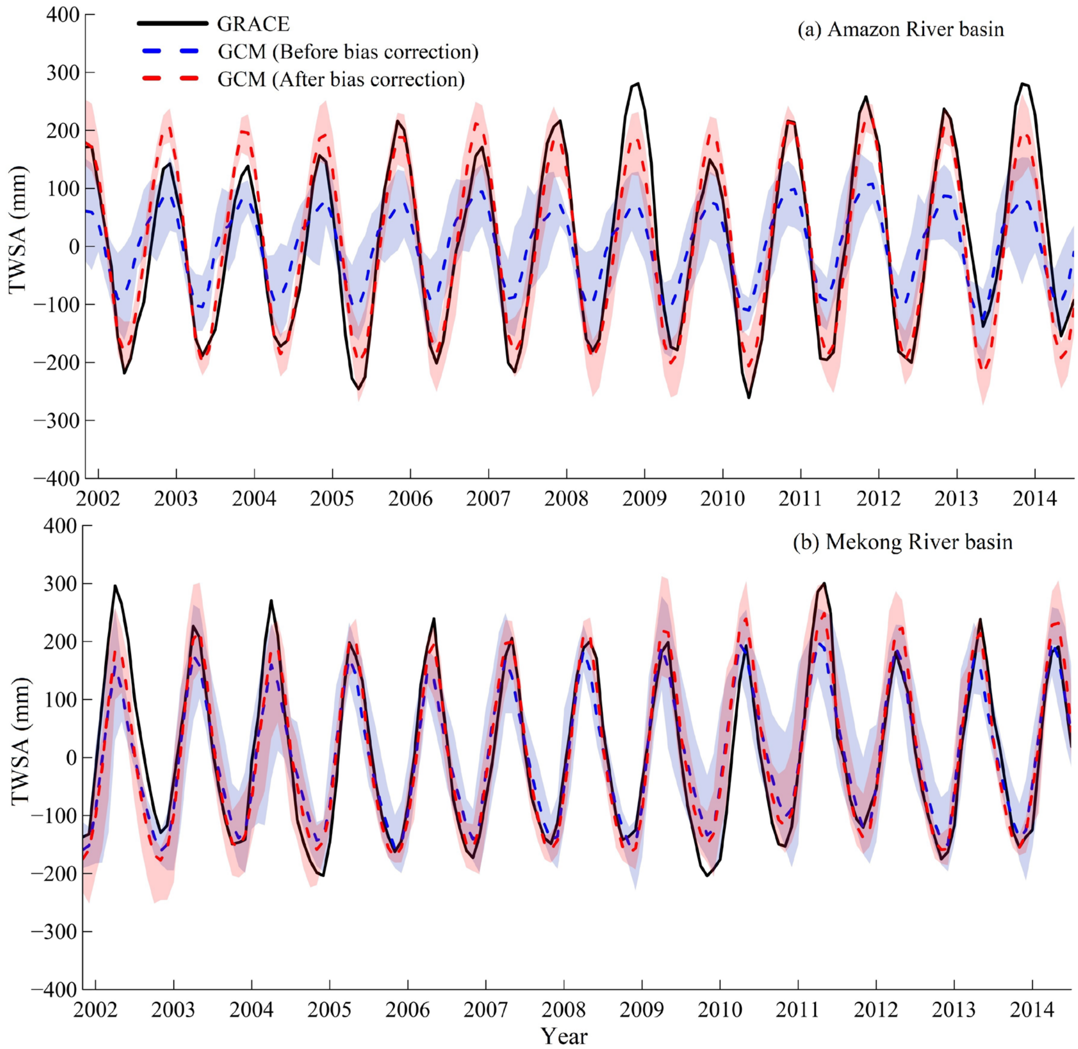

Figure R1. Monthly TWSA from GRACE and GCMs with and without bias correction in (a) Amazon and (b) Mekong River basins during 2002-2014. Note: The shading region means the spread of the GCM ensemble.

Table R1. Summary of attributes of different datasets used in this study.

Dataset	GRACE	WGHM	VIC	Noah	CLSM	CMIP6
Parameter	Satellite	GHM		LSM		GCM
Surface water storage	√	√	×	×	×	×
Soil moisture	√	√	√	√	√	√
Groundwater storage	√	√	×	×	√	×
Canopy water	√	√	√	√	√	×
Snow water	√	√	√	√	√	√
Soil layers (no.)	/	1	3	4	10	5~10
Soil depth (m)	/	2	2	2	1	2~10

Table R2. Summary of the changes in the DDWW test results over global land during 1985-2014.

Model/dataset	Previous results (ensemble mean of DATASET)	Updated results (individual datasets) [range]	Remark
DD	16.7%	6.47%-20.17%	From the perspective of TWSA, the DDWW is still challenged based on both the ensemble mean (previous version) and the individual datasets (current version) used in this study.
DW	8.4%	5.42%-16.13%	
WW	11.4%	4.54%-20.67%	
WD	14.9%	4.79%-19.3%	
TD	2.1%	0.95%3.88%	
TW	1.8%	0.73%-2.63%	
Non-significant	45.1%	17.2%-72.42%	

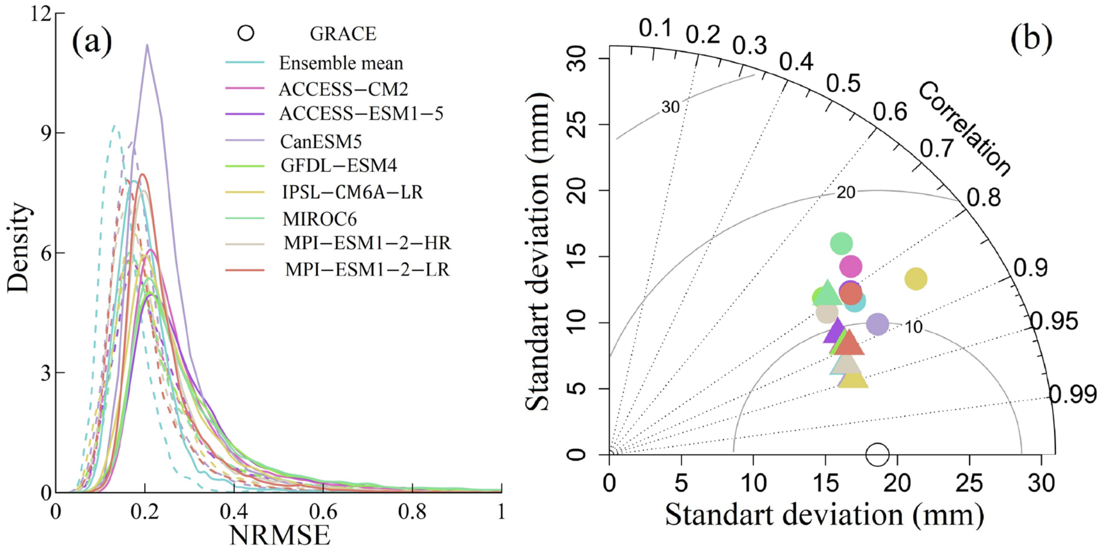

Figure R2. (a) Probability density function and (b) Taylor diagram of NRMSE between TWSA derived from the GRACE mission and each member and the ensemble mean of eight GCMs during the period April 2002-December 2014. Solid and dashed lines in sub-figure (a) and corresponding filled circles and triangles in sub-figure (b) denote the original and bias-corrected time series.

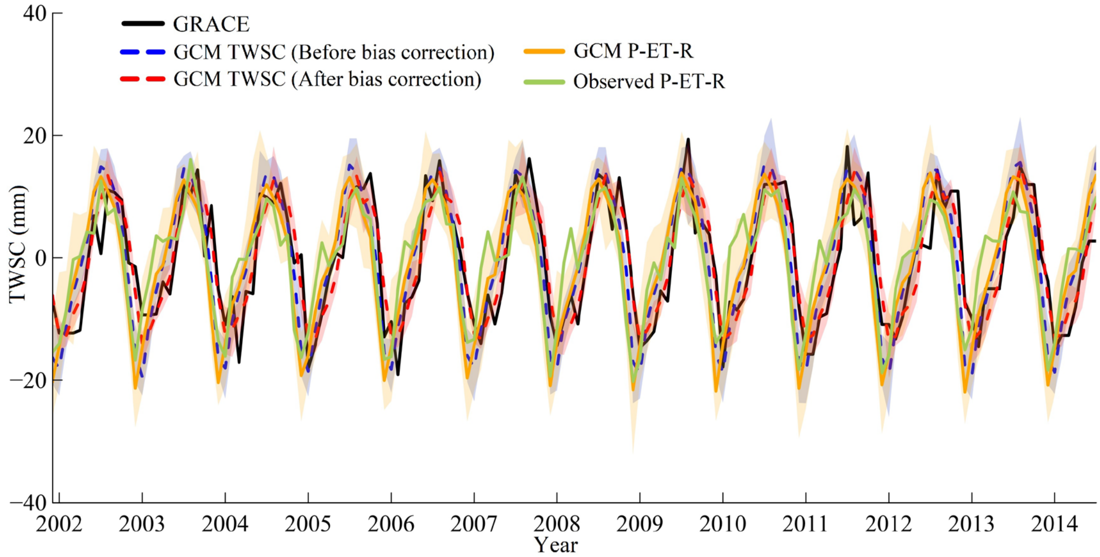

Figure R3. Time series of the monthly changes in TWSA (TWSC) and water balance estimates (i.e., P-E-R) derived from GRACE, GCM, and observations during 2002-2014. Note: The shaded regions represent the spread of the CMIP6 ensemble.

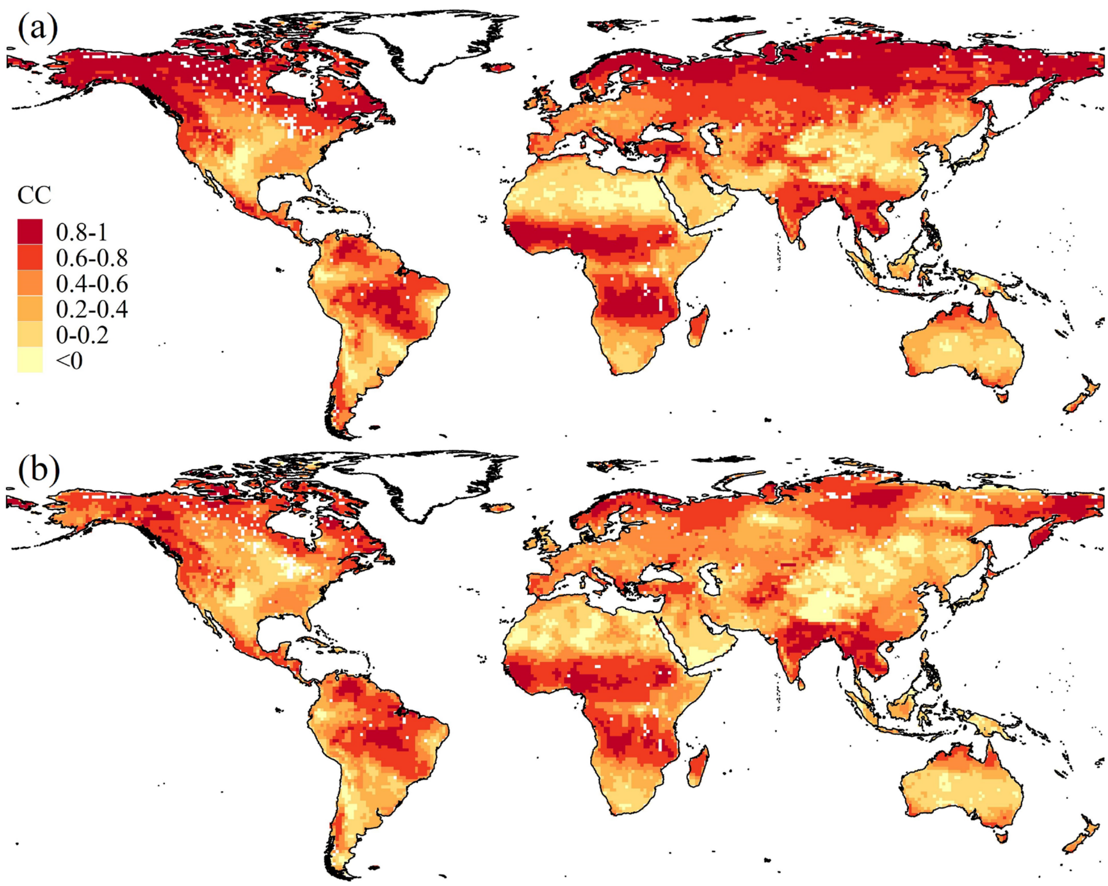

Figure R4. Spatial distribution of correlation coefficient between monthly water balance estimates of TWSA changes and the ensemble mean of GCM data (a) before and (b) after bias corrections during 1985-2014. The blank grids indicate the missing values of the datasets.

References:

Burke, E.J., Zhang, Y., Krinner, G. 2020. Evaluating permafrost physics in the coupled model intercomparison project 6 (CMIP6) models and their sensitivity to climate change. Cryosphere., 14 (9) , pp. 3155-3174

Pham-Duc, B., Papa, F., Prigent, C., Aires, F., Biancamaria, S., and Frappart, F. 2019. Variations of surface and subsurface water storage in the Lower Mekong Basin (Vietnam and Cambodia) from multisatellite observations. Water, 11(1). https://doi.org/10.3390/w11010075

Hamed, K. H., Rao, A. R. 1998. A modified Mann-Kendall trend test for autocorrelated data. Journal of hydrology, 204(1-4), 182-196.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., Piontek, F., 2013. A trend-preserving bias correction: the ISI-MIP approach. Earth Syst. Dyn. 4, 219–236. https://doi.org/10.5194/esd-4-219-2013

Xiong, J., Yin, J., Guo, S., He, S., Chen, J, Abhishek. 2022. Annual runoff coefficient variation in a changing environment: a global perspective. 6, 064006. 10.1088/1748-9326/ac62ad.

Wu, R.-J., Lo, M.-H., Scanlon, B.R., 2021. The Annual Cycle of Terrestrial Water Storage Anomalies in CMIP6 Models Evaluated against GRACE Data. J. Clim. 34, 8205–8217. https://doi.org/10.1175/JCLI-D-21-0021.1

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. J. Clim. 30, 6297–6308. https://doi.org/10.1175/JCLI-D-16-0458.1

<u>Due accreditation and comparison of other approaches</u>: We regret the unintentional impression that we criticize or curtail the value (whether methods or applicability) of other approaches/metrics (e.g., P, P-E, etc.). We just intend to highlight the differences in the governing processes and hence the applicability of the various metrics. We have modified or rather weakened such instances in the revised manuscript. We also calculate a new metric called P-E-R as the residual of precipitation, evapotranspiration, and runoff, which represents the changes in TWSA in terms of the water balance equation (Famiglietti and Rodell, 2013). It means different from the TWSA and its derived metric (TWS-DSI), the latter represents the actual status of land over the long-term baseline, while the P-E-R means its changes. Thus, we have added the cross-comparison between the two indexes, attempting to investigate the mechanism of the variations in dryness/wetness of the land by comparing their differences and bridging the total mass changes to the land surface water balance. For example, a dry year in an agriculturally dominant basin (e.g., the Ganges basin in India) will trigger more groundwater extraction leading to a more acute decrease in TWSA (primarily due to evaporation losses) than the corresponding decline in P itself. In this case, the ‘soil moisture’ may exhibit positive trends, thus providing ambiguous interpretations. Such issues are not prevalent in our TWSA-based assessment. Therefore, although other indices (e.g., based on P, ET, soil moisture, or a combination thereof) may undoubtedly perform at par for the specific variable in question, they tend to present equivocal inferences for the total water storage. It can be easily understood by the example of soil moisture or evapotranspiration-based indices in a highly irrigated area such as the Ganges river basin. TWS is unremittingly declining due to the overexploitation of groundwater for agriculture in this region (Rodell et al., 2009), while E or soil moisture may have positive trends, thus attenuating the actual TWS situation. Such ambiguities across the prevailing metrics further strengthen our research hypothesis and objectives.

We have clarified our workflow in the method section. We hope our revisions will put forward our results in a more robust way.

References:

Famiglietti, J. S. Rodell, M. 2013. Water in the balance. Science. 340 (6138), 1300–1301. doi:10.1126/science.1236460.
Rodell, M., Velicogna, I., Famiglietti, J.S., 2009. Satellite-based estimates of groundwater depletion in India. Nature 460, 999–1002. https://doi.org/10.1038/nature08238

Results and mechanism: This study does not involve mechanism analysis, and does not analyze why some typical places are getting drier or wetter. Fig. 2 and Fig. 4 make no sense as they are another displays of the same results in Fig. 1 and Fig. 3. Although this division method was used in the IPCC6 and even considered popular by the authors, it did not bring any innovative insights to this study. Moreover, they are difficult to interpret. Instead, the readers are more care about how dryness/wetness changes in time and why there are changes happen.

Response: As suggested, we have added mechanism analysis based on the comparison with the new metric P-E-R and presented the temporal changes of dryness/wetness over the selected typical region of Qinghai-Tibetan-Plateau. Please also see the third subsection of the response above for further details. Moreover, we have removed Figures 2 and 4 in the updated manuscript and have restrained our analysis on the global land only. We have shared the data used in the manuscript figures, as well as the historical datasets and bias-corrected CMIP6 members to enable the reproducibility of the results at the required spatial scales (e.g., basin scales).

Innovation and significance: The authors need to rethink and justify what are the new results or developments reported in this study? Why are these new results or developments significant?

Response: We have highlighted the new findings and the significance reported by our study in the conclusion of the revised manuscript as follows:
Conclusion:

In this study, the historical TWS-DSI monthly time series over global land during 1985-2014 is calculated from an ensemble of two GHMs (VIC and WGHM), two LSMs (Noah and CLSM), and one GRACE reconstruction. In addition, future projections of TWS-DSI from 2071 to 2100 under SSP126, SSP245, and SSP585 scenarios are derived from the average of eight selected CMIP6 GCMs after bias-correction using GRACE observations. Subsequently, we detect the long-term trends in dryness/wetness in both the past and future periods based on TWS-DSI. Further, the DDWW paradigm has been evaluated with a significance level of 0.05 from the perspective of terrestrial water storage change. We also establish the metric P-E-R based on multiple observational products and from the same models as the TWS-DSI for comparison. 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, the percentages of global land area presenting significant ($p<0.05$) drying and wetting trends range from 13.06% (WGHM)-43.35% (GRACE reconstruction) and 13.7% (CLSM)-39.43% (GRACE reconstruction), respectively. The wetting trends are mainly in North Australia, North and South Africa, South and Northwest China, western South America, central United States, and East

Russia. While the drying trends are found in Arab region, West Brazil, Northeast Asia, and southern and northern American continent. During the future period under climate change, the proportion of drying areas (always ~10% higher than wetting) with a significant slope increases from SSP126 (19.52%) to SSP585 (29.04%) scenario. A similar change is detected in the percentage with significant wetting trends, which reaches 11.48%, 13.01%, and 18.42% under SSP126, SSP245, and SSP585 scenarios, respectively.

(2) A total of 11.01% (VIC) to 40.84% (GRACE reconstruction) of the global land area shows the DDWW paradigm valid, in which the drying and wetting area account for 6.47% (VIC)-20.17% (GRACE reconstruction) and 4.54% (VIC)-20.67% (GRACE reconstruction), respectively during the period 1985-2014. However, the area showing the opposite patterns, like "dry gets wetter" (DW) or "wet gets drier" (WD), account for the 10.21% (WGHM)-35.43% (GRACE reconstruction) of the global land, respectively. The proportion of areas supporting (opposing) the DDWW paradigm is 14.66% (16.76%), 14.26% (18.72%), and 17.08% (26.64%) under SSP126, SSP245, and SSP585 scenarios, respectively.

(3) Parallel estimates of the water balance variables and their comparison with the TWSA-based analysis, on the one hand, shed light on the governing mechanisms and translation of hydrometeorological fluxes to the land water storage, on the other hand, outline additional insights into the varying and sometimes even contrasting behavior of the various metrics.

(4) Sensitivity analysis on different choices of significance levels from 0.01 to 0.1 for the long-term trends indicates similar patterns, in which the maximum decrease (increase) in the DDWW-validated regions reaches –7.4% (4.47% historically under the 0.01 (0.1) level, respectively. Such consistency is also evidenced by the projected TWS-DSI in the future under various scenarios. Moreover, independent experiments based on the individual TWSA datasets suggest that the divergent data sources might lead to model-variable biases for both the DDWW-agreed and DDWW-opposed patterns. The use of distinctive GCMs also suggests slightly overrated (e.g., GFDL-ESM4) and underrated (e.g., CanESM5) percentages of such patterns in the future under multiple emission scenarios.

New insights from the TWSA perspective highlight that the widely-used DDWW paradigm is still challenged in both historical and future periods under climate change. The differences between test results based on P-E-R imply the robustness of our developed TWS-DSI in capturing the total land water variations induced by climate changes and human activities, suggesting potentially new knowledge in the land hydrology field. The regional aggregation of our study in the Qinghai-Tibetan Plateau can provide important inferences for decision-makers and stakeholders for the sustainable management and efficient utilization of water resources under global change.

Specific comments:

(1) Line 9-10 and Line 17-18: These statements are contradicted. You are saying the

DDWW is challenged due to the choice of different metrics and datasets used, but you also stated the different data sources have subtle influences on the evaluation results.

Response: We have modified the introductory statement in the revised manuscript as follows:
However, the paradigm is largely conditioned by the choice of different metrics and datasets used and is still unexplored from the perspective of terrestrial water storage anomaly (TWSA).

(2) Line 21: "The hydrological conditions of the land surface have experienced...". The first sentence of this manuscript is talking about land surface condition. This is contrary to the author's argument that they are not concerned with the surface dryness/wetness, but with the entire land system.

Response: We regret the misleading articulation. We have modified the text, which now follows the order: Introduction and importance of hydrological cycle>DDWW paradigm>literature review dealing with P-E>introducing TWS>research hypothesis and objectives of our study.

(3) Line 37: What are oceanic records?

Response: It means the oceanic observations that provide environmental information for marine management, such as air temperature, precipitation, and evaporation (OOPC, Ioc-goos-oopc.org. Retrieved 11 June 2022). We have revised it.

(4) Line 45: P-ET is the amount of water remaining in the land system, but the components in the GCMs (soil moisture and snow water) and VIC (moisture), Noah (soil moisture, snow, and canopy water) are parts of the water stored in the land system (Table S2), and thus the models lack some components of the terrestrial water storage.

Response: We understand the reviewer's point of view. However, we would like to take this opportunity to explain the difference that hinges on 'hydroclimate' variables and the overall status of the land water storage, i.e., 'TWS'. Although absolute hydroclimatic variables and their changes are interrelated by the conservation of water mass and energy, their magnitude of change may not be consistent (Roth et al., 2021; Huntington et al., 2006; Dirmeyer et al., 2016; Labat et al., 2004). For example, an increase in precipitation in time does not necessarily imply an increase in river water availability—if accompanied by a steep increase in evaporation by more thermal energy availability, runoff can, in fact, decrease (Bosson et al., 2012, Katul et al., 2022).

Therefore, based on changes in precipitation or evapotranspiration, or runoff alone, it cannot be concluded how will be the variability of the total water storage. Since all these hydroclimate variables are intricately affected by natural or human factors or a combination thereof, an out-and-out separation of these two convoluted factors is almost not possible. This becomes acute in the regions of dominant human activities. For example, a dry year in an agriculturally dominant basin (e.g., the Ganges basin in India) will trigger more groundwater extraction leading to a more acute decrease in

TWSA than the corresponding decline in P itself.

Moreover, as we discuss above, the residual of precipitation, evapotranspiration, and runoff could be considered as changes in TWSA (TWSC) in terms of the water balance equation, instead of the TWSA itself. In other words, TWSA represents the current status of the land system over the long-term average, while TWSC denotes its change. Given the divergent meaning of the two variables, we conduct cross-comparisons between TWS-DSI and P-E-R, and analyze the mechanisms involved by comparing their differences. Lastly, we have explicitly discussed the inevitable uncertainties arising from various climate forcing, inadequate model physics, or the simplified assumptions of various processes.

References:

Roth, N., Jaramillo, F., Wang-Erlandsson, L., Zamora, D., Palomino-Ángel, S., Cousins, S. A. 2021. A call for consistency with the terms ‘wetter’and ‘drier’in climate change studies. Environmental Evidence, 10(1), 1-7.

Huntington TG. 2006. Evidence for intensification of the global water cycle: review and synthesis. J Hydrol. 319(1):83–95.

Dirmeyer PA, Yu L, Amini S, Crowell AD, Elders A, Wu J. 2016. Projections of the shifting envelope of water cycle variability. Clim Change. 136(3):587–600.

Labat D, Goddéris Y, Probst JL, Guyot JL. 2004. Evidence for global runoff increase related to climate warming. Adv Water Resour. 27(6):631–42.

Bosson E, Sabel U, Gustafsson LG, Sassner M, Destouni G. 2012. Influences of shifts in climate, landscape, and permafrost on terrestrial hydrology. J Geophys Res Atmos. 117(5):1–12.

Katul GG, Oren R, Manzoni S, Higgins C, Parlange MB. 2012. Evapotranspiration: a process driving mass transport and energy exchange in the soil-plant-atmosphere-climate system. Rev Geophys. 50(3):RG3002.

(5) Line 70: Is long-term P-ET approximately equal to the change in terrestrial water storage (TWS)? Why the authors do not use P-ET to construct an index and to perform the prediction of TWSA? Instead, this study uses partial outputs of soil moisture/snow data in the GCMs.

Response: As per the water balance equation, P-ET equals R+TWSC, i.e., summation of runoff and change in TWSA between two subsequent months (i.e., $TWSA_i$-$TWSA_{i-1}$). To be consistent in the variables (TWSA and TWSC) and to account for the water balance closures (as suggested by Reviewer#2), we used the metric P-E-R (=TWSC). However, as we discuss in the previous responses, they have different meanings for the terrestrial water cycle. Generally, it is difficult to detect significant trends in TWSC due to the slight interannual variability at the yearly time scale (Lv et al., 2021). Thus comparing their differences, as done in the revised manuscript, appears better than the individual investigation based on either of them.

Reference

Lv, M., Ma, Z., Yuan, N. 2021. Attributing terrestrial water storage variations across China to changes in groundwater and human water use. Journal of Hydrometeorology, 22, 3– 21. https://doi.org/10.1175/jhm-d-20-0095.1

(6) Table 1: What are the differences between GRACE reconstructions and GRACE

mascons solutions?

Response: The GRACE mascon solution is a type of GRACE-observed TWSA solution that has been widely used in the hydrology community (Scanlon et al., 2018). GRACE reconstruction is an ML-based TWSA product based on the GRACE observations and multi-source meteorological inputs, which was proposed to overcome the relatively short time span of the GRACE and GRACE-Follow On missions (~20 years). We have added more descriptions for the GRACE solutions and GRACE reconstructions in the revised manuscript as follows:

The GRACE (and GRACE Follow-On) missions have provided unprecedented estimates of monthly TWSA worldwide from April 2002 up to the present, however, with the 33 months missing because of the instrumental issues and mission interruption (Tapley et al., 2004). We use the GRACE mascon solution from the Center for Space Research at the University of Texas at Austin (UTCSR) to serve as the benchmarking product from the period 2002-2014 (Watkins et al., 2015). Compared to conventional GRACE products (e.g., 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). However, the GRACE observational products were not adequate to assess the long-term trends of TWSA due to relatively short temporal coverage (~20 years). Therefore, we obtain the GRACE reconstruction provided by Li et al. (2021b) for evaluation of the DDWW paradigm, which is generated using state-of-the-art machine learning and statistical methods and is also trained by the consistent GRACE mascon product from the UTCSR institution.

Reference:

Scanlon, B.R., Zhang, Z., Save, H., Wiese, D.N., Landerer, F.W., Long, D. Longuevergne, L., Chen. J. 2016. Global evaluation of new GRACE mascon products for hydrologic applications. Water Resour. Res., 52 (12), pp. 9412-9429

(7) Figure 2 makes no sense and it is hard to interpret. Instead, the manuscript can present dryness/wetness changes in some key regions here.

Response: We have removed Figure 2 in the revised manuscript and have restrained our analysis on the global land only. Since we have provided the processed data publicly available, further studies may focus on different regions (e.g., SREX regions, basin scales) as per the question in focus. Moreover, we have provided the temporal changes over the Qinghai-Tibetan Plateau, and the global land in the revised version to better establish the mechanism analysis as suggested by the reviewer.

(8) Line 273-276: Why did the authors use AI derived from CRU data to define wet and dry zones rather than TWS-DSI? The following DDWW analyses are based on the changes of TWS-DSI.

Response: Because the TWS-DSI, as a normalised drought index, is zero when looking at the long-term average and therefore not appropriate for the static classification for the climate zones, though we define the increase/decrease in TWS-DSI as wetting/drying signals. TWSA broadly shares a similar variability since it represents the anomaly over the long-term baseline. Therefore, in this case, we use commonly used AI to classify regions as arid, humid, or transitional following previous studies encountering similar issues (e.g., Feng and Zhang, 2015, Hu et al., 2019, and Yang et al., 2019). We have clarified the reasons for using AI for climate classification in the methods section. In addition, we also compare our AI-based results with the widely used Köppen-Geiger climate classification maps in the revised manuscript as follows:

To evaluate the DDWW paradigm over global land, the effective Aridity index (AI) is used to classify a grid cell as an arid, humid, and transitional region following Yang et al. (2019) because TWS-DSI/TWSA approximates zero for the long-term mean. The AI is calculated as the ratio of annual precipitation to potential evapotranspiration provided by the CRU TS-v4.06 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 S3, which is also highly consistent with the widely used Köppen-Geiger climate classification maps (Beck et al. 2018) (Figure S2). It can be seen that most of the arid regions (AI<0.5) are located in southwestern America, north and south Africa, central Asia, Arabian regions, and Australia, accounting for 39.3% of the land. The percentage of humid areas (AI>0.65) that are mainly located in east America, the Amazon region, central Africa, south China, west Europe, and Russia reaches 52.8% of the land. An approximate 7.9% of the land area is defined as the transitional region, referring to an intermediate between arid and humid climates. The transitional region generally lies in the shared boundaries of the humid and arid regions (e.g., western America, northern Canada, central Asia, western Africa, East Russia, and Australia). The DDWW paradigm is evaluated at a 5% significance level in this study, combined with the standard AI-derived climate classifications. We calculate the global mean trends of TWS-DSI using a spatially weighted method to account for the changing area of grid cells with latitudes.

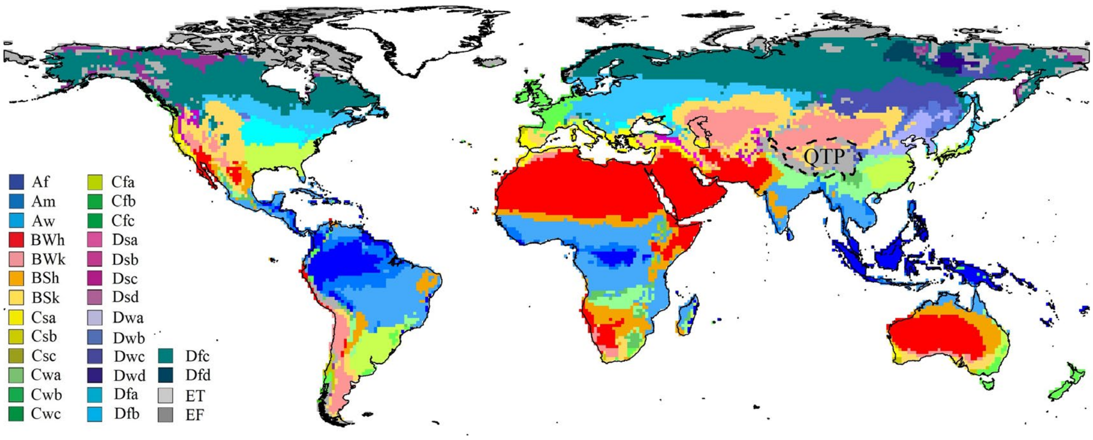

Figure S2. Global distribution of the improved Köppen-Geiger classifications during the period 1980-2016. Note: Please refer to Beck et al. (2018) for the details of the classification criteria. The dashed boundary represents the Qinghai-Tibetan Plateau.

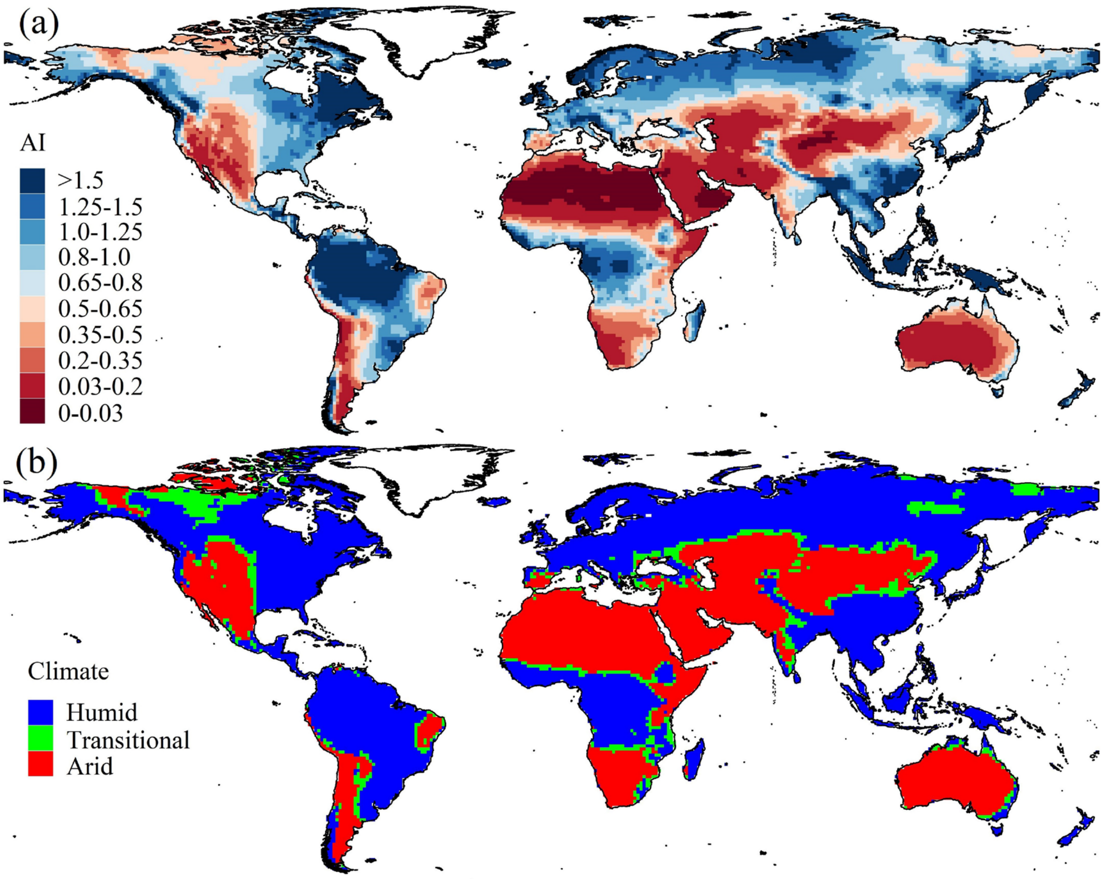

Figure S3. Global distribution of the (a) multi-year average aridity index (AI) and (b) climate type during the period 1985-2014. Note: The regions where AI>0.65 and <0.50 are defined as humid and arid regions, respectively.

References:

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F. 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci. Data, 5, https://doi.org/10.1038/sdata.2018.214

Feng, H., Zhang, M., 2015. Global land moisture trends: drier in dry and wetter in wet over land. Sci. Rep. 5, 18018. https://doi.org/10.1038/srep18018

Hu, Z.Y., Chen, X., Chen, D.L. Li, J.F., Wang, S., Zhou, Q., Yin, G., Guo, M. 2019. "Dry gets drier, wet gets wetter": a case study over the arid regions of central Asia. Int J Climatol 39(2):1072–1091

Yang, T., Ding, J., Liu, D., Wang, X., Wang, T., 2019. Combined Use of Multiple Drought Indices for Global Assessment of Dry Gets Drier and Wet Gets Wetter Paradigm. J. Clim. 32, 737–748. https://doi.org/10.1175/JCLI-D-18-0261.1

(9) Line 281: "We compare AI and TWSA derived from DATASET and CMIP6 between 1985 and 2014 in Figure S6". Is it a result comparison between Figure S5 and Figure S6 here?

Response: Yes, it is a result comparison between Figures S5 and S6. However, we have removed this because we do not simply take the ensemble mean of the DATASET as suggested by the reviewer. Moreover, we have added a comparison between AI and the Köppen-Geiger climate classification maps in the new version, as discussed above.

(10) Line 273-283: These contents about how to operate should be adjusted to the method section?

Response: As suggested, we have moved these contents to the method section in the revised version.

(11) Line 284: "Figure 3 illustrates the test results...". This is not a good way to express the content of figures.

Response: We have revised this statement in the new version as follows:
Combined with the climate regions classified by AI, we further test the DDWW paradigm at a 5% significance level using both TWS-DSI and P-E-R over global land in the past and future (Figures 3 and 4).

(12) Line 324-325: "Greve and Senevirtne (2015) used climate projections from CMIP5 to establish the measure for assessment of the DDWW paradigm...". The method used by Greve (2015) is more acceptable to peers.

Response: As suggested, we have established the new metric P-E-R for comparison with the TWS-DSI in the revised manuscript and have thoroughly discussed the observed similarities and differences.

(13) The presentation of Figure 4 makes no sense and it is hard to interpret. It may be more interesting to modify it to temporal changes over key areas.

Response: We have removed Figure 4 in the revised manuscript and have restrained our analysis on the global land only. Since we have provided the processed data publicly available, further studies may focus on different regions (e.g., SREX regions, basin scales). Moreover, we have provided the temporal changes over the selected key area of Qinghai-Tibetan-Plateau and the global land in the revised version as suggested by the reviewer.

(14) Line 355-358: "...resulting in the lack of certain TWSA components.". Why the authors do not use P-ET derived from the models to represent TWSA? In this case, none of the flaws discussed here exist.

Response: As discussed above, we have additionally established the metric P-E-R for comparison with TWS-DSI. However, it is noteworthy that the counterpart of precipitation, evapotranspiration, and runoff is changes in TWSA (i.e., TWSC) rather than TWSA in terms of the water balance equation. They reflect different aspects of the water cycle and can help to reveal the mechanisms involved by highlighting the difference between these metrics. To this end, we reserve the evaluation based on TWS-DSI and link these differences with some suggestions for future research in the discussion section of the revised manuscript.

(15) Line 417-419: Please explain what are the advantages of the developed TWS-DSI?

Response: We have added the advantages of the developed TWS-DSI in the revised manuscript as follows:

The non-dimensional TWS drought severity index (TWS-DSI) is established at both 1°×1° grid cell and regional/global scales, which is normalised by the regional hydroclimatological variability because a given magnitude of TWS deficit could indicate different dryness/wetness conditions in different climate regions. TWS-DSI has clear classification categories based on U.S. Drought Monitor (USDM) and is suitable for comparing dryness/wetness status for different locations and periods (Table S2). It has been widely used in hydrology and climate fields due to its simple structure and effective ability to capture drying and wetting conditions (Pokhrel et al., 2021).

(16) Line 447-462: The authors need to refine the conclusions. These do not look like conclusions, at least not serving for the purpose of this study.

Response: We have added some key points to the conclusion and refined the structure as suggested by the reviewer. Please also see our response to the major comment "Innovation and significance".

(17) The authors need to recheck and simplify the expression and logic of the entire manuscript, as many expressions seem redundant and use uncommon words, making it difficult to read.

Response: While revising the manuscript according to the reviewer's comments, we have thoroughly checked, simplified the expressions, and proofread the text to improve its readability.