

Reply to Reviewers' comments (Reviewer#3)

Legend

Reviewers' comments

Authors' responses

Direct quotes from the revised manuscript

We thank the reviewer for his/her time in reading our manuscript and for 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 #3: This is my second review of the manuscript, which has been resubmitted after the previous discussion round.

The authors present a re-examination of the dry gets drier, and wet gets wetter paradigm over global land, based on terrestrial water storage estimates from different sources. They make use of GRACE reconstructions, global hydrological models, and land surface models, as well as CMIP6 models for the future perspective. They conclude that the DDWW paradigm is challenged both in the historical period but also in the future.

Overall, the authors took into account my points previously made and the manuscript considerably improved compared to the initial submission. In particular, the authors corrected the calculation of the percentage values by area-weighting the grid boxes and added more discussion on the uncertainties inherent in their analysis.

As such, I'm happy with the changes made. However, there are still some open methodological points that need to be addressed (see specific comments). Also, in parts the manuscript might need to be checked for grammar and wording by a native speaker.

Response: We thank the reviewer for recognizing the potential of the manuscript's new perspective, highly encouraging feedback, and his/her detailed suggestions for improvement. All the concerns raised have been addressed in the revised manuscript. We hope the modified text, along with the supplementary analyses and discussions, will put forward the results in a much more robust way.

Specific comments:

(1) Line 81: I suggest changing the title to "Data pre-processing"

Response: Changed as suggested.

(2) Line 84: Change to “(see Table 1 and next sections)”

Response: Thank you. We have changed it as suggested.

(3) Line 86: Change to “resampled to 1° x 1° resolution to compare against the average”

Response: Changed as suggested.

(4) Line 89: Change to “As for DATASET, the members of the CMIP6 ensemble have been resampled to ...”

Response: We have merged this sentence with the previous line as follows:

The members of the CMIP6 ensemble and the historical datasets have been resampled to 1°×1° scale using a bilinear interpolation approach for consistency and better comparison in the spatial domain.

(5) Line 106: “We have implemented an ensemble ...” It’s not clear in this sentence that you take the mean of the reconstructions derived from the three forcing datasets. Please rephrase.

Response: We have removed this statement since only the CSR reconstructions from Li et al. (2021) are used for the evaluation of the DDWW paradigm in the new version owing to high reliability and robustness (produced using a combination of three data-driven approaches and compared against various hydroclimatic indices and water storage component outputs by global hydrological models) of this product compared to the other ones (e.g., JPL and GSFC reconstructions from Humphrey and Gudmundsson (2019)) (Li et 2020; 2021)

References:

- Humphrey, V., Gudmundsson, L., 2019. GRACE-REC: a reconstruction of climate-driven water storage changes over the last century. *Earth Syst. Sci. Data* 11, 1153–1170. <https://doi.org/10.5194/essd-11-1153-2019>
- Li, F., Kusche, J., Chao, N., Wang, Z., Loecher, A., 2021. Long-Term (1979-Present) Total Water Storage Anomalies Over the Global Land Derived by Reconstructing GRACE Data. *Geophys. Res. Lett.* 48, e2021GL093492. <https://doi.org/10.1029/2021GL093492>
- Li, F., Kusche, J., Rietbroek, R., Wang, Z., Forootan, E., Schulze, K., Lück, C. 2020. Comparison of Data-driven Techniques to Reconstruct (1992-2002) and Predict (2017-2018) GRACE-like Gridded Total Water Storage Changes using Climate Inputs. *Water Resources Research*, 56(5), e2019WR026551. <https://doi.org/10.1029/2019wr026551>.

(6) Line 107/108: Why only listing a subset of the variables used for the derivation of the CSR reconstruction and not all?

Response: We only listed a subset of input variables for CSR reconstruction since the other two reconstructions trained with JPL and GSFC are not used anymore, as we explained previously.

(7) Line 109: Change to “these three GRACE reconstructions”

Response: We changed it to “the GRACE reconstruction” since only one subset of GRACE reconstruction is used.

(8) Line 154: Change to “for which TWSA outputs are”

Response: Changed as suggested.

(9) Line 158: “and future periods (Krishnan and Bhaskaran, 2020)” Why citing this specific paper about wind speed in the Bay of Bengal? This does not appear to be the standard reference for CMIP6. Please use Eyring et al. 2016 instead.

Response: Thank you for the suggestion. We have updated the reference to Eyring et al. 2016.

(10) Line 162/163: “the sum of total soil moisture and snow water, which has been proven reliable to assess the TWS changes”. This is already mentioned a few lines above, please merge and rephrase.

Response: As suggested, we have merged this statement with previous lines in the new version.

(11) Line 192/193: Change to “... (NRMSE) between the mean GRACE TWSA and the ensemble means of DATASET and CMIP6 data after bias correction during the period April 2002-December 2014, with the NRMSE calculated as the ratio of RMSE ...”

Response: Thank you for the suggestion. We have changed it as suggested.

(12) Line 193: What is meant with "change range" of TWSA? The range of the TWSA values (i.e., max. minus min.) from DATASET and CMIP6 respectively? Please specify.

Response: Yes, it means the differences between the maximum and minimum TWSA. We have specified it in the revised version.

(13) Line 201: Change to “uncertainties in the CMIP6 simulations that remain even after undergoing the bias correction”

Response: Thank you for the suggestion. We have changed it as suggested.

(14) Line 210: Change to “The GRACE TWSA ranges from roughly –20 to 20 mm and shows ...”

Response: Changed as suggested.

(15) Line 211: Change to “A similar temporal pattern is captured ...”

Response: Thank you for the suggestion. We have changed it as suggested.

(16) Line 214/215: Change to “Moreover, the fluctuation range of DATASET is generally greater than the CMIP6 range before 2010. After 2010, DATASET tends to underestimate TWSA compared to CMIP6 and GRACE, and shows an increase in range.”

Response: Since we have excluded the usage of the specific datasets (e.g., two other GRACE reconstructions from JPL and GSFC), the apparent underestimation of the datasets disappear now. So we have changed this sentence as follows:

The GRACE TWSA ranges from roughly -20 to 20 mm and shows obvious seasonal characteristics with relatively higher uncertainty in the dry season than that in the wet season. A similar temporal pattern is captured by various models, with the change spread covering the variations of GRACE data.

(17) Line 217/218: Please rephrase, simplify, or merge with the following sentence. You look at long-term trends here.

Response: We have merged this sentence with the following text to focus on the long-term trends here.

(18) Line 219: “the historical period 1985-2014” Please clarify that the trend estimate for the historical period is based on DATASET (both in the text and in the figure captions). It would be interesting to see how the CMIP6 historical trends compare with the ones from DATASET.

Response: We have clarified the data source for the trend estimate during the historical and future periods in both text and figure captions. Moreover, we have compared the CMIP6 historical trends with the DATASET in Figures R1 and R2. The CMIP6 GCMs reasonably capture the decreasing trends in North America, Northeast Asia, North Africa, and South Australia together with increasing trends over central Africa, eastern Russia, and central North America, which are consistent with the DATASET. However, it also presents differently increasing TWS-DSI in central Asia and eastern Europe as well as depletion of South Africa. The differences might be arising from the bias in the atmospheric forcings and different TWSA components within these models. Based on this newly added comparison among CMIP6 historical trends and the ones from different models/products, we have appended the discussion in the revised manuscript.

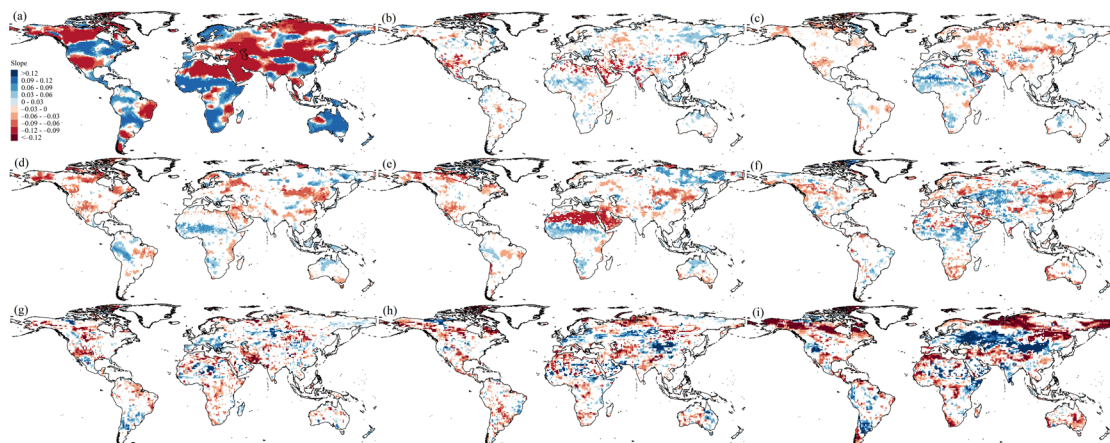


Figure R1 Global distribution of the significant ($p < 0.05$) long-term trends in TWS-DSI during (a-f) the historical (1985-2014) and future (2071-2100) period under (g) SSP126, (h) SSP245, and (i) SSP585 scenarios. Note: The historical results are based on the (a) GRACE reconstruction, (b) WGHM, (c) VIC, (d) CLSM, (e) Noah, and (f) ensemble mean of eight GCMs, respectively. The future results are based on the ensemble of eight GCMs.

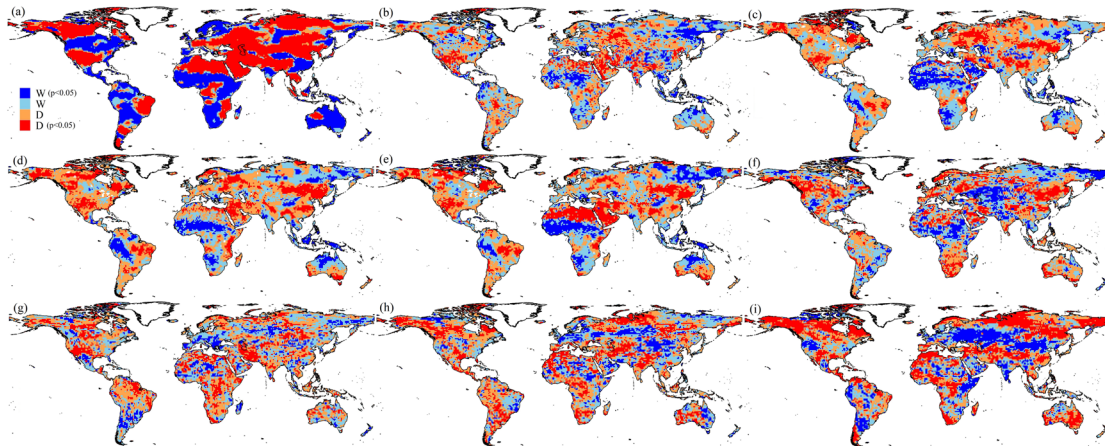


Figure R2 Global distribution of the classification in long-term trends in TWS-DSI during (a-f) the historical (1985-2014) and future (2071-2100) period under (g) SSP126, (h) SSP245, and (i) SSP585 scenarios. Note: The historical results are based on the (a) GRACE reconstruction, (b) WGHM, (c) VIC, (d) CLSM, (e) Noah, and (f) ensemble mean of eight GCMs, respectively. The future results are based on the ensemble of eight GCMs. “D” and “W” indicate regions with drying and wetting trends, respectively.

(19) Line 237: Change to “become wetter because”

Response: Changed as suggested.

(20) Line 261, Figure 1: Please clarify in the caption what data source (i.e., DATASET or CMIP6) is used for the individual temporal subset.

Response: We have clarified the data source for the trends estimations during individual temporal subsets (Figures R1 and R2 above).

(21) Line 269, Figure 2: Change to “the bar plot displays the global percentage.”

Response: We have removed this Figure according to the suggestions from Reviewers #1 and #2.

(22) Line 284: “Figure 3 illustrates the test results of DDWW paradigm at a 5% significance level ($p = 0.05$)” Based on what test? Does the mentioned significance level relate to the test results for long-term trends? Please clarify how you derive these results in the methods section.

Response: The significance of the trend estimates is based on the modified Mann-Kendall test at a 5% significance level to avoid autocorrelation (Hamed and Rao, 1998). As rightly recognized by the reviewer, the mentioned significance here for DDWW evaluation indicates the test results for long-term trends, which has been clarified in the methods section as follows:

We convert the monthly TWS-DSI into annual means to calculate the long-term trends using the linear regression method. We examine the first-order autocorrelation of each TWSA dataset using the Durbin-Watson test (Durbin and Watson, 1950, 1951) (Figure S1). We find a total of 20% (GRACE reconstruction), 43% (WGHM), 41% (VIC), 23% (CLSM), 29% (Noah), and 20% (GCM) of the grid cells not presenting autocorrelation during 1985-2014, respectively. For the future period, the percentage is 25%, 26%, and 22% under the SSP126, SSP245, and SSP585 scenarios, respectively. In this case, the significance of the long-term trends is evaluated using the modified Mann-Kendall trend test at a 5% level to avoid autocorrelation (Hamed and Rao, 1998).

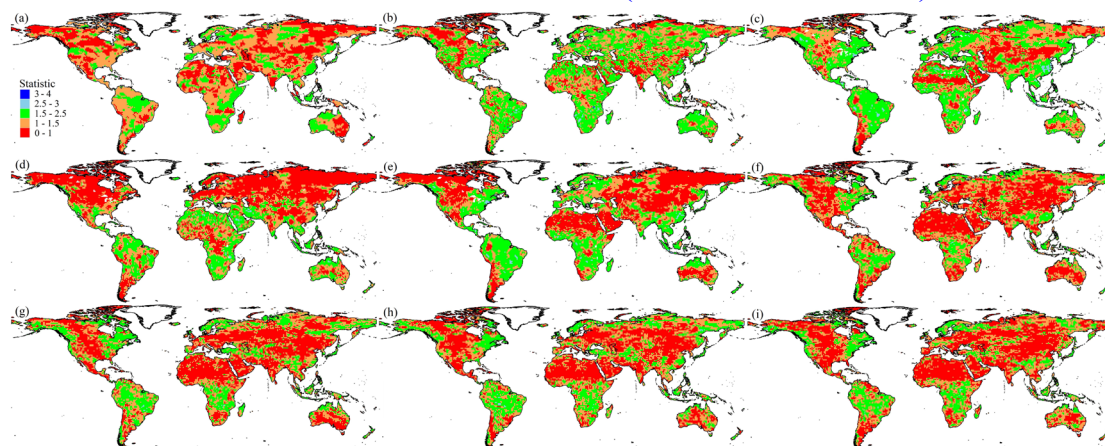


Figure S1 Global assessment of the autocorrelation during the (a-f) historical (1985-2014) and future (2071-2100) period under (b) SSP126, (c) SSP245, and (d) SSP585 scenarios. Note: The historical results are based on the (a) GRACE reconstruction, (b) WGHM, (c) VIC, (d) CLSM, (e) Noah, and (f) ensemble mean of eight GCMs, respectively using the Durbin-Watson test. The future results are based on the ensemble of eight GCMs. Generally, the residuals are considered not auto-correlated when the Durbin-Watson test statistic has a value between 1.5 and 2.5. If the statistic is below 1 or above 3, then there is definitely autocorrelation among the residuals.

Reference:

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

(23) Line 332, Figure 3: Please clarify the data source (i.e., DATASET or CMIP6) used for the different temporal subsets. It would be interesting to see how the results for DDWW based on the two data sources compare during the historical period. This could help to shed more light on the applicability of the CMIP6 ensemble for investigating the DDWW paradigm also in future periods.

Response: We have clarified the data source used for different temporal subsets in Figure R3. Moreover, the comparisons between the CMIP6 ensemble and different datasets during the historical period (1985-2014) are also performed. The CMIP6 historical results compare well with the multiple models. Specifically, the DD regions are mainly in North Africa, Northeast Asia, Arab region, and southwestern America. The WW area is generally located in the eastern Russia and coastal regions of West Africa. However, the CMIP6 data seems to fail to identify the WW pattern in the Spratly Islands of the South China Sea as well as Malaysia and Philippines. Moreover, the DD pattern is also slightly alleviated in central Asia. Overall, the CMIP6 ensembles show the potential to detect the large-scale changes in dryness/wetness globally, at least for

the long-term trends by comparing with different data sources (i.e., GRACE reconstruction, GHMs, and LSMs). Based on this newly added comparison among CMIP6 historical trends and the ones from the historical datasets, we have appended the discussion in the revised manuscript.

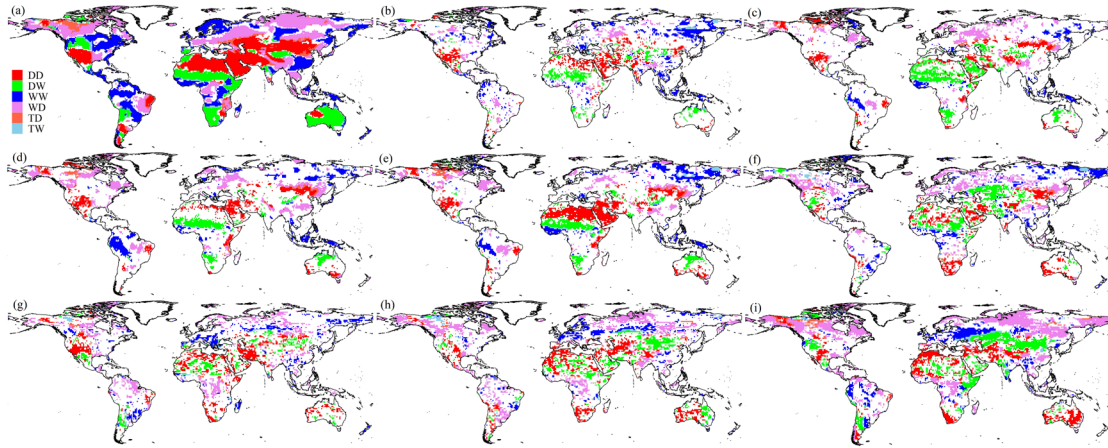


Figure R3 Global assessment of the DDWW paradigm during the (a) historical (1985-2014) and future (2071-2100) period under (b) SSP126, (c) SSP245, and (d) SSP585 scenarios. Note: The historical results are based on the (a) GRACE reconstruction, (b) WGHM, (c) VIC, (d) CLSM, (e) Noah, and (f) ensemble mean of eight GCMs, respectively. The future results are based on the ensemble of eight GCMs. 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.

(24) Line 340, Figure 4: “D” and “W” indicate regions with drying and wetting trends, respectively.” I guess this does not belong to this figure caption?

Response: We have removed this Figure according to the suggestions from Reviewer #1 and #2.

(25) Line 364: “reported to show underestimation or overestimation” -> variable-specific biases?

Response: Thank you for the suggestion. We have changed it as suggested.

(26) Line 421/422: “Despite the magnitude bias from satellite products, simulations of LSMs and GHMs, and GCMs projections, ...” Not sure what is meant here? Please rephrase.

Response: We have rephrased and simplified this statement given that we opt to present the DDWW evaluation results of each subset of datasets instead of only the ensemble mean.

(27) Line 453: “significance levels from 0.01 to 0.1” For the test on long-term trends?

Response: Yes, it indicates the test on long-term trends. We have revised this sentence in the new version for clarity.