Reply to Reviewers' comments (Reviewer#2)

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.

Reviewer #2: 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. I think that the topic is interesting and fits well the journal, and the use of the complete terrestrial water storage for the analysis of the DDWW paradigm adds another perspective compared to previous studies. However, the paper currently lacks some important information and has a substantial methodological issue which requires major revision.

Response: We thank the reviewer for recognizing the potential of the manuscript's new perspective 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.

Major comments:

(1) The use of percentage of grid cells for the presentation of many of the results is not appropriate and hinders the proper interpretation. It's necessary to present the corresponding numbers as percentage per land area (i.e., by weighing the grid boxes according to their effective km^2 area) in order not to give excessive weight to higher latitudes. This will most likely have impacts on the overall conclusions of the paper.

Response: Thank you very much for pointing out this mistake. As suggested, we have performed the re-calculation based on the actual area instead of the number of grid cells to evaluate the DDWW paradigm over global land (see Figure R1). The main findings have been summarized in the conclusion section of the revised manuscript.

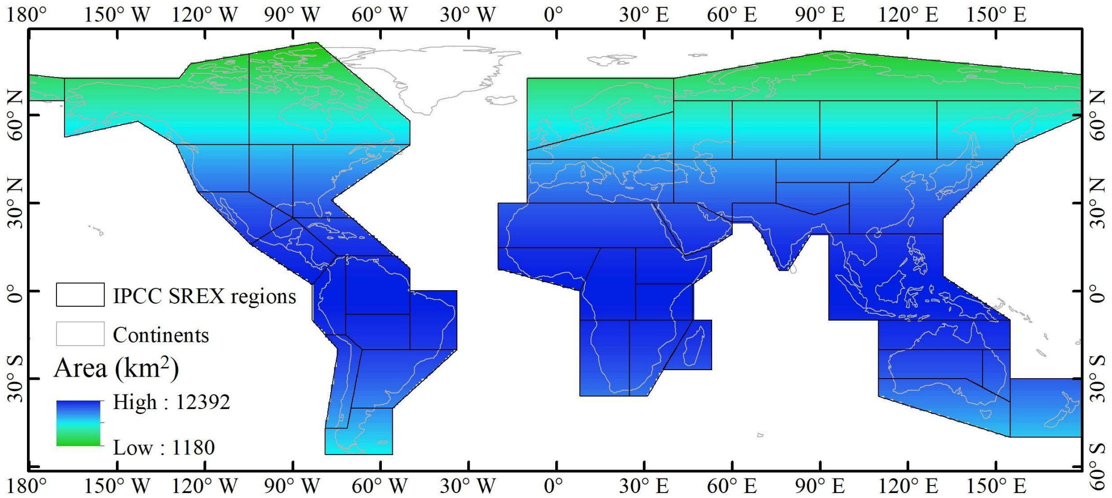

Figure R1. Spatial distribution of actual area of the 1° grid cells over global land.

The revised conclusions of the manuscript are as below (same as section 4 in the revised manuscript):

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 were summarized 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 with a significant slope increases from SSP126 (23.6%) to SSP585 (30.1%) scenario. 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 from 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 utilization of water resources under global change.

(2) The choice of the eight CMIP6 models is not transparent. Why didn't the authors choose a larger model ensemble based on the CMIP6 archive? Based on which criteria were these eight models selected?

Response: We thank you for the instructive comment. We have clarified the criteria to select these eight CMIP6 models in the Data section of the revised manuscript:

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 period under multiple emission scenarios (see Table S1). 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).

Reference:

Freedman, F.R., Pitts, K.L., Bridger, A.F.C., 2014. Evaluation of CMIP climate model hydrological output for the Mississippi River basin using GRACE satellite observations. J. Hydrol. 519, 3566–3577. https://doi.org/10.1016/j.jhydrol.2014.10.036.

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

(3) Also, given the large uncertainties between the CMIP6 models on the one hand, but also within

the DATASET ensemble (cf. Fig. S8), the impact of the applied ensemble mean approach on the results should be discussed in more detail.

Response: Thank you for the comment. We have discussed the results utilizing the individual model outputs and clarified the rationale for employing the ensemble mean approach. Moreover, we have added discussions on the uncertainty and implications of our study in the updated manuscript, as below (same as newly added section 3.3 of the revised manuscript).

Each ensemble member of the DATASET has embedded uncertainties inherently originating from one or more of forcing variables, simplified assumptions of complex processes in the models and their physical structure, retrieval algorithms, and systematic biases, which inevitably have propagated to the results presented herein. For example, the original GRACE mascon observations contain the measurement error and signal leakage at gridded scale, and they further spread into 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, the TWSA simulations from the GHMs, LSMs, and GCMs are featured by incomplete representation (Table S2). They are generally based on the simplified hydrological processes, resulting in the missing some of the TWSA components. For example, the widely used Noah model lacks the surface water and groundwater storage in TWSA, and all the GCMs can only simulate the snow water and soil moisture within a limited depth from 2 to 10 m below land surface (Xiong et al., 2021b; 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 showing underestimation or overestimation nearly over the global land (Eyring et al., 2016; Kim et al., 2020). Despite employing bias correction with GRACE data, propagated 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 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 the reasonably good accuracy with the NRMSE generally below 0.2, highlighting the reduced uncertainty compared with the individual solution. It is not surprising that 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 TWSA in the past (April 2002-December 2014) is conducted (see Figure S9). The spatial distributions clearly show that the ensemble-mean of eight GCMs outperforms each member globally, particularly in Australia, southern Africa, and North America. An overall decrease in NRMSE is observed according to the probability density functions, which is also detected from the Taylor diagram results (see Figure S10).

Further, we carried out an independent analysis at the individual member level (see Figure S11). For the historical period, 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 modelled 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, 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 from 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 S11). 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.

Despite the multisource uncertainties, our study can provide 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). The new insights benefit the comprehensive evaluation of terrestrial conditions over regions where some parts of TWSA (e.g., groundwater storage and snow water) have been rapidly depleting due to intensive human activities and warming climate worldwide, including the Qinghai-Tibet Plateau and northwest India (Rodell et al., 2009; Xing et al., 2021). Furthermore, the projected changes in global TWSA and associated TWS-DSI improve our understanding of the large-scale hydrological response under 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 can supply valuable inferences for policymakers and stakeholders for better water resources management in a changing environment (Iturbide et al., 2020).

Specific comments:

(1) Line 20: "and freshwater availability" instead of "and fresh availability"

Response: We regret the error. We have revised the sentence as follows:

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 (Shugar et al., 2020; Gampe et al., 2021).

(2) Line 83/84: Which of the available meteorological forcings for these reconstructions did you consider (MSWEP, GSWP3 or ERA5)? Did you take the mean over the three forcings for each of the two calibrations?

Response: Yes, we have opted for the ensemble mean of these reconstructions to avoid implicit biases to a single meteorological forcing an explanation of the GRACE reconstructions used in our study has been added as follows:

We note that the ensemble-mean of the NASA JPL and GSFC reconstructions forced with 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) datasets have been taken, respectively. The CSR reconstruction is derived from four kinds of meteorological variables (e.g., precipitation and 2 m temperature) and three kinds of hydrological variables (e.g., soil moisture and runoff) (Li et al., 2021).

(3) Line 125: Based on which criteria did you select these 8 models?

Response: Thank you for the suggestion. The selection criterion is constrained by the availability of the data. Please also see our response to Major comment 2 above for details.

(4) Line 136: "to match the observed data" instead of "to match the observed results"

Response: As suggested, we have revised the sentence.

(5) Line 144 (Equation 1): The dash in TWS – DSI could be confused with a "minus". I suggest changing it to an "_" or at least a short dash.

Response: Thank you for the suggestion. We have revised the representation of TWS-DSI in Equation 1 of the manuscript using a short dash.

(6) Line 152: "all the land area except for the Greenland and Antarctica"

Response: Following your constructive comment, we have revised this sentence as follows:
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 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 below).

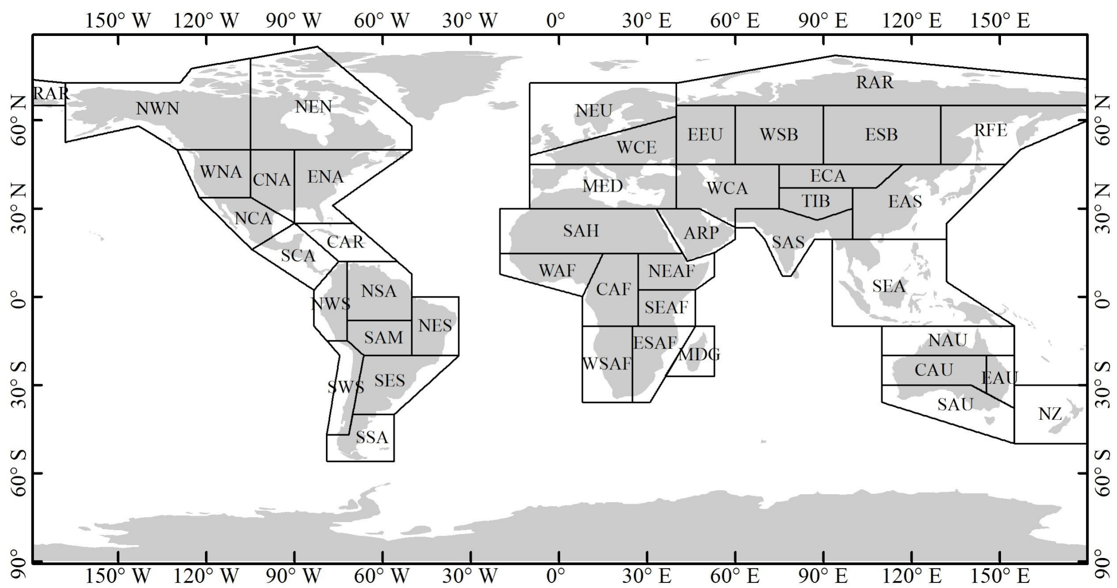

Figure S1. Location of the 43 selected Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Adaptation (SREX) regions from the Intergovernmental

Panel on Climate Change (IPCC) Sixth Assessment Report (AR6). The regional abbreviations are listed in Table S3.

(7) Line 157: “CMIP6 archive” instead of “CMIP6 achieve”

Response: We have rectified this typo.

(8) Line 168-170: Why poorer performance when NRMSE is lower?

Response: Thank you so much for the comment. It is an inadvertent error. We have revised the sentence as follows:

Most 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.

(9) Line 180: “greater” instead of “slighter”? The fluctuations of CMIP6 are larger than the ones of DATASET (Fig. S4).

Response: Considering the fluctuations of CMIP6 are larger than the ones of DATASET, we have revised this statement according to your suggestion as follows:

Moreover, the fluctuation range of CMIP6 data is generally greater than the DATASET, highlighting the considerable uncertainty sourced from different forcing variables and model parameterizations.

(10) Line 180: “the effective bias correction performance” Why effective? CMIP6 deviates more from GRACE than DATASET.

Response: We thank you for the comment. This sentence has been modified in the updated version of the manuscript.

(11) Line 212: Here and throughout the results the use of percentage of grid cells is not appropriate and needs to be changed to percentage of land area for proper interpretation.

Response: As suggested, we have re-calculated all the results based on the actual area instead of the number of grid cells and updated the results through the manuscript. Subsequently, the relevant text has been modified throughout the manuscript to reflect the changes.

(12) Figure 2: Nice plot, however quite crowded. The region names are often barely readable. You could just refer to the Supplementary Figure 1 for the definition and naming of the regions. The same applies to Figure 4.

Response: Following your constructive suggestion, we have removed the region names in Figure 2 and Figure 4 of the original manuscript, as shown in Figure R2 and R3 below.

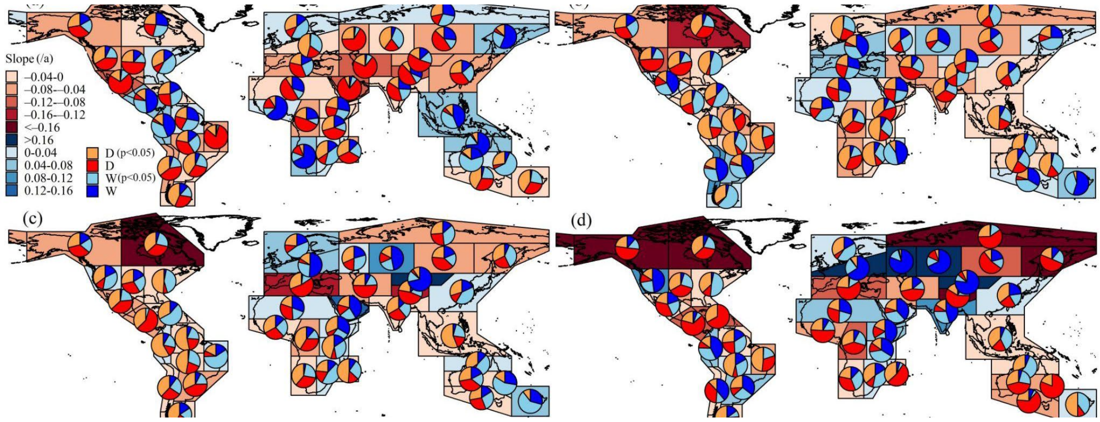

Figure R2 Global distribution of the long-term trends in TWS-DSI in 43 selected IPCC SREX regions during the (a) historical (1985-2014) and future (2071-2100) period under (b) SSP126, (c) SSP245, and (d) SSP585 scenarios. Note: The pie chart represents the regional proportion of area with different trends. "D" and "W" indicate regions with drying and wetting trends, respectively. Please refer to Figure S1 (added below) for abbreviations of the IPCC SREX regions.

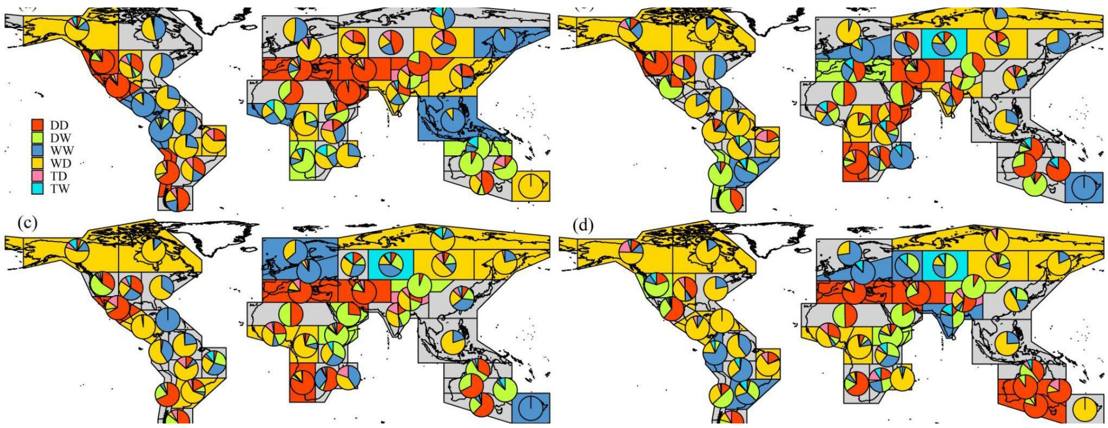

Figure R3 Global assessment of the DDWW paradigm in 43 selected IPCC SREX regions during the (a) historical (1985-2014) and future (2071-2100) period under (b) SSP126, (c) SSP245, and (d) SSP585 scenarios. Note: The light grey colour represents an insignificant pattern. The pie chart represents the regional proportion of area with different patterns to the total area with significant ($p<0.05$) patterns. "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 (added below) for abbreviations of the IPCC SREX regions.

(13) Line 234: There's no stippling on these figures? Please revise the caption and also explain the meaning of the pie charts.

Response: Thank you for this valuable comment. We have revised the caption and added an explanation of these figures as Figure R2 above.

(14) Line 235: Same title as for Section 3.1. I guess this is an oversight.

Response: We apologize for this oversight. We have replaced the title for section 3.2 with "Assessment of the DDWW Paradigm".

(15) Line 257 and following: I assume these percentages are again based on the grid cells only, not based on the actual area?

Response: Thank you again for the kind reminder. We have updated the manuscript using the newly calculated results based on the actual area instead of the number of grid cells.

(16) Conclusion: The conclusions need to be extended. What's new compared to previous studies? What are the implications?

Response: We thank you for the enlightening suggestion. The conclusions have been systematically extended and itemized for better comprehension in the new version of the manuscript. Please find the modified conclusions in Major comment 1.