HESS-2023-163

Supplement in Response to Referee #1

We added a sentence in the "Abstract" to explain the importance of incorporating groundwater simulation for basins with underlying aquifer (highlighted in yellow):

Results show that the simulated seasonal total water storage change (TWSC) does not compare well with GRACE even in basins within the same climate zone. The models overestimated the seasonal amplitude in most boreal basins and underestimated it in tropical, arid, and temperate zones. In basins with major underlying aquifers, the models that incorporate groundwater simulations provide a better representation of the water storage dynamics. In cold basins, the modeled phase of TWSC precedes that of GRACE by up to 2-3 months. However, it lags the GRACE phase by one month over temperate, arid to semi-arid basins. There was good agreement between GRACE and model amplitudes in the tropical zone. With the findings and analysis, we concluded that R2 models with optimized parametrizations have a better correlation with GRACE than the reverse scenario. This signifies an enhancement in the predictive capability of models regarding the variability of TWSC. The seasonal amplitude and phase-difference analysis in this study provide new insights into the future improvement of large-scale hydrological models and TWS investigations.

At the end of the "Introduction section" we added (highlighted in yellow):

In this study, we take advantage of Water Resource Reanalysis tier-1 (RI) and tier-2 (R2) products which provide a large set of LSMs and GHMs (Schellekens et al., 2017). We investigate the performance of 13 models in simulating the amplitude and phase at seasonal cycle relative to the latest release (RL06) of GRACE TWS for 29 major river basins under different climates.

Unique aspects of this study include:

1. Benchmark seasonal TWSC amplitudes and phase based on 13 GHMs and LSMS against GRACE.
2. Compare high-resolution forcing and more optimized structured R2 models against R1 models and access their ability to simulate TWSC variability and replicate water storage against the latest release (RL06) of GRACE TWSC.

At the end of section "3.1Comparison of seasonal amplitude between GRACE and models" we added (highlighted in yellow):

Over the four tropical basins (Fig. 4), all models underestimated the TWSC amplitude against GRACE. In the Mekong River basin, GRACE signals were very strong and the TWSC amplitude was at 230 mm while the GHMs and LSMs greatly underestimated it and TWSC amplitude ranged between ~134 mm to 191 mm below the GRACE. Over the Congo River basin, LSMs R1 and GHMs (R1 and R2) amplitude were at ~16 mm to 18 mm below the GRACE TWSC of 40 mm, though LSM R2 showed relatively better performance where TWSC amplitude was 6 mm below the GRACE. Surfex-Trip-R2 and PCR-GLOBWB -R1 were the relatively best-performing models over this basin where seasonal amplitudes were at 44 mm and 35 mm respectively. In the Orinoco basin, the GRACE amplitude was at 178 mm while all the models underestimated it by ~112 mm to 160 mm. In the Amazon basin, the GRACE amplitude was recorded at 178 mm while models greatly underestimated it by ~149 mm to 156 mm. Generally, over the Orinoco and Amazon River basins, PCR-GLOBWB-R1 gave better estimates of TWSC amplitude as compared to other models. Table 2 exhibits amplitude derived from GRACE, and mean amplitude derived from LSMs and GHMs.

To investigate the benefit of explicitly represent groundwater compartment in models, we identify five basins with major underlying aquifers, namely Congo, Amazon, Orinoco, Ganges-Brahmaputra, and California basins (https://www.un-igrac.org/resource/whymap-groundwater-resources-world-incl-thematic-maps). The models with groundwater simulations showcase more accurate representations of the seasonal water storage cycle (Fig. 5). This suggests that considering groundwater storage change and the interaction between surface water and groundwater leads to better alignment with GRACE measurements. In contrast, models without groundwater simulations often underestimate or even fail to capture the intricate dynamics of water storage changes (Fig. 5). Groundwater, being a vital component of the hydrological system in these basins, significantly impacts the seasonal variations in water storage. Models that do not account for groundwater dynamics missed essential aspects of the TWSC, resulting in significant underestimation of TWSC against GRACE. Ultimately, the inclusion of groundwater simulations in global hydrological and land surface models enhances their ability to accurately represent the TWSC in these critical river basins. This improved agreement with GRACE TWSC is essential for better monitoring and management of water resources, especially in regions where groundwater plays a significant role in sustaining the water cycle and socioeconomic wellbeing.

At the end of section "3.2 Phase difference between GRACE and models" we added (highlighted in yellow):

The seasonal cycles of the boreal basins show TWS peaks in spring, which are largely generated by snowmelt. In snow-dominated basins (Fig. 1) seasonal TWSC variations from models and GRACE exhibited consistency in the timing of crest except over the Saint Lawrence River basin where Surfex-Trip-R1and SWBM-R1peaks appeared one month earlier than GRACE, while troughs were inconsistent with GRACE TWSC over all the basins. The model TWS precedes GRACE by 3-4 months. The trough in GRACE for all the basins started in September (except for the Kolyma and Amur basins where they started in October) while in models trough began in June, giving models a 3-month lead. Similarly, over Yenisei and Amur basins (July) and Saint Lawrence (where most of the models showed ditch in May), models were 4 months ahead of GRACE observations.

There was no phase difference between modeled and GRACE TWSC in the temperate zone except for the Yellow River and Rio Grande River basin where GRACE peaks were ahead of modeled TWSC by one month (Fig. 2).

In arid basins modeled TWSC peaks have an identical phase with GRACE TWSC over Niger and Nile River basins. While over the Zambezi and São Francisco River basins modeled TWSC peaks appeared in April, resulting in a one-month time lag over these two basins compared to GRACE where peak storage was recorded in March (Fig. 3).

Models and GRACE TWSC phase were quite consistent with GRACE over the Orinoco and Amazon River basins in the tropical zone. However, the GRACE peak over the Congo River basin was observed earlier in April while modeled peaks were noted in May. Similarly, over the Mekong Rivers, GRACE observed peak water storage change was observed in September while the models' peak appeared in October (Fig. 4). Figure 6 shows time lag between GRACE and average lag derived from LSMs and GHMs.

In the "Conclusions" section, we added:

13 models (GHMs, LSMs) were evaluated using different resolutions of Water Resources Reanalysis (WRR1 and WRR2) to compare simulated Total Water Storage Change (TWSC) against GRACE observations over 29 major river basins. Model performance differs significantly across basins, even within the same climatic region. In snow-dominated basins, LSMs generally underestimate the TWSC amplitude and GHMs overestimate. Models and GRACE exhibited inconsistency in the phase with modeled TWSC preceding GRACE with 3-4 months lags.

In temperate, arid, and tropical basins GHMs and LSMs generally underestimate the amplitude. However, the modeled TWSC phase is identical to those of GRACE with few exceptions. Furthermore, in basins with major underlying aquifers like the Congo, Orinoco, and Ganges-Brahmaputra those models incorporating groundwater simulations show substantially better agreement with GRACE and provide a more accurate depiction of the seasonal TWSC compared to the models without groundwater simulations.

The following are newly added figures 5 and 6, and Table 2 in the manuscript:

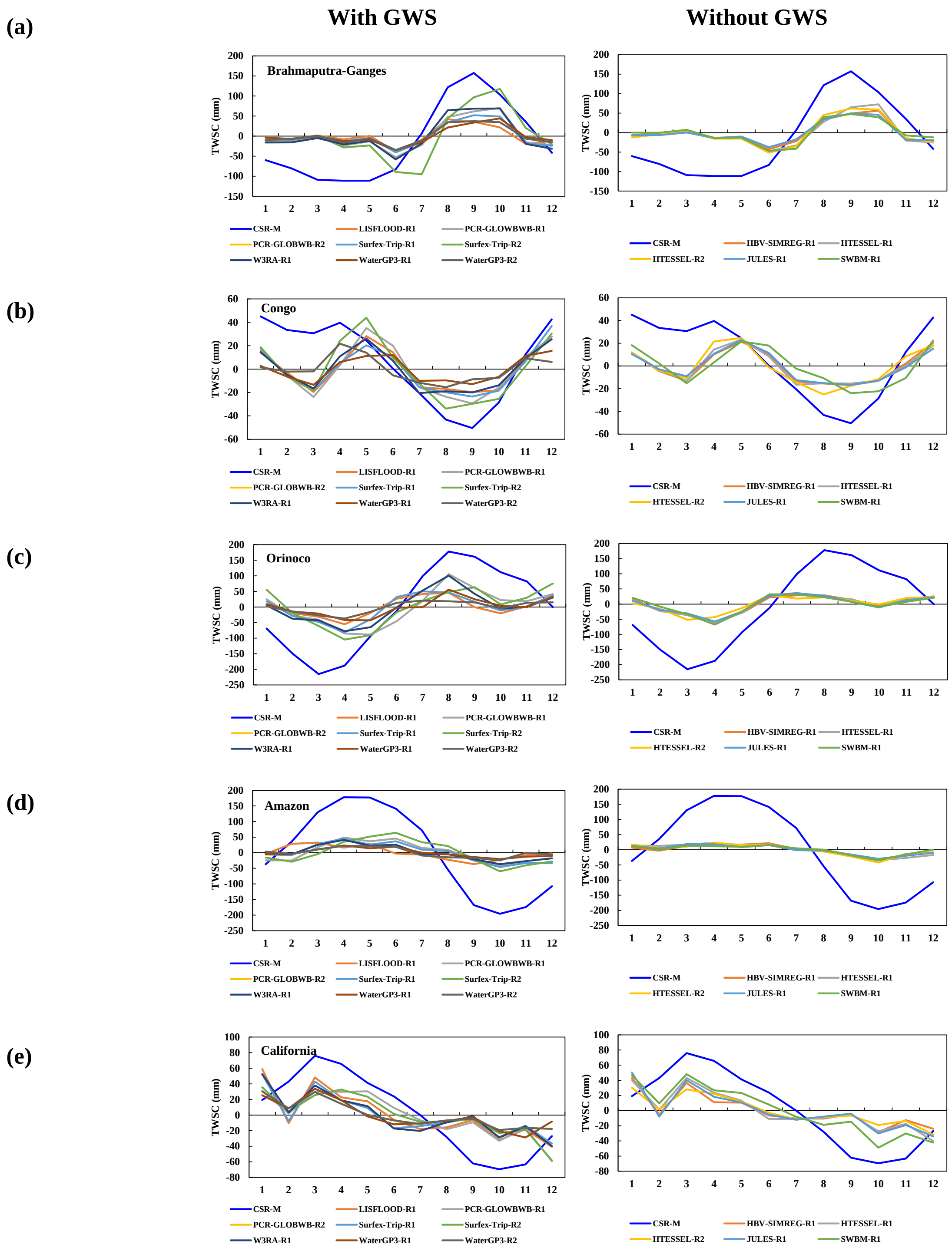

Figure 5: Comparison between seasonal TWSC from GRACE and models with and without groundwater simulations in (a) Ganges-Brahmaputra, (b) Congo, (c) Orinoco, (d) Amazon and (e) California basins.

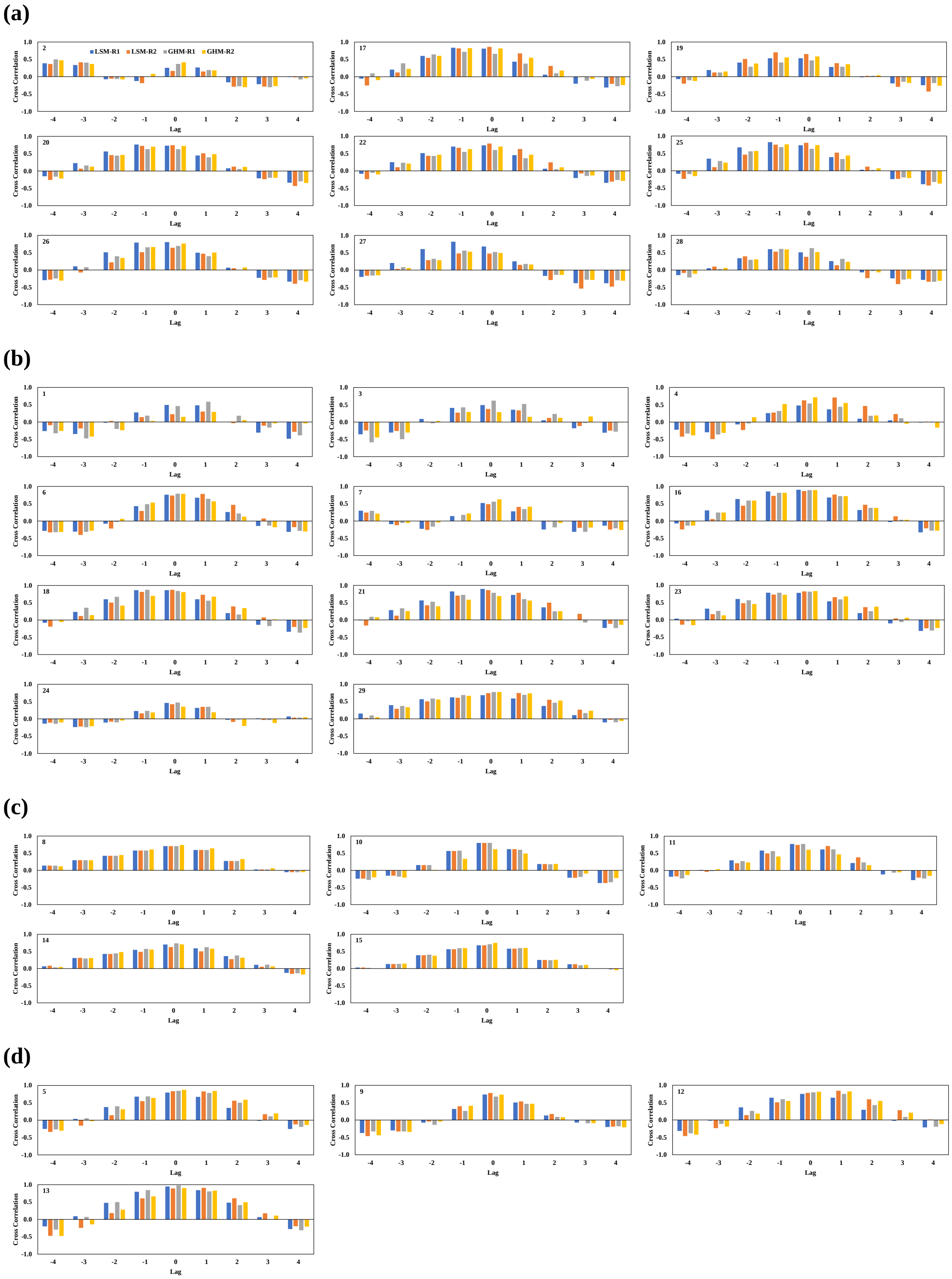

Figure 6: The values of correlation coefficients at different time lag (months) between GRACE and the LSM (blue and orange bars for R1 and R2) and GHM (grey and yellow bars for R1 and R2 respectively).

Table 2: Amplitude (mm) derived from GRACE, LSM_R1, LSM_R2, GHM_R1 and GHM_R2. ± is the standard deviation.

Climate Zone	Basin ID	CSR-M	LSM-R1	LSM-R2	GHM-R1	GHM-R2
Boreal	2	14	8.4±0.8	5±0.3	20±3.4	16±3.1
	17	30	26±0.8	18±0.3	54±3.5	28±3.1
	19	62	20±0.8	24±1.0	36±2.2	35±0.5
	20	47	25±0.8	17±1.0	36±2.2	29±0.5
	22	67	57±0.8	41±0.3	96±3.4	62±3.1
	25	59	33±0.8	21±0.3	54±3.4	36±3.1
	26	49	27±0.4	20±1.0	42±2.2	32±0.5
	27	30	20±0.4	45±1.0	54±2.2	40±0.5
	28	30	19±0.4	29±1.0	37±2.2	37±0.5
Temperate	1	6	12±1.1	19±2.7	9±1.5	9±0.6
	3	8	15±1.1	19±2.8	10±14.5	4.9±0.6
	4	42	16±1.2	19±2.2	16±3.9	14±0.6
	6	157	56±0.5	89±2.3	49±1.2	49±0.1
	7	36	29±1.0	40±1.1	31±4.0	31±5.2
	16	114	28±0.8	38±0.3	39±3.4	38±3.1
	18	63	11±1.0	12±0.5	15±4.0	9.3±5.0
	21	79	25±1.0	27±1.1	31±4.0	22±5.0
	23	53	12±0.4	15±1	16±2.2	19±0.5
	24	56	10±2.3	11±1.6	11±1.0	11±2.0
	29	76	41±3.1	27±1.7	40±8.3	30±0.3
Arid	8	187	55±0.8	72±0.3	52±3.4	48±3.1
	10	112	42±0.5	38±2.7	42±14	34±0.1
	11	59	27±0.5	29±2.3	30±1.2	25±0.1
	14	70	36±0.8	38±0.3	39±3.4	32±3
	15	65	18±1	17±1.1	18±3.9	19±5
Tropical	5	230	65±1.1	96±2.7	39±14	29±15
	9	40	22±0.4	34±1.0	24±2.2	25±0.5
	12	178	38±2.2	37±2.3	66±3.9	62±0.6
	13	178	26±0.8	40±0.5	28±3.4	36±3.0