Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Surface waters expose ambiguity in global satellite-based groundwater estimates

Augusto Getirana

augusto.getirana@nasa.gov

NASA Goddard Space Flight Center https://orcid.org/0000-0001-9635-7220

Hiroko Beaudoing

NASA Goddard Space Flight Center

Sujay kumar

NASA GSFC https://orcid.org/0000-0001-8797-9482

Matthew Rodell

NASA Goddard Space Flight Center https://orcid.org/0000-0003-0106-7437

Article

Keywords:

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11816360/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Surface waters expose ambiguity in global satellite-based groundwater estimates

Augusto Getirana*[1,2], Hiroko Kato Beaudoing[1,3], Sujay Kumar[1], Matthew Rodell[1]

[1]Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, United States

[2] Science Systems and Applications, Inc., Lanham, Maryland, USA

[3]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD

*Corresponding author (augusto.getirana@nasa.gov)

Abstract

Groundwater storage is commonly estimated by subtracting modeled soil moisture, snow and canopy water from terrestrial water storage (TWS) observed by the GRACE satellite missions, implicitly assuming that residuals represent groundwater. This assumption is problematic because land surface models (LSMs) partition water differently and often omit surface water storage (SWS). Here, we couple four global LSMs with a hydrodynamic model and quantify how model-consistent river and floodplain storage alters TWS during 2003–2014. The models generate markedly different runoff and SWS contributions, with global SWS impact indices ranging from 14% to 27% of modeled seasonal TWS variability. Nevertheless, adding each model's corresponding SWS generally improves agreement with GRACE and increases multi-model consistency. Degradation is concentrated in snow- and ice-dominated regions and strongly regulated basins, where cryospheric processes remain uncertain and reservoir operations are not represented. These results show that SWS is a necessary, model-dependent component of TWS and that residual differences between GRACE and model-consistent TWS (i.e., accounting for SWS) should not be interpreted as groundwater without consideration for various sources of uncertainty. These may include model structural limitations, forcing and parameterization errors, unresolved human influences and satellite uncertainty. This challenges the common practice of estimating groundwater variations as a GRACE-LSM residual.

Main

The Gravity Recovery and Climate Experiment (GRACE) and GRACE Follow-On missions measure monthly variations in terrestrial water storage (TWS), i.e., the vertically integrated mass of water in snow, vegetation, soils, aquifers and surface water bodies[1,2]. These observations have revealed widespread changes in freshwater availability associated with hydroclimatic variability, groundwater abstraction, cryospheric loss and surface water management[3]. However, satellite gravimetry cannot distinguish among the reservoirs producing these changes. Groundwater storage is therefore commonly inferred by subtracting modeled soil moisture, snow and canopy water from GRACE TWS, an approach that has enabled influential assessments of groundwater depletion in regions with limited well observations[4–6]. Such residual estimates are uniquely attributable to groundwater only if every other varying water reservoir is represented accurately – an assumption that cannot be verified from agreement with GRACE alone.

This limitation creates a fundamental identifiability problem. Land surface models (LSMs) differ substantially in their storage capacities, process formulations and explicit representation – or omission – of groundwater, and can, consequently, partition TWS differently even when their total storage variations are similar[7–10]. Any omitted storage, together with errors in modeled soil moisture, snow and other components, is transferred to the GRACE residual and may be misinterpreted as groundwater variability[11]. In particular, certain LSMs overestimate soil moisture variations at the expense of other TWS component variations, particularly those of groundwater and surface water, so that a LSM might produce TWS variations that compare well with GRACE TWS variations without accurately portraying the individual components. The practical consequences are substantial: an evaluation of 320 combinations of GRACE and modeled storage products across 37 major aquifer systems found that the fraction of physically plausible groundwater estimates varied from below 10% to above 60% among individual realizations[12]. Such screening can identify inconsistent residual estimates but cannot determine which omitted or misrepresented reservoir produced them. Hence, agreement in total TWS does not validate a model's internal partitioning, nor does it guarantee that the remaining signal represents groundwater[11].

Surface water storage is one potentially consequential source of this ambiguity. TWS estimates assembled from commonly archived land-model states generally exclude dynamically varying water stored in rivers and floodplains, although these reservoirs can strongly influence both the amplitude and timing of continental storage variations. Global routing experiments have shown that including river storage improves modeled TWS relative to GRACE across many large basins[13]. Independent satellite observations indicate that surface water storage contributes approximately half of the annual TWS variation in the Amazon and Ganges-Brahmaputra basins[14–16], while large-scale hydrodynamic simulations estimate that rivers and floodplains account for, approximately, 8% of TWS variability globally – lower than other water storage components –, with much larger contributions in many regions globally[17]. However, the latter analysis relied on a single groundwater-enabled LSM and could not determine whether the influence of surface water is robust across models with fundamentally different subsurface structures. Note that the definition of SWS impact on TWS varies across these studies.

An earlier multi-model analysis of the Amazon showed that LSMs with contrasting storage formulations, including models without an explicitly diagnosed groundwater reservoir, could reproduce much of the GRACE-observed storage variability when storage change was calculated from precipitation, evapotranspiration and routed river discharge[18]. The analysis indicated that

realistic aggregate TWS variability could emerge without a unique representation of groundwater. However, because storage change was inferred from fluxes rather than evaluated from the modeled storage states, it could not isolate the contribution of surface water, determine how storage was partitioned internally, or establish whether this behavior extends globally.

Here, we test the global generality and physical basis of this result using storage states from four land surface models (LSMs) in the Global Land Data Assimilation System Version 2 (GLDAS-2) [19], including models with and without explicit groundwater reservoirs. Runoff from each model independently drives the HyMAP-3 global hydrodynamic model[20,21], providing corresponding estimates of river and floodplain storage that are added to the land storage components in a controlled experiment. We find that including surface water storage generally improves the correlation and seasonal amplitude of modeled TWS relative to GRACE and reduces global multi-model spread. Importantly, all models, with and without explicit groundwater storage, also reproduce much of the GRACE-observed TWS variability across most non-snow-dominated regions once model-consistent SWS is included, despite their fundamentally different subsurface storage formulations.

Controls of surface water storage impacts on TWS

Multi-model differences in LSM-derived runoff fields are a first-order control on the contribution of surface water storage to modeled terrestrial water storage, but not the sole control. The four LSMs differ substantially in their representation of infiltration, soil-water dynamics, evapotranspiration, snow processes, runoff generation, and subsurface drainage[9,18,22]. These structural differences translate into a wide range of global runoff rates, spanning more than a factor of two across the four simulations, from 208 $mm \cdot yr^{-1}$ in CLSM to 449 $mm \cdot yr^{-1}$ in VIC, with even wider ranges regionally (Fig. 1). The different subsurface-storage formulations in Noah-MP and CLSM can also affect both the magnitude and timing of runoff supplied to the river system[10,23,24].

Yet, the resulting SWS impacts on TWS do not follow the same ordering as the annual runoff totals. Both CLSM and Noah, despite having lower runoff than Noah-MP, exhibit larger SWS impacts (16% and 21%, respectively; Noah-MP = 14%). VIC shows both the highest runoff and the largest SWS impact (27%). This non-monotonic behavior indicates that the effect of SWS on TWS is not determined solely by how much runoff is generated, but also by when runoff occurs and how its seasonal distribution shapes the timing and amplitude of routed surface water storage variability compared to other storage components. The larger relative contribution of subsurface storage (i.e., soil moisture and explicit representation of groundwater) in Noah-MP and CLSM may decrease SWS impact.

SWS impact reflects the interaction among runoff magnitude, runoff seasonality, the resulting variability of routed surface water, and the variability of the other storage components that compose TWS. The comparatively stronger SWS impacts found in Noah and VIC, which do not include explicit groundwater storage, are not necessarily evidence that surface water compensates for missing groundwater variability. Rather, they may reflect greater SWS variability, or lower baseline land water storage variability (see Supplementary Figs. S1 and S2 for impact index maps for other water storage components). It is worth noting that, in dry areas, such as the Sahara, components may be small, resulting in noisy ratios.

Surface water storage improves global multi-model consistency

At each grid cell, correlation, standard deviation ratio, 2003–2014 linear trend error, and RMSE were calculated independently for each LSM before and after adding its corresponding HyMAP-3 SWS, and the model-specific changes were summarized across the four LSMs (Fig. 2a–d). Accounting for SWS generally improves agreement with GRACE. Median correlation increases by ~0.01 globally and by 0.01 when regions north of 50°N are excluded, with improvements reaching 0.31 in the Amazon Basin. Degradation is concentrated across Arctic and sub-Arctic regions and parts of the Himalayas. Localized degradations are observed near the Mississippi River outlet, lower Paraná Basin, and Nile, and are consistent with unmodeled processes, such as sea level rise, land subsidence and land loss, as widely reported in the Mississippi outlet[25–27], and river regulation (e.g., Itaipu and Aswan dams, in Paraná and Nile Rivers, respectively). The strongest decrease, -0.5, occurs near the Suez region, in Egypt. Consequently, widespread high-latitude degradation partially offsets predominantly positive changes at lower latitudes.

The representation of TWS variability generally improves with the addition of SWS. Using the log-symmetric standard deviation ratio improvement index, which treats proportional underestimation and overestimation of variability equally, the area-weighted median improvement is 0.082 globally and increases to 0.165 south of 50°N (Fig. 2b); both values are statistically different from zero ($p<0.001$). Degradations are again concentrated at high northern latitudes and the Himalayas. An additional east–west degradation band crosses the African Sahel. These global patterns are consistent with known sensitivity of simulated runoff and surface water dynamics to regional land surface processes in high latitudes[28–32], Asian mountains[33–35] and the Sahel[36–41].

The effect of SWS on agreement in 2003–2014 linear trends is substantially more spatially heterogeneous (Fig. 2c). The bounded trend-error improvement index has an area-weighted global median of −0.002, which is not statistically different from zero ($p=0.743$), indicating no systematic global improvement or degradation in trend agreement. South of 50°N, the median is slightly positive (0.002; $p=0.036$), but its small magnitude indicates only a weak broad-scale tendency toward improved trend agreement. Larger improvements and degradations occur regionally, reinforcing that the influence of SWS on study-period trends is predominantly spatially dependent rather than globally systematic. Regional study-period trends provide no evidence that routed SWS systematically compensates for the absence of an explicit groundwater reservoir (Supplementary Text S2, Fig. S3 and Table S1).

RMSE changes largely reproduce the spatial structures identified by the correlation and amplitude diagnostics (Fig. 2d). The largest improvement reaches -257 mm in the Amazon Basin, whereas the strongest degradation reaches 108 mm at the Mississippi River delta. High-latitude degradation offsets widespread improvements elsewhere, while excluding regions north of 50°N yields a median RMSE improvement of, approximately, -0.05 mm. Assuming 20mm as a practical sensitivity threshold[8,42–45], 93.2% of the global analyzed area and 91.2% of the area south of 50°N showed RMSE changes within ±20 mm. Among grid cells exceeding this threshold, improvements were substantially more frequent than degradations, occurring over 5.2% versus 1.6% of the global area and 6.8% versus 2.0% south of 50°N. The recurrence of extremes around the Mississippi River delta, Itaipu, Niger Inner Delta, Nile, and Himalayas across multiple metrics indicates coherent errors in the timing or amplitude of simulated storage rather than sensitivity to an individual performance measure.

SWS also reduces multi-model spread-to-signal ratio independently of agreement with GRACE (Fig. 2e). The median change in the multi-model spread-to-signal ratio is approximately -0.06 globally, and -0.09 south of 50°N indicating greater convergence after SWS is included.

Convergence is widespread south of 50°N, with divergence pockets in parts of Paraná, Niger, Nile, Volga and Australia. The recurrence of several of these features in both the GRACE-based metrics and the multi-model spread diagnostic indicates that their origin likely involves differences in the simulated SWS itself rather than solely discrepancies between modeled TWS and GRACE. Overall, the inclusion of surface water storage improves both the observational and multi-model consistency of modeled TWS across most of the global domain. Most importantly, the overall spread convergence across the globe demonstrates that SWS reconciles part of the storage differences among models with contrasting subsurface structures.

Basin-scale surface water accounting across contrasting climate and human influences

The basin-scale analysis reinforces the predominantly positive influence of SWS while identifying where degradation is concentrated. Across the 18 basins and four LSMs (see basins in Fig. 2a), correlation with GRACE improves in 39 model–basin combinations and degrades in 33; excluding Arctic and sub-Arctic basins changes this balance to 32 improvements and only 16 degradations. Similar behavior occurs for the standard-deviation ratio and normalized RMSE (nRMSE), indicating that much of the apparent basin-scale degradation originates at high northern latitudes (Fig. 3).

The Mackenzie, Yenisei, Lena, Ob, Amur, and Volga exhibit predominantly degraded performance after SWS is included, consistent with the global high-latitude patterns; the Volga is a partial exception. These results are examined further through their seasonal cycles in the next section. Outside high northern latitudes, degradation is more localized, particularly in the Yangtze, Niger, Nile, and Paraná. The Nile and Paraná are classified as very highly and highly regulated, respectively[46], and reservoir operations are not represented in these HyMAP-3 simulations.

In contrast, the comparatively less regulated Amazon, Congo, and Mekong exhibit substantial improvements across correlation, variability, and nRMSE for most or all LSMs. Their large rivers, floodplains, and strong seasonal SWS variability make omission of surface water particularly consequential. Where runoff and surface water dynamics are reasonably represented and anthropogenic storage changes are more limited, including SWS brings modeled TWS closer to GRACE across independent metrics, regardless of whether the underlying LSM explicitly represents groundwater storage.

Surface water storage reconciles seasonal TWS across models

Seasonal TWS variability is comparatively consistent among the four modeling systems in predominantly rainfall-driven basins such as the Amazon, Congo, Niger, and Mekong, where the climatologies exhibit limited multi-model spread and generally reproduce the timing of GRACE (Fig. 4). In contrast, the Mackenzie, Yenisei, Lena, Ob, and Amur display pronounced differences in both phase and amplitude and comparatively large normalized mean absolute error (nMAE). Because all four simulations use the same meteorological forcing, these differences primarily reflect model-dependent representation of the processes controlling seasonal water storage and runoff.

Large multi-model spread also occurs in the Indus and Yangtze, where high-elevation snow, ice, and frozen-ground processes exert important controls on seasonal hydrology. Differences in the timing of snow accumulation and melt propagate through runoff and routing into phase differences in SWS and TWS, whereas such phase offsets are substantially smaller in tropical basins. The

seasonal climatologies reinforce the association between large multi-model divergence and regions where cryospheric processes are represented differently among LSMs[29,31,35].

Model–GRACE discrepancies provide complementary information. Relatively high nMAE also occurs in the Paraná, Nile, Tigris–Euphrates, Yangtze, and Indus, several of which are strongly affected by reservoir regulation that is absent from the simulations. Reservoir operation can explain disagreement with GRACE but cannot directly explain differences among the four modeled climatologies because it is absent from all simulations. Interpretation of the remaining discrepancies depends on model structure. Noah-MP and CLSM explicitly partition part of subsurface storage into groundwater, whereas Noah and VIC do not. In the latter models, water that might otherwise be represented as groundwater is expressed through the soil-water states and associated fluxes permitted by their formulations. The ability of Noah and VIC to reproduce much of the observed TWS seasonality in many basins, once model-consistent SWS is added, should not be interpreted as evidence that groundwater variability is negligible, but rather that similar aggregate TWS variability can arise from different internal storage partitions. Agreement with GRACE at the total-storage level alone cannot determine which partition is correct.

Discussion

Accounting for the surface water storage generated from each LSM's own runoff generally moves modeled TWS closer to GRACE while increasing multi-model consistency, despite substantial differences in runoff generation and storage partitioning among the four models. Regional degradation is concentrated where identifiable limitations in cryospheric, floodplain, or regulated-river processes occur. Where SWS varies significantly, omitting it leaves the modeled TWS incomplete. SWS is an intrinsic component of TWS, and deterioration after its inclusion may be an indicator of limitations in model or observational representations rather than a reason to omit surface water. Groundwater and surface water are hydrologically connected and have long been considered as a single resource[47]. Indeed, this interpretation was proposed by Rodell et al.[48], who introduced the GRACE-groundwater residual method. This physical connection, however, does not make their storage variations interchangeable when partitioning GRACE TWS among all individual components.

GRACE-based studies have often addressed missing SWS by combining land-model storage with independently estimated surface water from observations or separate hydrological models[6,49–51]. Such estimates can provide an otherwise missing physical reservoir but do not necessarily restore the internal consistency of an LSM, even if the external SWS were error free. SWS integrates the magnitude and timing of model-generated runoff together with routing and surface water dynamics, and the present results demonstrate that its contribution to TWS is generally beneficial, but also strongly model dependent. As a result, a realistic and observation-based external SWS estimate can complete the inventory of TWS components without producing a storage estimate dynamically consistent with each LSM. On the other hand, deriving SWS from each model's own runoff preserves this land–river connection[17].

This distinction is critical when groundwater is inferred as a GRACE residual. The absence of an explicit groundwater state in Noah and VIC does not mean that their modeled storage changes represent changes in TWS minus groundwater's contribution. These models partition water among the states and fluxes available in their formulations, so variability that might be represented as groundwater in another model may instead appear in soil moisture or other components. After model-consistent SWS is included, the remaining model–GRACE difference reflects groundwater-

related variability, but also includes differences in model structure, hydrological partitioning, forcing and parameterization, unresolved processes, human influences, and satellite retrieval uncertainty, largely characterized in the literature[7,11,12,52]. Consequently, the residual cannot be uniquely interpreted as groundwater. Rather, it incorporates disagreement between observed and modeled TWS whose internal attribution remains uncertain.

Trends provide a direct test of one possible alternative explanation for this result. One could expect that models without an explicit groundwater reservoir would, in principle, express missing subsurface storage changes through altered runoff and routed SWS. In other words, missing trends in LWS components could be detected in SWS as a result of compensations in runoff fluxes. Our trend tests show no systematic evidence for such compensation. Noah and VIC do not exhibit consistently stronger or more negative SWS trends than Noah-MP and CLSM, and, within each examined region, the sign of the SWS trend is common to all four simulations. So, the ability of Noah and VIC to reproduce aggregate TWS variability after SWS is included does not appear to arise from systematic transfer of missing groundwater trends into surface water. Instead, large remaining model–GRACE trend discrepancies evidence storage changes or errors not represented by the modeled system.

Explicit groundwater representation does not eliminate this ambiguity. Noah-MP and CLSM partition subsurface storage differently, and previous work shows large model dependence in the simulated contribution of groundwater and soil moisture to TWS[10]. Thus, even when an explicit groundwater state exists, it remains a model-specific partition of total subsurface storage rather than an independently observed groundwater quantity. More complete groundwater formulations and GRACE data assimilation can substantially alter simulated subsurface dynamics[11,53–55]. Consequently, the physical realism of groundwater representation is important for understanding groundwater processes, but it does not make the residual between GRACE and another model's TWS uniquely attributable to groundwater.

Implications for GRACE data assimilation and storage attribution

These results have direct implications for GRACE data assimilation. Because GRACE observes vertically integrated TWS rather than individual reservoirs, the modeled observation operator should include all storage components contributing appreciably to TWS variability. Where adding SWS improves agreement with GRACE, the simulated TWS is already closer to the observation and smaller assimilation corrections should be required. Where SWS degrades agreement, the larger mismatch instead identifies regions where GRACE assimilation and model development have greater potential to improve the simulated storage distribution. Thus, such degradation is not a reason to omit SWS; doing so would force the assimilation system to assign genuine river and floodplain storage variability to other reservoirs such as soil moisture, snow, or groundwater.

A second implication is that GRACE TWS assimilation is not restricted to LSMs with an explicit groundwater reservoir. Noah and VIC reproduce much of the observed TWS variability once model-consistent SWS is included, showing that GRACE can constrain total storage across models with different subsurface structures. This does not reduce the importance of explicit groundwater modeling when groundwater dynamics, pumping, irrigation, or groundwater-specific storage changes are the quantities of interest. In other words, GRACE can constrain total water storage even when a model does not represent groundwater explicitly, but interpreting the individual storage components still requires physically realistic representations of those reservoirs.

Limitations

Meteorological forcing remains an important source of uncertainty because errors in precipitation and atmospheric conditions propagate through evapotranspiration, snow, soil water, runoff, and SWS. Using the same Princeton forcing[56] for all four simulations provides a controlled comparison that better isolates differences in LSM structure, but does not quantify forcing uncertainty or model–forcing interactions. A previous 12-member forcing experiment with a single land surface and routing system demonstrated appreciable uncertainty in TWS component contributions[17]. A comprehensive extension would combine multiple LSMs and meteorological forcings. However, convergence across four structurally distinct models indicates that the central SWS result is not specific to a single LSM formulation.

Human water management is another important limitation. Groundwater abstraction, irrigation withdrawals, or reservoir operations, all of which can alter SWS and TWS variability[57–59], are absent in the simulations. Their omission is particularly relevant in intensively managed regions, such as Northern India, California, and the Tigris–Euphrates and near major reservoirs such as Aswan and Itaipu. Incorporating human water management would improve the physical attribution of regional storage changes, but its omission reinforces rather than resolves the ambiguity of interpreting the model–GRACE residual as groundwater.

Methods

GLDAS–HyMAP framework

We analyzed four global land surface model (LSM) simulations from the Global Land Data Assimilation System version 2 (GLDAS-2): Noah[60], VIC[61], Noah-MP[62], and the Catchment Land Surface Model (CLSM[63]). The simulations were driven by the Princeton meteorological forcing dataset, and evaluated over 2003–2014, during which all datasets overlap. The four LSMs differ in their representation of infiltration, soil-water storage, snow, evapotranspiration, runoff generation, subsurface drainage, and groundwater processes. Noah-MP and CLSM include separately diagnosed subsurface/groundwater storage terms, whereas Noah and VIC do not diagnose an explicit groundwater storage reservoir in the configurations analyzed here. Table 1 summarizes key LSM physics (a more detailed description of models is provided in Supplementary Text S1).

Surface and subsurface runoff from each LSM independently forced the Hydrological Modeling and Analysis Platform version 3 (HyMAP-3[20,21]) global hydrodynamic model. Thus, four HyMAP-3 simulations were generated, one for each LSM, and every land storage estimate was paired only with the surface water storage generated from its own runoff. The HyMAP-3 configuration with GLDAS-2 adopts the local inertia formulation[64], accounting for a more stable and computationally efficient representation of backwater effects. The Courant–Friedrichs–Levy (CFL) condition is used in order to determine HyMAP-3's optimal time steps for numerical stability. Coastal boundary conditions are set at a zero-meter constant sea level. One-degree global HyMAP-3 river geometry and roughness, as well as floodplain roughness were estimated following Getirana et al.[20]. River network, topology, length and elevation, as well as floodplain topography were derived from the Multi-Error-Removed Improved-Terrain (MERIT) Hydro DEM at 3-arcsec spatial resolution[65].

Modeled storage components

For LSM m, land water storage (LWS) was defined as

$$\mathrm{LWS}_m(t) = \sum_{\ell=1}^{L_m} \mathrm{SM}_{m,\ell}(t) + \mathrm{SWE}_m(t) + \mathrm{GWS}_m(t)$$

where $\mathrm{SM}_{m,\ell}$ is soil moisture in layer ℓ, L_m is the number of soil layers, SWE is snow water equivalent, and GWS is the separately diagnosed groundwater or catchment subsurface-storage term, when available. All units are in millimeters. For model configurations without such a diagnosed state, $\mathrm{GWS} = 0$, meaning that subsurface water storage is represented within soil moisture states instead. Canopy-interception water storage was neglected due to its insignificant contribution relative to the other water components[66].

HyMAP-3 calculates surface water storage (SWS), in millimeters, from river channel and floodplain water interactions. Two modeled TWS definitions were evaluated:

$$\mathrm{TWS}_m^0(t) = \mathrm{LWS}_m(t)$$

and

$$\mathrm{TWS}_m^1(t) = \mathrm{LWS}_m(t) + \mathrm{SWS}_m(t)$$

where superscripts 0 and 1 denote simulations without and with SWS, respectively. Comparing TWS^0 and TWS^1 isolates the influence of river and floodplain storage while preserving the forcing, parameterization, and land water storage structure of each LSM.

Surface water storage impact on modeled TWS

The relative contribution of SWS to modeled TWS variability was quantified using the impact index introduced by Getirana et al.[17]. For each storage component k, the contribution C_k was calculated from the absolute anomalies of its mean seasonal cycle:

$$C_{m,k} = \sum_{t=1}^{12} \left| \overline{S}_{m,k,t} - \overline{\overline{S}}_{m,k} \right|$$

where t denotes calendar month, $\overline{S}_{m,k,t}$ is the 2003–2014 climatological mean storage for calendar month t, and $\overline{\overline{S}}_{m,k}$ is the mean of the 12 monthly climatological values. The impact of component k was then

$$I_{m,k} = 100\ \frac{C_{m,\mathrm{k}}}{\sum_{j=1}^{nc} C_{m,j}}$$

Where *nc* contains the storage components represented in model m: aggregated soil moisture, snow water equivalent, SWS, and the explicitly diagnosed groundwater storage term when available. The component impacts sum to 100% at each grid cell. Right panels in Fig. 1 report I_{SWS}. Global impact values were obtained from area-weighted summaries of the grid-cell impact field. Because the impact index is based on the relative amplitudes of separately diagnosed storage components, it should be interpreted as a component contribution to the modeled seasonal storage cycle rather than as a formal decomposition of TWS variance.

Evaluation of modeled terrestrial water storage

Grid-based modeled storage variations were evaluated against the Center for Space Research (CSR) Release 06.3 GRACE mascon TWS product[67]. All model and satellite fields are available

at the same 1-degree monthly spatiotemporal resolution – Greenland and Antarctica excluded. All storage anomalies are referenced to a fixed 2004–2009 baseline. Modeled LWS and SWS anomalies were filtered separately using a spherical Gaussian kernel with a 300-km half-width at half maximum, ensuring spatial and temporal consistency with the GRACE data.

Each LSM was evaluated independently against GRACE before and after adding its corresponding HyMAP-3 SWS. Temporal agreement was measured using the Pearson correlation coefficient. For model m, $r_m = r_m^1 - r_m^0$, so positive values indicate improved correlation after SWS is included. The representation of TWS amplitude was evaluated using the model-to-GRACE standard deviation ratio

$$\gamma_m^j = \frac{\sigma\left(\mathrm{TWS}_m^j\right)}{\sigma(\mathrm{TWS_{GRACE}})}, \quad j \in \{0,1\}$$

A value of one denotes equal temporal amplitude.

Because proportional underestimation and overestimation represent equivalent departures from observed variability, the SWS effect was quantified using the bounded log-symmetric improvement index

$$I_{\gamma,m} = \frac{|ln(\gamma_m^0)| - |ln(\gamma_m^1)|}{|ln(\gamma_m^0)| + |ln(\gamma_m^1)|}$$

The index ranges from -1 to 1. Positive values indicate that adding SWS moves modeled variability closer to GRACE, zero indicates no change, and negative values indicate degradation. A value of one indicates complete removal of the original amplitude error.

Linear storage trends over 2003–2014 were estimated at each grid cell from the seasonality-removed monthly anomaly time series using ordinary least-squares regression against time, expressed in mm·yr^{-1}. For each LSM m, trends were calculated separately for modeled LWS, SWS, and TWS, and GRACE TWS, using the same common valid months.

We first quantified the contribution of SWS to the pre-existing model–GRACE trend discrepancy using a trend-gap closure fraction

$$\mathrm{F_m} = \frac{\beta_{\mathrm{SWS,m}}}{\beta_{\mathrm{G}} - \beta_{\mathrm{m}}^0}$$

where β_m^0 is the LWS trend, $\beta_{SWS,m}$ is the SWS trend, and β_G is the GRACE TWS trend. Values $0 < F_m < 1$ indicate partial closure of the original GRACE–LWS trend gap, $F_m = 1$ represents complete closure, and $1 < F_m < 2$ indicates overshoot while still leaving the modeled trend closer to GRACE than before SWS was added. Negative values indicate that the SWS trend acts in the opposite direction from the original discrepancy, whereas values above 2 indicate an overshoot large enough to increase the model–GRACE trend error.

Also, the absolute model error in trend was defined as

$$E_m^j = \left|\beta_m^j - \beta^G\right|, \quad j \in \{0,1\}$$

where β_m^1 denotes the modeled TWS trend. Adding SWS changes the modeled trend by $\beta_{SWS,m}$; equivalently, it either closes part of the original GRACE–LWS trend gap, closes it completely, or overshoots it. The SWS effect on trend agreement was expressed as the bounded symmetric improvement index

401 $$I_{\mathrm{trend},m} = \frac{E_m^0 - E_m^1}{E_m^0 + E_m^1}$$

402 This index ranges from -1 to 1. Positive values indicate that adding SWS brings the modeled TWS
403 trend closer to GRACE, zero indicates no change in absolute trend error, and negative values
404 indicate degradation. A value of one indicates complete removal of the original trend error. The
405 bounded form avoids the very large magnitudes that can occur in the raw closure fraction when
406 the original GRACE–LWS trend difference approaches zero.

407 Overall disagreement was quantified with root mean square error (RMSE) as $\Delta\mathrm{RMSE}_m =$
408 $\mathrm{RMSE}_m^1 - \mathrm{RMSE}_m^0$, where negative ΔRMSE indicates improvement.

409 At each grid cell, GRACE-based fields were summarized as the median of the four model-specific
410 paired changes or improvements.

411 The influence of SWS on consistency among the four models was evaluated independently of
412 GRACE. At each grid cell and month, multi-model spread was defined as the standard deviation
413 across the four modeled TWS anomalies. Then, the time-varying spread, S, was summarized by
414 its temporal root mean square (RMS). To account for differences in the amplitude, *A*, of the
415 modeled storage signal, the RMS multi-model spread was normalized by the temporal standard
416 deviation of the four-model ensemble mean TWS anomaly. The multi-model spread-to-signal ratio
417 was then calculated as

418 $$R^j = \frac{S^j}{A^j}, \qquad j \in \{0,1\}$$

419 And the influence of adding SWS was expressed as the absolute change in this ratio as
420 $\Delta R = R^1 - R^0$, for which negative values denote convergence.

421 *Significance of local changes*

422 Grid-cell significance for changes in correlation, standard-deviation ratio error, RMSE, and multi-
423 model spread-to-signal ratio was evaluated using 5,000 Monte Carlo realizations. Each GRACE
424 and modeled storage time series was represented by its linear trend, mean seasonal cycle, and
425 residual seasonality-removed monthly variability. For each Monte Carlo realization, the observed
426 trend and seasonal cycle were retained, while the remaining monthly variability was randomly
427 resampled in a way that preserves its month-to-month persistence and spatial coherence. For each
428 realization, TWS was reconstructed as LWS + SWS and the corresponding metric change was
429 recalculated. Two-sided p-values were adjusted independently for each metric using the
430 Benjamini–Hochberg false-discovery-rate procedure[68] at q=0.05. Hatching in Fig. 2a, b, d, and e
431 identifies statistically significant improvement or degradation.

432 Trend-error robustness in Fig. 2c was evaluated separately. At each grid cell, the change in absolute
433 trend error relative to GRACE was calculated independently for the four LSMs after adding SWS.
434 Hatching is shown only where the GRACE TWS trend is statistically significant after false-
435 discovery-rate correction and at least three of the four LSMs agree on the direction of the SWS
436 effect. Agreement among at least three models that SWS reduces trend error indicates robust
437 improvement, whereas agreement that it increases trend error indicates robust degradation.

Trend-error robustness in Fig. 2c was evaluated separately. The significance of the GRACE TWS trend was tested after removing the mean seasonal cycle using the Hamed–Rao modified Mann–Kendall test[69], which accounts for temporal autocorrelation. The resulting p-values were adjusted across grid cells using the Benjamini–Hochberg false-discovery-rate procedure at q=0.05. At each grid cell, the change in absolute trend error relative to GRACE was then calculated independently for the four LSMs after adding SWS. Hatching is shown where the GRACE trend is significant and at least three of the four LSMs agree on the direction of the SWS effect. Agreement that SWS reduces trend error indicates robust improvement, whereas agreement that it increases trend error indicates robust degradation.

Significance of global and regional changes

Uncertainty in the global and <50°N area-weighted median changes was evaluated using the same 5,000 Monte Carlo realizations. For each realization, the complete metric fields were recalculated and their area-weighted medians were computed using grid-cell areas. The resulting distributions were used to calculate 95% confidence intervals and two-sided p-values for the observed median changes. False-discovery-rate correction was not applied to these predefined spatial summaries.

Regional study-period trend diagnostics

To examine whether routed SWS could compensate for storage changes not represented within the LSMs, we analyzed eight regions representing contrasting GRACE trends and hydrological settings[3,4,6,40,57,70,71]: California, Texas, Central Amazon, Itaipu dam, Eastern Brazil, Aswan dam, Tigris–Euphrates, and Northern India (Supplementary Fig. S3). Area-weighted monthly GRACE TWS and modeled LWS, SWS, and TWS anomalies were calculated for each region using common valid months. The mean seasonal cycle was removed before estimating linear trends over 2003–2014. Trends were calculated independently for each LSM, with modeled TWS defined as LWS + SWS. The influence of SWS was quantified as the change in absolute model–GRACE trend error after SWS was included. Model-specific trends were retained to determine whether the effect of SWS differed systematically between models with and without explicitly represented groundwater storage.

Basin-scale evaluation

Basin-scale analyses were performed for 18 major river basins. Correlation, standard deviation ratio, and RMSE were calculated from basin-average TWS and LWS time series. Basin-scale normalized RMSE was defined as

$$\mathrm{nRMSE}_{m,b}^{j} = \frac{\mathrm{RMSE}_{m,b}^{j}}{\sigma\left(\mathrm{TWS}_{\mathrm{GRACE},b}\right)}$$

Improvement corresponds to increased correlation, a standard deviation ratio closer to one, or reduced nRMSE.

Arctic/sub-Arctic basins were defined as basins for which at least 25% of analyzed grid cells are located at or north of 50°N. Basin degree of regulation was taken from BasinATLAS and used only as an independent diagnostic of the potential influence of river regulation.

Seasonal climatologies and normalized MAE

For each basin, month climatologies were calculated for GRACE and modeled TWS and LWS using the same valid monthly samples across all five time series (modeled LWS climatologies are presented in Supplementary Fig. S4). Mean values were subtracted from each climatology.

Both seasonal-cycle mean absolute error (MAE) and normalized MAE (nMAE) were used to evaluate climatologies. nMAE is defined as

$$\mathrm{nMAE}_{m,b} = \frac{\mathrm{MAE}_{m,b}}{\sigma\left(\mathrm{TWS}^{\mathrm{clim}}_{\mathrm{GRACE},b}\right)}$$

As a result, a normalized MAE of one corresponds to a climatological MAE equal to one standard deviation of the GRACE seasonal cycle.

Figures

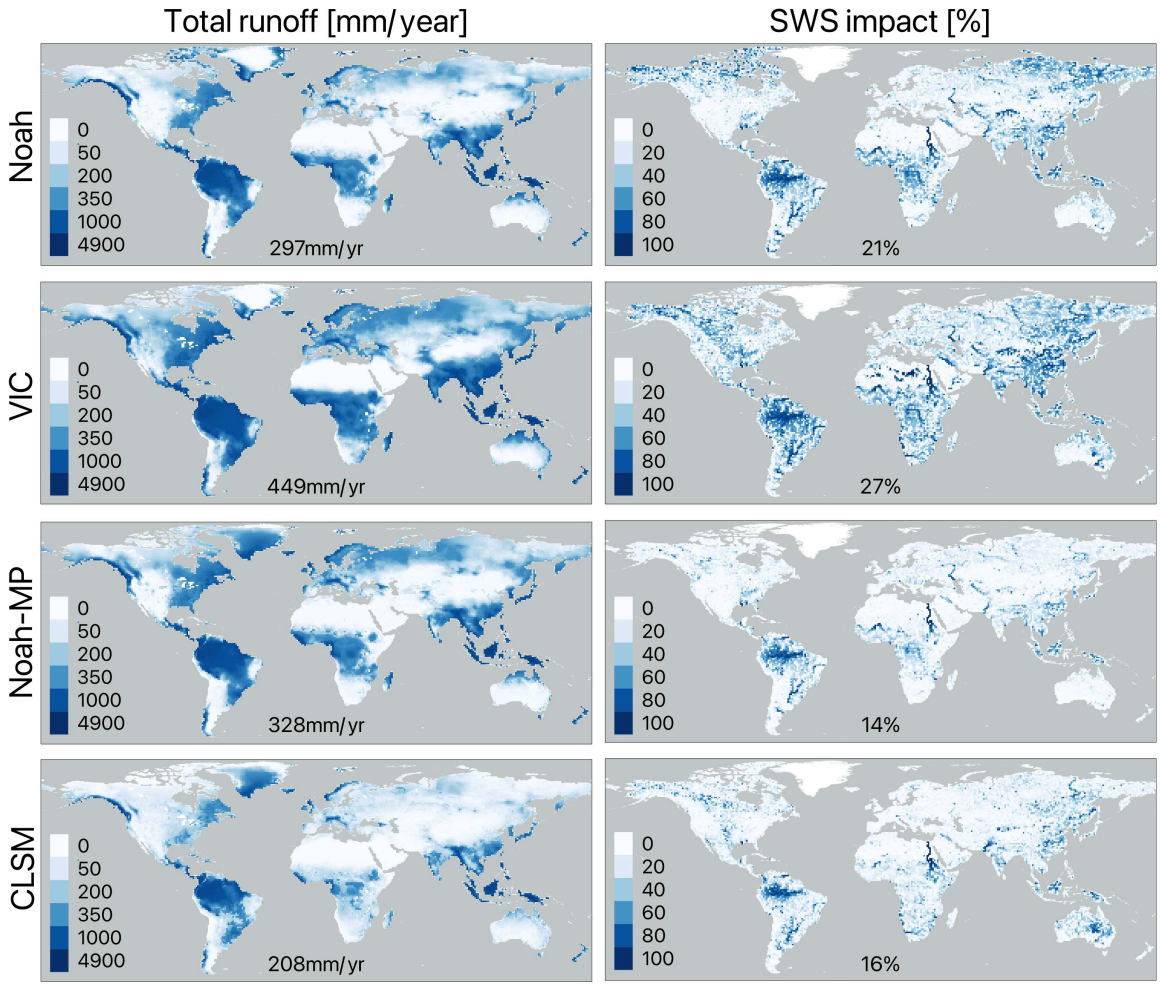
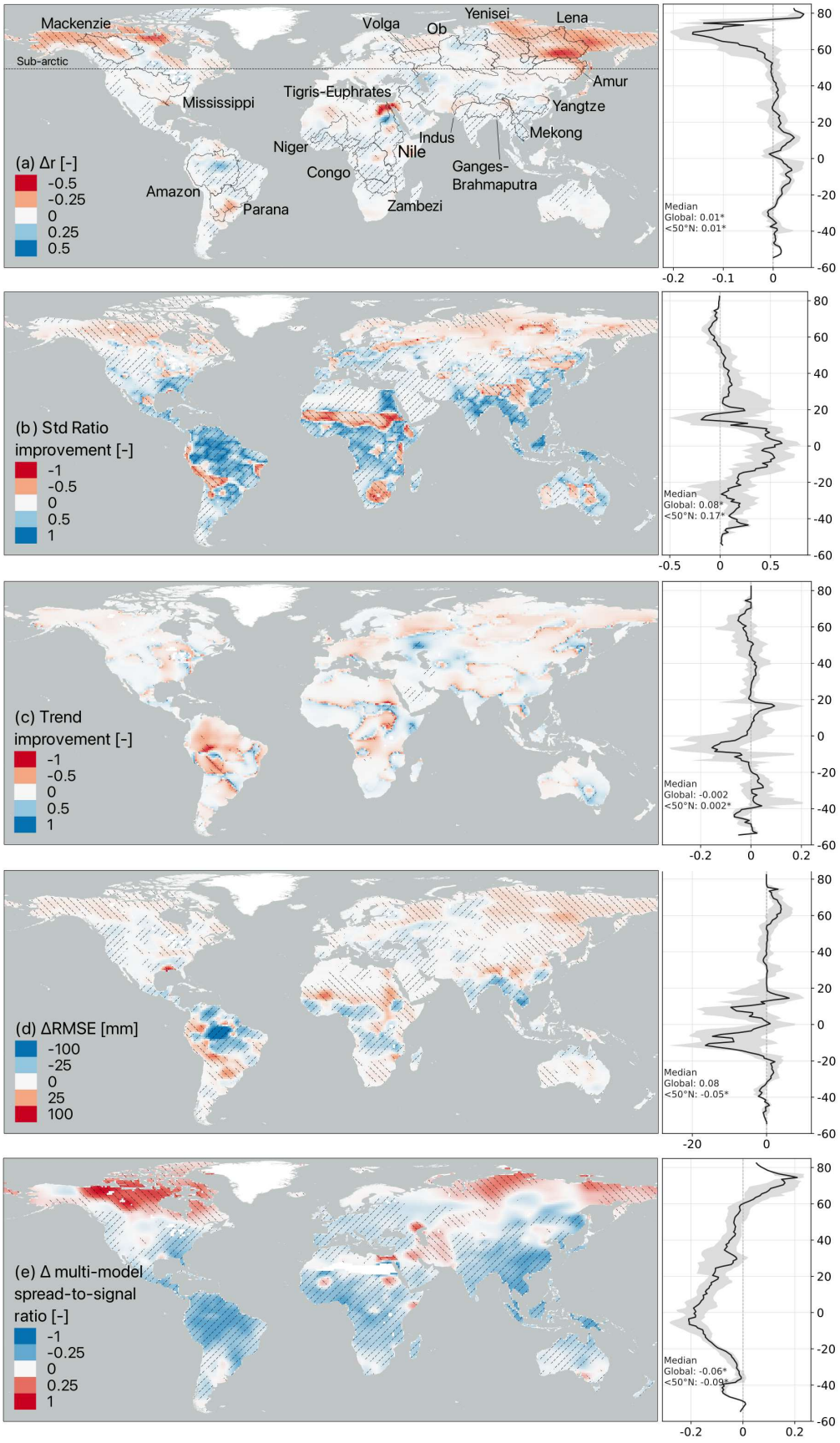


Figure 1. Global patterns of total runoff and surface water storage (SWS) impact index [%] on modeled seasonal terrestrial water storage (TWS) variability. The left column shows 2003-2014 mean annual total runoff, calculated as the sum of surface runoff and baseflow, in $mm \cdot yr^{-1}$. The right column shows the impact of SWS on modeled TWS variability. Rows correspond, from top to bottom, to the Noah, VIC, Noah-MP, and CLSM LSMs. All maps represent the 2003–2014 analysis period, with Greenland excluded.


Mackenzie
Volga
Ob
Yenisei
Lena
Sub-arctic
Amur
Mississippi
Tigris-Euphrates
Yangtze
Indus
Mekong
Niger
Nile
Congo
Ganges-
Brahmaputra
Amazon
Parana
Zambezi
(a) Δr [-]
-0.5
-0.25
0
0.25
0.5
Median
Global: 0.01*
<50°N: 0.01*
(b) Std Ratio
improvement [-]
-1
-0.5
0
0.5
1
Median
Global: 0.08*
<50°N: 0.17*
(c) Trend
improvement [-]
-1
-0.5
0
0.5
1
Median
Global: -0.002
<50°N: 0.002*
(d) ΔRMSE [mm]
-100
-25
0
25
100
Median
Global: 0.08
<50°N: -0.05*
(e) Δ multi-model
spread-to-signal
ratio [-]
-1
-0.25
0
0.25
1
Median
Global: -0.06*
<50°N: -0.09*

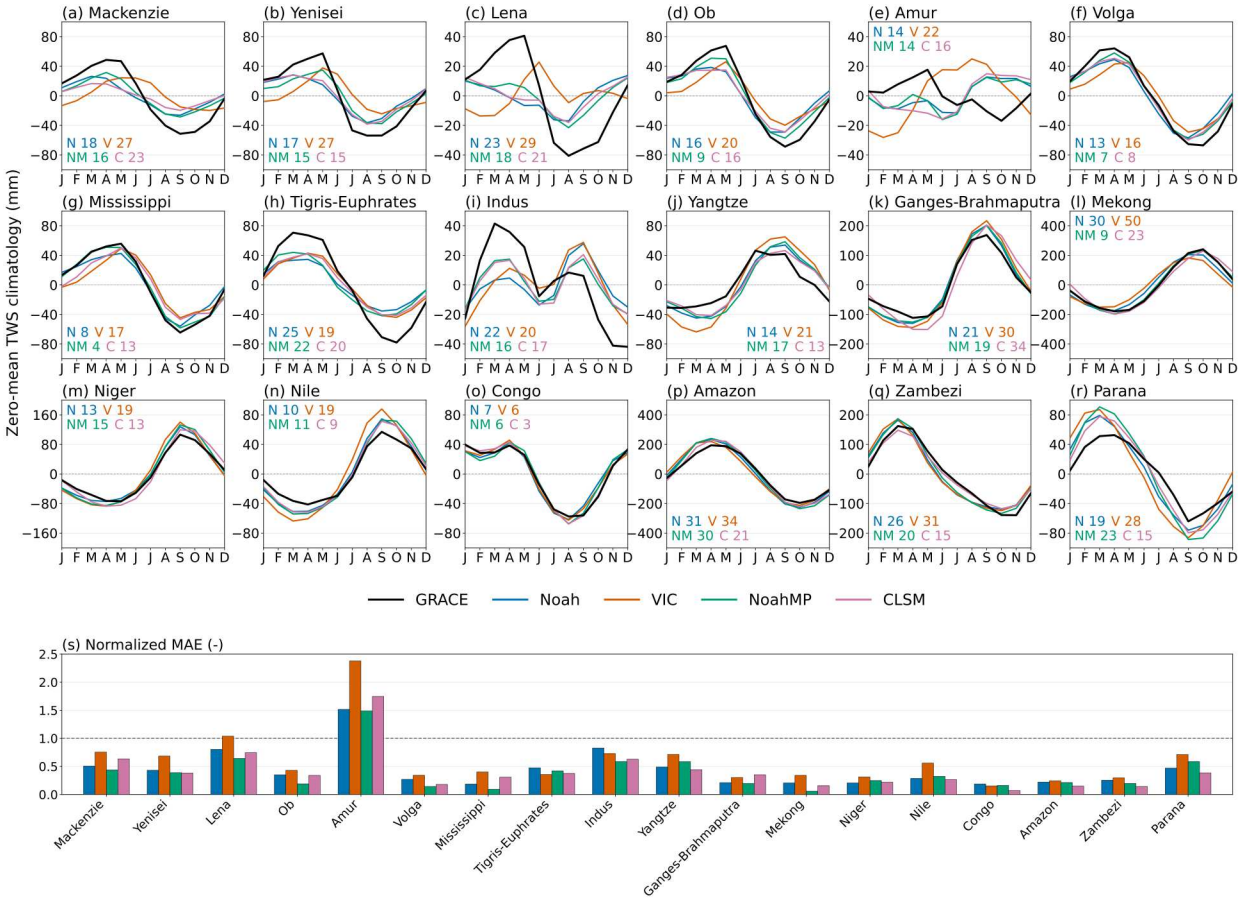
494

Figure 2. Global effects of surface water storage on terrestrial water storage simulations. (a) median change in correlation with GRACE across the four GLDAS-2 LSMs, (b) median log-symmetric improvement in standard deviation ratio error relative to GRACE, (c) bounded symmetric improvement in 2003–2014 linear trend error relative to GRACE, (d) median change in RMSE relative to GRACE, and (e) change in multi-model spread-to-signal ratio (multi-model consistency). Blue tones indicate improvement in all panels. For panels (a), (b), (d), and (e), hatching denotes grid cells where the corresponding change is statistically significant. In all panels, diagonal hatching "/" indicates improvement with the inclusion of SWS, whereas diagonal hatching "\" indicates degradation. Panel (a) also shows 18 major river basins and the sub-Arctic boundary at 50°N. The accompanying latitudinal profiles show the median across longitudes at each latitude, with shading representing the interquartile range of the longitudinal distribution. Global and <50°N median values are provided. Asterisks (*) indicate medians that are statistically different from zero ($p<0.05$). Greenland is excluded from the analysis.



509

510 **Figure 3. Basin-scale effects of surface water storage on modeled terrestrial water storage**
511 **anomalies.** Paired performance metrics are shown for four GLDAS-2 LSMs relative to GRACE:
512 (a) correlation, (b) standard-deviation ratio, and (c) normalized root mean square error (nRMSE).
513 Metrics were calculated from area-weighted basin-average time series. Open and filled symbols
514 represent simulations without and with HyMAP-3 surface water storage (SWS), respectively;
515 connecting lines show the change produced by adding SWS. Improvement corresponds to
516 increased correlation, a standard-deviation ratio closer to unity, or reduced nRMSE. Light-blue
517 rows identify Arctic/sub-arctic basins, defined here as basins with at least 25% of their grid cells
518 at or north of 50°N. The degree of regulation (DOR) column reports BasinATLAS regulation
519 classes[46]. Panel summaries report improved and degraded model–basin pairs globally, with values
520 in parentheses excluding Arctic/subarctic basins.



521

522 **Figure 4. Basin-scale seasonal cycles of modeled and GRACE terrestrial water storage**
523 **anomalies.** Panels (a–r) show zero-mean monthly terrestrial water storage (TWS) climatologies
524 for 18 major river basins from GRACE and four GLDAS-2 LSMs coupled with HyMAP-3.
525 Modeled TWS includes land water storage and HyMAP-3 surface water storage. For each basin,
526 monthly means were calculated using the same valid monthly samples across GRACE and all four
527 models, and the annual mean of each climatology was removed to isolate the seasonal cycle.
528 Numbers within each panel report the mean absolute error (MAE; mm) of each modeled
529 climatology relative to GRACE. Panel (s) summarizes normalized MAE for each model and basin.

530

Acknowledgements

Computing resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at NASA's Goddard Space Flight Center. GLDAS-2 and HyMAP-3 outputs are freely available through GESDISC (https://disc.gsfc.nasa.gov/datasets?keywords=GLDAS).

References

1. Tapley, B., Bettadpur, S., Watkins, M. & Reigber, C. The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters* **31**, 1–4 (2004).
2. Landerer, F. W. *et al.* Extending the Global Mass Change Data Record: GRACE Follow-On Instrument and Science Data Performance. *Geophysical Research Letters* **47**, e2020GL088306 (2020).
3. Rodell, M. *et al.* Emerging trends in global freshwater availability. *Nature* **557**, 651–659 (2018).
4. Rodell, M., Velicogna, I. & Famiglietti, J. S. Satellite-based estimates of groundwater depletion in India. *Nature* **460**, 999–1002 (2009).
5. Famiglietti, J. S. *et al.* Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophysical Research Letters* **38**, 2–5 (2011).
6. Voss, K. A. *et al.* Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region. *Water Resources Research* **49**, 904–914 (2013).
7. Scanlon, B. R. *et al.* Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National Academy of Sciences* **115**, (2018).
8. Scanlon, B. R. *et al.* Tracking Seasonal Fluctuations in Land Water Storage Using Global Models and GRACE Satellites. *Geophysical Research Letters* **46**, 5254–5264 (2019).

9. Cai, X. *et al.* Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed. *Journal of Geophysical Research: Atmospheres* **119**, 13,751-13,770 (2014).
10. Xia, Y. *et al.* Comparison and assessment of three advanced land surface models in simulating terrestrial water storage components over the United States. *Journal of Hydrometeorology* **18**, 625–649 (2017).
11. Getirana, A., Kumar, S. & Rodell, M. Inconsistencies in GRACE-Based Groundwater Storage Estimation—A Call for a Proper Use of Land Surface Models. *Geophysical Research Letters* **52**, e2025GL119197 (2025).
12. Arifin, A. *et al.* Plausibility Criteria for GRACE-Derived Groundwater Storage Changes From Aquifers Globally. *Geophysical Research Letters* **52**, e2025GL118580 (2025).
13. Kim, H., Yeh, P. J.-F., Oki, T. & Kanae, S. Role of rivers in the seasonal variations of terrestrial water storage over global basins. *Geophysical Research Letters* **36**, (2009).
14. Papa, F. *et al.* Surface freshwater storage and variability in the Amazon basin from multi-satellite observations, 1993-2007. *Journal of Geophysical Research Atmospheres* **118**, (2013).
15. Salameh, E. *et al.* Fifteen years (1993-2007) of surface freshwater storage variability in the ganges-brahmaputra river basin using multi-satellite observations. *Water (Switzerland)* **9**, 1–19 (2017).
16. Paiva, R. C. D. *et al.* Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. *Water Resources Research* **49**, 1226–1243 (2013).

17. Getirana, A., Kumar, S., Girotto, M. & Rodell, M. Rivers and Floodplains as Key Components of Global Terrestrial Water Storage Variability. *Geophysical Research Letters* **44**, 10,359-10,368 (2017).
18. Getirana, A. *et al.* Water balance in the amazon basin from a land surface model ensemble. *Journal of Hydrometeorology* **15**, 2586–2614 (2014).
19. Rodell, M. *et al.* Overview of Global Land Data Assimilation System Datasets and Ongoing Development. *Journal of Hydrometeorology* (2027).
20. Getirana, A. C. V. *et al.* The hydrological modeling and analysis platform (HyMAP): Evaluation in the Amazon basin. *Journal of Hydrometeorology* **13**, 1641–1665 (2012).
21. Getirana, A., Kumar, S., Arsenault, K., Yoon, Y. & Geiger, J. *The HyMAP-3 Hydrodynamic Model*. 30 https://ntrs.nasa.gov/api/citations/20250009546/downloads/HyMAP3_tech_memo_v3.pdf (2025).
22. Li, B. *et al.* Groundwater Recharge Estimated by Land Surface Models: An Evaluation in the Conterminous United States. https://doi.org/10.1175/JHM-D-20-0130.1 (2021) doi:10.1175/JHM-D-20-0130.1.
23. Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E. & Su, H. Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. *Journal of Geophysical Research* **112**, D07103 (2007).
24. Miguez-Macho, G. & Fan, Y. The role of groundwater in the Amazon water cycle: 1. Influence on seasonal streamflow, flooding and wetlands. *Journal of Geophysical Research Atmospheres* **117**, 1–30 (2012).

25. Jankowski, K. L., Törnqvist, T. E. & Fernandes, A. M. Vulnerability of Louisiana's coastal wetlands to present-day rates of relative sea-level rise. *Nature Communications* **8**, 14792 (2017).
26. Blum, M. D. & Roberts, H. H. Drowning of the Mississippi Delta due to insufficient sediment supply and global sea-level rise. *Nature Geoscience* **2**, 488–491 (2009).
27. Getirana, A. *et al.* Climate and Human Impacts on Hydrological Processes and Flood Risk in Southern Louisiana. *Water Resources Research* **59**, (2023).
28. Kim, R. S. *et al.* Snow Ensemble Uncertainty Project (SEUP): quantification of snow water equivalent uncertainty across North America via ensemble land surface modeling. *The Cryosphere* **15**, 771–791 (2021).
29. Maina, F. Z. & Kumar, S. V. Global patterns of rain-on-snow and its impacts on runoff from past to future projections. *Nat Commun* **16**, 4731 (2025).
30. Farhoodi, S., Trudel, M. & Leconte, R. The effect of hydrological model structure on spring flow forecasts when assimilating a distributed snow product. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques* 1–20 (2024) doi:10.1080/07011784.2024.2434517.
31. Musselman, K. N. *et al.* Projected increases and shifts in rain-on-snow flood risk over western North America. *Nature Clim Change* **8**, 808–812 (2018).
32. Slater, A. G., Bohn, T. J., McCreight, J. L., Serreze, M. C. & Lettenmaier, D. P. A multimodel simulation of pan-Arctic hydrology. *Journal of Geophysical Research: Biogeosciences* **112**, (2007).
33. Maina, F. Z. & Kumar, S. V. Diverging Trends in Rain-On-Snow Over High Mountain Asia. *Earth's Future* **11**, e2022EF003009 (2023).

34. Maina, F. Z. *et al.* Development of a multidecadal land reanalysis over High Mountain Asia. *Sci Data* **11**, 827 (2024).
35. Bhattacharya, A. *et al.* High Mountain Asian glacier response to climate revealed by multi-temporal satellite observations since the 1960s. *Nat Commun* **12**, 4133 (2021).
36. Getirana, A. *et al.* Streamflows over a West African basin from the ALMIP2 model ensemble. *Journal of Hydrometeorology* **18**, 1831–1845 (2017).
37. Grippa, M. *et al.* Modeling Surface runoff and water fluxes over contrasted soils in the pastoral sahel: Evaluation of the ALMIP2 land surface models over the Gourma Region in Mali. *Journal of Hydrometeorology* **18**, 1847–1866 (2017).
38. Grippa, M. *et al.* Land water storage variability over West Africa estimated by Gravity Recovery and Climate Experiment (GRACE) and land surface models. *Water Resources Research* **47**, 1–18 (2011).
39. Dadson, S. J. *et al.* Wetland inundation dynamics in a model of land surface climate: Evaluation in the Niger inland delta region. *Journal of Geophysical Research* **115**, D23114 (2010).
40. Getirana, A., Kumar, S., Konapala, G. & Ndehedehe, C. E. Impacts of fully coupling land surface and flood models on the simulation of large wetlands' water dynamics: the case of the Inner Niger Delta. *Journal of Advances in Modeling Earth Systems* https://doi.org/10.1029/2021MS002463 (2021) doi:10.1029/2021MS002463.
41. Neal, J., Schumann, G. & Bates, P. A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. *Water Resources Research* **48**, 1–16 (2012).

42. Kumar, S. V. *et al.* Assimilation of Gridded GRACE terrestrial water storage estimates in the North American land data assimilation system. *Journal of Hydrometeorology* **17**, 1951–1972 (2016).
43. Zaitchik, B. F., Rodell, M. & Reichle, R. H. Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River basin. *Journal of Hydrometeorology* **9**, 535–548 (2008).
44. Swenson, S., Yeh, P. J. F., Wahr, J. & Famiglietti, J. A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois. *Geophysical Research Letters* https://doi.org/10.1029/2006GL026962 (2006) doi:10.1029/2006GL026962.
45. Springer, A. *et al.* A review of current best practices and future directions in assimilating GRACE/-FO terrestrial water storage data into numerical models. *Hydrology and Earth System Sciences* **30**, 985–1022 (2026).
46. Linke, S. *et al.* Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Sci Data* **6**, 283 (2019).
47. Winter, T. C., Harvey, J. W., Franke, O. L. & Alley, W. M. *Ground Water and Surface Water: A Single Resource*. *Circular* https://pubs.usgs.gov/publication/cir1139 (1998) doi:10.3133/cir1139.
48. Rodell, M. *et al.* Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeol J* **15**, 159–166 (2007).
49. Shamsudduha, M., Taylor, R. G. & Longuevergne, L. Monitoring groundwater storage changes in the highly seasonal humid tropics: Validation of GRACE measurements in the Bengal Basin. *Water Resources Research* **48**, (2012).

50. Khandu, Forootan, E., Schumacher, M., Awange, J. L. & Müller Schmied, H. Exploring the influence of precipitation extremes and human water use on total water storage (TWS) changes in the Ganges-Brahmaputra-Meghna River Basin. *Water Resources Research* **52**, 2240–2258 (2016).
51. Satish Kumar, K., AnandRaj, P., Sreelatha, K. & Sridhar, V. Reconstruction of GRACE terrestrial water storage anomalies using Multi-Layer Perceptrons for South Indian River basins. *Science of The Total Environment* **857**, 159289 (2023).
52. Landerer, F. W. & Swenson, S. C. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research* **48**, 1–11 (2012).
53. Maina, F. Z. *et al.* Integrating the Interconnections Between Groundwater and Land Surface Processes Through the Coupled NASA Land Information System and ParFlow Environment. *Journal of Advances in Modeling Earth Systems* **17**, e2024MS004415 (2025).
54. Abbaszadeh, P. *et al.* Coupling the ParFlow Integrated Hydrology Model within the NASA Land Information System: a case study over the Upper Colorado River Basin. *Hydrology and Earth System Sciences* **29**, 5429–5452 (2025).
55. Li, B. *et al.* Global GRACE Data Assimilation for Groundwater and Drought Monitoring: Advances and Challenges. *Water Resources Research* **55**, 1–23 (2019).
56. Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling. *Journal of Climate* **19**, 3088–3111 (2006).
57. Longuevergne, L., Wilson, C. R., Scanlon, B. R. & Crétaux, J. F. GRACE water storage estimates for the Middle East and other regions with significant reservoir and lake storage. *Hydrology and Earth System Sciences* **17**, 4817–4830 (2013).

58. Getirana, A., Jung, H. C., Hoek, J. V. D. & Ndehedehe, C. E. Hydropower dam operation strongly controls Lake Victoria's freshwater storage variability. *Science of The Total Environment* **726**, 138343 (2020).
59. Xie, J., Xu, Y.-P., Booij, M. J. & Guo, Y. Influences of reservoir operation on terrestrial water storage changes detected by GRACE in the Yellow River basin. *Journal of Hydrology* **610**, 127924 (2022).
60. Ek, M. B. *et al.* Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres* **108**, (2003).
61. Liang, X., Lettenmaier, D. P., Wood, E. F. & Burges, S. J. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres* **99**, 14415–14428 (1994).
62. Niu, G. Y. *et al.* The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research Atmospheres* **116**, 1–19 (2011).
63. Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M. & Kumar, P. A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *Journal of Geophysical Research: Atmospheres* **105**, 24809–24822 (2000).
64. Bates, P. D., Horritt, M. S. & Fewtrell, T. J. A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *Journal of Hydrology* **387**, 33–45 (2010).

65. Yamazaki, D. *et al.* MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset. *Water Resources Research* https://doi.org/10.1029/2019WR024873 (2019) doi:10.1029/2019WR024873.
66. Güntner, A. *et al.* A global analysis of temporal and spatial variations in continental water storage. *Water Resources Research* **43**, (2007).
67. Save, H., Bettadpur, S. & Tapley, B. D. High-resolution CSR GRACE RL05 mascons. *Journal of Geophysical Research: Solid Earth* **121**, 7547–7569 (2016).
68. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Royal Statistical Society. Journal. Series B: Methodological* **57**, 289–300 (1995).
69. Hamed, K. H. & Ramachandra Rao, A. A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology* **204**, 182–196 (1998).
70. Liu, P.-W. *et al.* Groundwater depletion in California's Central Valley accelerates during megadrought. *Nat Commun* **13**, 7825 (2022).
71. Famiglietti, J. S. & Rodell, M. Water in the balance. *Science* **340**, 1300–1301 (2013).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarymaterial.pdf