The global land water storage data set release 2 (GLWS2.0) derived via assimilating GRACE and GRACE-FO data into a global hydrological model

Helena Gerdener1 · Jürgen Kusche1 · Kerstin Schulze1 · Petra Döll2,3 · Anna Klos1,4

Received: 29 November 2022 / Accepted: 4 July 2023 / Published online: 26 July 2023 © The Author(s) 2023

Abstract
We describe the new global land water storage data set GLWS2.0, which contains total water storage anomalies (TWSA) over the global land except for Greenland and Antarctica with a spatial resolution of 0.5°, covering the time frame 2003 to 2019 without gaps, and including monthly uncertainty quantification. GLWS2.0 was derived by assimilating monthly GRACE/-FO mass change maps into the WaterGAP global hydrology model via the ensemble Kalman filter, taking data and model uncertainty into account. TWSA in GLWS2.0 is then accumulated over several hydrological storage variables. In this article, we describe the methods and data sets that went into GLWS2.0, how it compares to GRACE/-FO data in terms of representing TWSA trends, seasonal signals, and extremes, as well as its validation via comparing to GNSS-derived vertical loading and its comparison with a version of the NASA Catchment Land Surface Model GRACE Data Assimilation (CLSM-DA). We find that, in the average over more than 1000 stations globally, GLWS2.0 correlates better with GNSS observations of vertical loading at short-term, seasonal, and long-term temporal bands than GRACE/-FO. While some differences exist, overall GLWS2.0 agrees reasonably well with CLSM-DA in terms of TWSA trends and annual amplitudes and phases.

Highlights
• We describe the new global land water storage data set GLWS2.0, which contains total water storage anomalies over the global land with a spatial resolution of 0.5°, covering the period 2003 to 2019 without gaps, and including uncertainty quantification.
• GLWS2.0 synthesizes monthly GRACE/-FO mass change maps with daily precipitation and radiation data via the WaterGAP model framework, taking data and model uncertainty into account.
• Here we describe the methods and data sets that went into GLWS2.0 and its validation from a geodetic applications perspective. We find that, in the global average, GLWS2.0 fits better than GRACE/-FO to GNSS observations of vertical loading.

Keywords GLWS · GRACE · GRACE-FO · Total water storage anomalies · Data assimilation · Global

1 Introduction
The GRACE and GRACE-FO satellite missions (Tapley et al. 2019) have provided unique and unprecedented data sets for studying the terrestrial water cycle, its natural variability, and its response to precipitation, radiative forcing, land use change, and direct anthropogenic alterations (Rodell et al. 2018). GRACE/-FO-derived maps of monthly total water storage anomalies (TWSA) have been used to quantify long-term trends in glacier mass balance (Wouters et al. 2019), groundwater stress (Richey et al. 2015), large-scale droughts
(Zhao et al. 2017; Gerdener et al. 2020) and floods (Reager et al. 2014; Han et al. 2021), the vegetation response to moisture (A et al. 2017; Gerdener et al. 2022), to evaluate land surface and atmospheric models at different timescales (Springer et al. 2014), and to improve the representation of soil processes in models (Swenson and Lawrence 2014).

Apart from these climate and hydrology applications, GRACE/-FO data are increasingly employed in geodesy and geodynamics in lieu of model-based mass redistribution maps, e.g., for deriving environmental loading computations for satellite altimetry (Ray et al. 2013) and Global Navigation Satellite System (GNSS) (Chanard et al. 2018; Ménin et al. 2020), deriving geocenter estimates (Wu et al. 2017; Razeghi et al. 2019), simulating the hydrological (HAM) and ocean (OAM) angular momentum contributions to Earth rotation variations (Seoane et al. 2009; Jin et al. 2010), inferring rheological properties of the Earth (Broerse et al. 2015; Rovira-Navarro et al. 2020), or for deriving empirical or semi-empirical models of glacial isostatic adjustment (GIA) (Sun and Riva 2020).

Yet the effective GRACE/-FO spatial resolution of about 300 km, after appropriate post-processing, is still too coarse for several applications in hydrology and climate but also in geodesy and geodynamics (Pail et al. 2015; Wiese et al. 2022). In addition, GRACE errors are not stationary with respect to time due to switching off the thermal control of the accelerometers in 2010 and other hardware problems. Monthly data gaps exist, and an eleven-month gap between the two missions could not be avoided. All this means that the time series of TWSA maps are difficult to work with if either spatial resolution or temporal consistency is an issue. Several simulation studies have shown that improved mission concepts with multiple spacecraft and upcoming technology are expected to lead to notable improvements, but not to a quantum leap in spatial resolution (see Pail et al. (2015), Wiese et al. (2022) and references therein).

In recent years, studies made efforts to predict unobserved TWSA (e.g., during GRACE gaps or prior to the GRACE era) by combining GRACE/-FO maps with remote sensing or in-situ data via statistical or machine learning approaches (Humphrey and Gudmundsson 2019; Li et al. 2021). Also, many projects apply formal data assimilation (DA) techniques (Zaitchik et al. 2008; Tangdamrongsub et al. 2015; Girotto et al. 2016; Li et al. 2019; Springer et al. 2019) in order to disaggregate observed TWSA into changes in the groundwater, soil, snow, or surface water storages. But to our knowledge, they all focused on hydrological applications, and only a few (if any) data products are publicly available including quantification of their uncertainty. For example, Tangdamrongsub et al. (2015) and Tangdamrongsub et al. (2020) investigate the improvement via assimilation in individual storages and fluxes by comparisons with independent groundwater, soil moisture, and streamflow records, and Schumacher et al. (2016) and Khaki et al. (2017) tested different assimilation schemes, but all these studies are regionally limited. As one of only a few global products, groundwater and soil moisture conditions from GRACE/-FO TWSA assimilation into the CSLM model (Li et al. 2019) for the contiguous USA and for the global land surface can be downloaded from the National Drought Mitigation Center, University of Nebraska-Lincoln, archive (https://nasagrace.unl.edu/) in the form of indicator percentile maps. In addition, the underlying water storage grids (https://disc.gsfc.nasa.gov/) are provided, but to our knowledge no uncertainty quantification is available. Lastly, several assimilation projects make use of GRACE/-FO mascon (i.e. level-3) data grids—while we acknowledge that these are easy to use and competitive to the standard GRACE/-FO solutions for many applications, the use of level-2 GRACE/-FO data allows one to design more flexible and optimized analysis schemes since the smoothing within the mascon processing cannot be undone in post-processing.

Following methods outlined in Eicker et al. (2014), we developed a global data assimilation framework based on the ensemble Kalman filter approach (Evensen 2003) to produce a global land water storage (GLWS) data set with 0.5° spatial resolution, here in version 2.0 (GLWS2.0). Via the WaterGAP (Müller Schmied 2021) modeling framework, TWSA is at the same time vertically disaggregated into ten different water compartments including groundwater, soil, snow, and surface water (river, reservoir, wetland, and lake) storage changes. WaterGAP runs had been evaluated before with respect to GRACE/-FO TWSA data, both at grid and basin scale and at various timescales, and were found to perform with mixed results (Scanlon et al. 2018, 2019; Müller Schmied 2021). Similarly, GRACE TWSA maps had also been used to calibrate WaterGAP model parameters before (Werth and Günther 2010).

GLWS2.0 can be viewed as synthesized from GRACE/-FO mass change grids and precipitation and radiation data grids (the WaterGAP forcing data), taking data uncertainties into account, by including several flow-storage relations that are conceptualized in the hydrological model. In other words, with GLWS2.0 we provide a ‘level-3,’ downscaled TWSA data set (next to ‘level-2’ hydrological variables, such as groundwater and surface water storage anomalies) that we suggest can provide reference surface mass change maps for geodesy and geodynamic studies. Other studies (Eicker et al. 2014; Schumacher et al. 2018; Gerdener et al. 2022) had assimilated TWSA in WaterGAP but this is the first global assimilation with this model. As mentioned above, several others have assimilated TWSA data into hydrological or land surface models, but many early studies used basin-averaged in place of gridded TWSA data, the resulting downscaled maps are mostly not publicly available, and the models that were used employed very different representations of the
terrestrial water storages and fluxes (e.g., unlike WaterGAP most models miss anthropogenic water withdrawals).

We would like to point out that the new GLWS2.0 data set could be used as well in hydrology and climate modeling and validation exercises, having significantly higher effective spatial resolution than GRACE/-FO, but with the caveat that it is not a pure observational data set. We show in this contribution that GLWS2.0, in the spectral (spherical harmonic) domain, provides a smooth transition in power between GRACE data and WaterGAP simulation data. An earlier version of this data set (GLWS1.0) had been distributed within the GRACE community before; significant updates are a new WaterGAP model version with an improved representation of surface water and groundwater abstraction and several bug fixes.

This contribution focuses on the monthly GLWS2.0 TWSA data sets, how they were generated, and how they compare to the GRACE/-FO data and to the WaterGAP data. For an evaluation against independent data, we predict monthly average vertical station movements from TWSA grids for about 1020 IGS sites and compare to the Repro3 solution (Rebischung 2021), i.e., to vertical motion as derived from GPS, GLONASS, and Galileo observations. The first applications of GLWS in hydrological research were described in Gerdener et al. (2022) and Dibi-Anoh et al. (2022).

This paper is organized as follows: First, we introduce the gridded GRACE and GRACE-FO data that we use for assimilation and to which we later compare, as well as the WaterGAP model, forcing data, and the GNSS data used for evaluation. Then, our data assimilation framework is described. Next, we compare GLWS2.0 to GRACE/-FO and WaterGAP TWSA maps in the spatial as well as in the spectral domain, we will provide an uncertainty quantification, and we compare to another global assimilation-derived TWSA data set. Finally, we compare to GNSS observations and provide some conclusions and an outlook including future work.

2 Data and models

2.1 GRACE data

The Gravity Recovery and Climate Experiment (GRACE) satellite mission by the German aerospace center (DLR) and NASA’s Jet Propulsion Laboratory (NASA/JPL) carried a K-band instrument to perform range-rate measurements between two co-orbiting GNSS-tracked platforms. GRACE was launched in 2002; although it was planned as a five-year mission, it ended in 2017 after nearly 15 years of operation. Its successor GRACE-FollowOn (-FO) was launched in 2018 and is still in orbit.

In this study, we use level-2 spherical harmonic geopotential coefficients derived from GRACE/-FO instrument data at TU Graz (ITSG 2018 and operational, Kvas et al. (2019) and Mayer-Gürr et al. (2018)). We chose the ITSG data product because it is accompanied by full error information—i.e., the normal equations—which we deem important to represent the dominating uncertainty patterns of GRACE in the assimilation. Level 2 data are then further converted to TWAS maps following the standard processing steps (Gerdener et al. 2022): certain lower degree coefficients are replaced, a de-stripping procedure (DDK) (Kusche 2007) is applied, the temporal mean from 2003 to 2016 (time span was decided before the GRACE/-FO gap) is removed, and finally, TWAS maps are synthesized following Wahr et al. (1998). Gravity changes related to glacial isostatic adjustment are removed using the A et al. 2013 model.

In the standard GRACE/-FO post-processing at IGG/University of Bonn, we generate TWAS maps at 0.5° resolution, consistent with the spatial resolution of most global hydrology models. However, Schumacher et al. (2016) compared various spatial aggregation scales applied to TWAS maps, when including them into the assimilation, and found that 0.5° and even 1.0° can lead to numerical instabilities in the variance-covariance matrix, which is used to weight observations and model simulations. In more detail, with higher resolution the GRACE/-FO covariance matrices tend to become more ill-posed, which then poses obvious problems within the Ensemble Kalman filter because the (Pseudo)-inverse computations may get numerically unstable or in the worst case can even fail. Therefore, as a compromise, here we transform the spherical harmonic coefficients to a 4° grid prior to assimilation. This can be viewed as an aggregation to numerically regularize the estimation, but it clearly introduces fewer constraints than the use of basin-averaged TWAS.

As mentioned above, in order to represent the dominating error patterns of GRACE, i.e., the latitudinal dependency due to orbital coverage, we derive covariances from the level-2, fully populated normal equations, and propagate them through our post-processing until we arrive at a fully populated level-3 (TWAS grid) variance-covariance matrix. We note that after DDK3 filtering and aggregation to a 4° grid, correlations between TWAS in directly neighboring 4° grid cells can be significantly high, for example, 0.85 in North–South direction close to the equator and 0.44 in East-West direction in a latitude of approximately 45°. However, we found that remaining spatial error correlations for 4° TWAS maps can still induce numerical instability during the update step of the assimilation. This is why we decided to suppress any spatial correlation at the grid level in the representation of GRACE errors within the assimilation. In other words, when vector \( y \) represents observed TWAS on \( n \) grid points, the GRACE uncertainties are represented in the DA via
with the 1-sigmas $\sigma_y$ being, in first approximation, dependent on latitude. Nonetheless, it is important to understand that neglecting the correlations between spherical harmonic coefficients from the very beginning (i.e., propagating a diagonal variance matrix from level-2 coefficients to level-3 TWSA grids) would already lead to incorrect 1-sigmas for the grid TWSA values, which our procedure, in contrast, retains correctly. We stress that suppressing spatial correlation in assimilation frameworks is a common procedure, and what we describe above can be viewed as a form of observation localization (Houtekamer and Mitchell 1998). This is in line with simulation experiments by Schumacher et al. (2016) who found that suppressing spatial correlations, within a DA framework very similar to ours, did not negatively affect results on average.

Some studies have suggested assessing GRACE TWSA errors by comparing different GRACE solutions, either derived from level-2 data or from mascon processing. We refrain from this approach here since (1) these differences may not reflect real uncertainty since all processing centers use similar analysis schemes and the same input data, and (2) the GRACE uncertainties that we infer via rigorous error propagation may lack absolute calibration, they are likely more realistic in terms of spatial patterns. We also stress that for the assimilation here we could in fact ingest any GRACE solution together with any reasonable error covariance, since it is not straightforward to assume different GRACE solutions have vastly different spatial error patterns.

In the subsequent sections, we will compare the GRACE/-FO observation with the WaterGAP model simulations and new GLWS data set in the spatial and spectral domain. The spherical representation can be derived from the gridded TWSA (global land area $\mathcal{L}$, Load Love number $k_n'$, TWSA in equivalent water height $\Delta h$):

$$
\begin{align*}
\mathbf{C}_{yy} &= \begin{bmatrix}
\sigma_{y1}^2 & 0 & 0 & \cdots & 0 \\
0 & \sigma_{y2}^2 & 0 & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & 0 \\
0 & 0 & 0 & \cdots & \sigma_{yn}^2
\end{bmatrix} \\
&= \begin{bmatrix}
\sigma_{n1}^2 & 0 & 0 & \cdots & 0 \\
0 & \sigma_{n2}^2 & 0 & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & 0 \\
0 & 0 & 0 & \cdots & \sigma_{nn}
\end{bmatrix}
\end{align*}
$$

(1)

$$
\begin{align*}
c_{nm} &= \frac{3}{\rho_e} \frac{1 + k_n'}{2n + 1} \int_{\mathcal{L}} C_{nm} \Delta h \, d\omega \\
s_{nm} &= \frac{3}{\rho_e} \frac{1 + k_n'}{2n + 1} \int_{\mathcal{L}} S_{nm} \Delta h \, d\omega.
\end{align*}
$$

(2)

It is clear that the 'GRACE/-FO TWSA harmonics' derived in this way do not necessarily resemble the GRACE solutions, as they represent (DDK-) filtered solutions masked to the continents only and not including contributions from Greenland and Antarctica ice sheets.

### 2.2 WaterGAP global hydrology model and precipitation data

In this study, we make use of the WaterGAP global hydrology model (Müller Schmied 2021), version 2.2e. WaterGAP simulates water fluxes and storages on an 0.5° grid by solving vertical and horizontal water balance equations, representing ten different storage types in total: local lakes, global lakes, local wetland, global wetland, river, reservoirs, groundwater, soil moisture, snow, and canopy storage. Unlike many land surface models, WaterGAP accounts for anthropogenic water use for irrigation, livestock farming, manufacturing, domestic use, and thermal power plant cooling. In its standard configuration, the model is calibrated against long-term annual river discharge. The current version implements, as a major update with respect to version 2.2d, an improved algorithm for surface and groundwater abstraction.

Hydrological models generally partition precipitation (observed or reanalysis, i.e., simulated but constrained by observations) into a variety of fluxes, some of which, like evapotranspiration, are again largely controlled by observations. These forcing data are, for our runs, derived from the homogenized reanalysis GSWP3-W5E5 data set and they include precipitation, temperature, long-wave radiation, and shortwave radiation (Lange et al. 2022).

For data assimilation, we need to represent the uncertainty of the model states, i.e., for ten storage compartments on the global land, represented by the 0.5°-grid. In the EnKF, this is achieved via an ensemble that accounts for perturbations of the initial model states, of the forcing data, and of 22 WaterGAP model parameters as described in Döll et al. (2023, in review).

While WaterGAP appears as a fairly complete model in terms of processes that are represented conceptually, it is clear that certain mismatches must be expected when confronted with GRACE/-FO data. For example, glaciers are not represented as storage in the model and real glacier mass imbalance observed by GRACE/-FO will lead to mismatch and potentially updating other storages in the model (e.g., snow storage or groundwater). Likewise, anthropogenic water use is simulated via sectors models, but these rely on external data, e.g., for irrigation which will be inevitably incomplete (e.g., Müller Schmied 2021). As a result, we expect that in heavily engineered regions the model deviates from GRACE, and TWSA in GLWS should be pulled toward what GRACE observes. Döll et al. (2023, in review) show for the Mississippi river basin that assimilating TWSA has indeed some positive effect also on streamflow, especially in areas characterized by many small wetlands, which are poorly represented within WaterGAP. Other potential sources of mismatch are reservoirs (WaterGAP mimics reservoir operations in a very simplified form) and rivers...
(WaterGAP does consider river storage, but the river network topology itself is greatly simplified).

2.3 GNSS validation data

Even twenty years after the launch of GRACE, an independent assessment of the systematic and random uncertainty of satellite-derived land TWSA is challenging, due to the as-yet unprecedented nature of space-gravimetric data. Hydrological models are not sufficiently accurate, due to missing process representation but also uncertain forcing data, with spatially and temporally heterogeneous quality. ‘Test signals’ have been frequently proposed, e.g., the low TWSA variability over the Sahara regions, and the TWSA signal for large lakes and reservoirs where water levels, bathymetry, and area change can be monitored independently and volume changes thus can be derived straightforwardly; yet problems with this approach are suspected with groundwater flow in deep aquifers or reservoir seepage due to fractured geology. Measurements with superconducting gravimeters have been proposed, e.g., the low TWSA variability over the Sahara regions, and the TWSA signal for large lakes and reservoirs where water levels, bathymetry, and area change can be monitored independently and volume changes thus can be derived straightforwardly; yet problems with this approach are suspected with groundwater flow in deep aquifers or reservoir seepage due to fractured geology. Measurements with superconducting gravimeters have been suggested for comparison to, and validation of, GRACE-derived gravity variations (Crossley et al. 2012), but they are affected by local water table variations and other environmental factors, and their global coverage is quite limited.

Therefore, in this study, we decided to compare GRACE/-FO and GLWS2.0 TWSA gridded data to vertical loading derived from hundreds of globally distributed GNSS sites, using methods that were described in (Blewitt and Clarke 2003; Wu et al. 2003; Kusche and Schrama 2005), and applied to evaluate GRACE solutions as in Chen et al. (2018). GNSS loading evaluations rely on the validity of an underlying Earth model, which is usually assumed as radial-symmetric and purely elastic. Real poroelastic or thermoelastic deformations are difficult to separate from observations and may mask elastic loading in certain geological settings (Carlson et al. 2020), and the varying thickness of the lithosphere below topography and ocean is not taken into account. GNSS data are also affected by technique errors, however, with continuous reprocessing efforts the impact of these errors appears to be greatly reduced, and we believe the sheer number of displacement time series will eventually help to overcome the mentioned problems.

We use 1020 vertical displacement time series that were provided by the IGS as part of the Repro3 reprocessing (Rebischung 2021). Repro3 includes data from the GPS, GLONASS, and Galileo constellations and it was generated as the GNSS contribution to the recent realization of the International Terrestrial Reference Frame (ITRF2020) (Altamimi et al. 2021). To be consistent with the TWSA maps, we removed outliers using the three times Interquartile Range rule. We corrected jumps using epochs available in various GNSS databases and updated these in a manual procedure. In addition, we removed atmospheric and oceanic loading effects from the series using predictions made available by the Earth System and Modelling group at GeoForschungsZentrum (GFZ).

All time series (GNSS and predicted from GRACE/-FO, GLWS2.0, and WaterGAP) were detrended before comparison, in order to eliminate potentially incorrectly modeled glacial isostatic adjustment or local tectonic or anthropogenic subsidence and uplift. Finally, residual time series (i.e., explained variability) are split into short-term, seasonal, and long-term bands.

3 Data assimilation framework

The global land water storage data set version 2.0 (GLWS2.0) is produced by assimilating gridded GRACE and GRACE-FO-derived monthly TWSA maps into the WaterGAP global hydrological model, via an Ensemble Kalman Filter (EnKF) (Evensen 2003) approach as implemented in the Parallel Data Assimilation Framework (PDAF) (Nerger and Hiller 2013). The resulting data set represents thus an optimal synthesis of GRACE/-FO data and all data sets that went into the hydrological model. In other words, this synthesis seeks to fit observed TWSA grids (aggregated to 4°) within error ranges, and at the same time, it solves the horizontal and vertical water balances as represented in the hydrological model (i.e., on a 0.5° grid), again within error ranges.

To this end, the uncertainty of the hydrological model simulation is represented via an n-member ensemble, where we take into account the uncertainty of forcing data as well as the uncertainty of some model calibration parameters. As a result, when no GRACE or GRACE-FO data are available, the GLWS data set represents the mean of an ensemble where each member is dynamically consistent with the model. It is important to recognize that this mean depends on the ensemble creation and thus will differ from published WaterGAP standard runs, even if there are no GRACE data within a particular month. It is also important to understand that the assimilation-derived GLWS TWSA data set does not represent a simple downscaling of the GRACE data, i.e., spatial smoothing of GLWS will not necessarily correspond to GRACE/-FO TWSA.

The monthly level-3 GLWS data represent the total water storage anomaly on 0.5° grids; they are publicly available now from 01/2003 to 12/2019 and will be updated in the future. Additionally, the standard deviation of the ensemble members of GLWS TWSA at the grid cell level is provided, derived from the EnKF ensemble referred against the ensemble mean. The main updates with respect to GLWS1.0 (first version of the GRACE/-FO assimilation into WaterGAP, which was distributed to a limited audience), were the use of an updated version of WaterGAP as described above, as well as minor bug fixes in the assimilation. Model initialization,
spin-up period, and other details are described in Table 1. No model parameters were actually calibrated during this run.

In the EnKF, it is common to adjust the weighting between model prediction and data by 'inflating' the ensemble spread, either based on empirical considerations or on concepts very similar to variance component estimation. However, the inflation factor (the inverse of the 'forgetting' factor in Table 1) was set to 1.0 for GLWS2.0., i.e., no weighting adjustment was applied (see Sect. 4.4).

In order to simulate the combined effect of model structural errors and parameter uncertainty, for each of about 30 WaterGAP model parameters we define two alternative uncertainty bands, a 'small' (optimistic) and a 'large' (pessimistic) range. For each range, minimum and maximum plausible parameter values were defined based on expert knowledge (e.g., Schumacher et al. 2018; Werth et al. 2009; Hosseini-Moghari et al. 2020), and we then assign triangular probability distributions to multiplier (i.e., scaling) parameters and uniform distributions to all other parameters. This allows us then to generate ensemble runs with model parameter perturbations. In this study, we used the small range.

In the EnKF, an ensemble of model simulations is run in parallel to represent simulation uncertainty, with the model states collected in \( X \). In a nutshell, once TWSA observations are assimilated, the entire ensemble of model states is updated to \( \hat{X} \) following

\[
\hat{X} = X + K(Y - H(X))
\]

where each column of \( X \) contains all model states for a single ensemble member, i.e., the entries for all ten water storages in all \( x \) grid cells. Since we work with 32 ensemble members, \( X \) contains \( n_e = 32 \) columns. In Eq. (3), \( K = WW' + GG' \) is the gain matrix whose task is to weight model and observation errors, which we both represent through ensembles (i.e., covariance matrices approximated through \( WW' \) for simulated TWSA and \( GG' \) for the observed TWSA). \( Y \) contains an ensemble of aggregated TWSA observations where perturbations have been sampled from the error variance maps of Eq. (1), and \( H(\cdot) \) represents the (in our case linear, i.e., \( H(X) = HX \)) operator that maps the model states to aggregated observations of total water storage variability (Eicker et al. 2014).

The ensemble mean, for all ten storages in the model, is then provided by

\[
\hat{x} = \frac{1}{n_e} \hat{X} 1
\]

where \( 1 \) is a vector containing 1 in every entry. Then, the TWSA at the 0.5° model grid follows as

\[
\hat{z} = U\hat{x}
\]

where \( U \) is similar to \( H \) except that it accumulates the ten model storages to total storage anomaly TWSA. Meanwhile, we remove the temporal mean value (derived from the open loop run) from the simulated states to refer to simulated anomalies of TWSA on the 0.5° model grid. Here, we again choose the period for the temporal mean similar to the temporal mean of the GRACE/-FO observation (2003–2016) because it was decided before the gap. Similarly, the ensemble representation of GLWS TWSA at the model grid is

\[
\hat{Z} = U\hat{X}.
\]

From Eq. (6), it is straightforward to derive GLWS 1-sigma standard deviations or other error measures, including spatial correlation metrics. We note, however, that the ensemble median (for TWSA grids, \( \hat{z} = U\hat{x} \) where the tilde stands for the median) is often regarded as being more representative for the variability that individual ensemble members display. The ensemble median is considered in this study when we look at extremes, i.e., total water storage anomalies during drought and flood conditions.

With GLWS2.0, we can provide in addition three more data sets that we term 'level-2' accompanied with a 1-sigma characterization; i.e. surface storage anomalies (sum of wetlands, lakes, man-made reservoirs, and river storages), groundwater storage anomalies, and soil moisture anomalies. These are all derived via Eqs. (5) and (6) with the respective linear operators \( U \).

As discussed above, our GRACE/-FO error model seeks to represent coverage and orbital geometry effects (latitudinal growth of errors and anisotropic correlation) as well as the temporally varying data quality, inevitably limited by the quality of the error models used in the level-2 processing. This is complemented by an error model for the hydrological
simulations that accounts for forcing data uncertainty (which depends, e.g., on season and topography) and model parameter errors (which depend strongly on climate regions, soil types, and other geographical factors). As a result, the GLWS error model is expected to reflect all of these patterns and may be thus difficult to interpret. A caveat is, however, that several conceptual equations in the modeled water balances (e.g., inequality constraints whenever a variable cannot be negative, wetland flooding once a threshold storage is reached) can lead to multimodal probability distributions which are possibly not well represented through the limited-size ensemble, and for which the EnKF is known to be not optimal. In lieu of other feasible approaches (e.g., numerical costs seem prohibitive for a particle filter in our application), we accept this limitation.

4 Characterization of the GLWS2.0 global land 0.5° total water storage anomaly data set

4.1 Comparing GLWS2.0 to GRACE data and to WaterGAP model simulations

In this section, we will first compare the total water storage anomaly grids derived by accumulating over all storages in the assimilation product (and removing the temporal mean) to the original GRACE data (at 0.5° resolution), as well as the WaterGAP simulations. We will first look at linear trends, the entire signal variability expressed as RMS, and the seasonal cycle, and then add a comparison in terms of large-scale surplus and deficit water storage averaged over latitudinal bands.

Figure 1 shows the left column the linear trends in total water storage anomalies, computed for years 2003–2019 from the WaterGAP model simulations (top), the original GRACE/-FO-derived TWSA data (center), and the new GLWS2.0 data set (bottom). Root mean square (RMS) maps in the right column display the total variability of TWSA in the model, GRACE/-FO, and GLWS2.0 data set. In the GRACE data, we identify large negative trends i.e. increasing water deficit with −15 mm/a or more, e.g., in the Parana-Sao Francisco-Atlantico river basins in South America, the Caspian Sea region, in parts of Northern India and in glacier regions such as Alaska, Patagonia, and Northern Canada. This is not surprising, as in previous studies (e.g., Gardner et al. 2011; Luthcke et al. 2013; Getirana 2016; Panda and Wahr 2016; Chen et al. 2017; Rodell et al. 2018); these regions were identified to have been affected by severe droughts, decreasing surface water levels, groundwater depletion and melting of glaciers. In contrast, large positive trends (i.e., wetting) can be found in Eastern Canada, West Africa, the Zambesi basin and African Great Lake region, and in the La Plata basin in South America, which were all associated with regions that are either experiencing increasing precipitation and/or transform from dry to wet conditions, e.g., recovery after a drought during the observation period takes place (e.g., Rodell et al. 2018).
As has been noticed in many studies (e.g., Scanlon et al. 2018, and references therein), the WaterGAP global hydrology model generally simulates much lower TWSA long-term trends as compared to GRACE/-FO. However, we notice that there are many similarities in the signs, i.e., drying vs. wetting, e.g., in the Amazon basin, West Africa, Northwest America, and other regions. In comparison with GRACE, the model displays the trend maps with much higher spatial detail due to its intrinsically higher spatial resolution, which in turn is due to the higher spatial resolution of its forcing data. We note that drying trends in our model simulations with WaterGAP2.2e appear less pronounced as in Müller Schmied (2021) (cf. Fig. 9 for basin-averaged total water storage trends in the 2.2d standard version).

Linear trends in GLWS seem to confirm the anticipated advantage of assimilating GRACE data into the model: The effective spatial resolution of the assimilated data set appears visually much higher than the GRACE data, while the large-scale patterns from GRACE/-FO seem inherited for many regions. We thus hypothesize that we obtain a more detailed picture of the distribution of negative and positive trends. For example, a pattern of large negative trend at the Southern tip of Patagonia that one identifies in the GRACE data appears more localized in the actual glacier location in the GLWS data. At the same time, we also identify regions, where GLWS does not inherit large-scale GRACE/-FO patterns, for example, for parts of the Middle East. This demonstrates that data assimilation does not represent a simple downscaling exercise, but a weighted combination of model prediction and observation. We note that if trends for GLWS would be inferred from the ensemble median instead of the ensemble mean, this would lead to negligible average changes of about 0.02%.

Hydrological models are not expected to represent small trends in the water balance well (Scanlon et al. 2018), while many studies have shown that they compare well with GRACE/-FO data at seasonal and monthly timescales. From the right column of Fig. 1, we notice that WaterGAP seems to display variability at large spatial scales comparable to observed TWSA, with generally fewer signals in arid regions. The model cannot reproduce glacier mass changes (e.g., Alaska glacier chain) as has been pointed out already, but for many larger river basins variability is similar as seen in GRACE (as has been described in countless studies). We suspect that glacier mass loss signals missing in the model simulation will also dominate the differences in RMS. We notice that in GLWS2.0, trends in arid regions are somewhat pulled toward GRACE, and spatial detail is added in particular for large lakes, reservoirs, and riverine regions such as the Amazon, Mississippi, and Zambezi river basins. We do see stronger variability in GLWS2.0 as compared to GRACE/-FO in island regions such as the Indonesian Archipelago, which are difficult to assess from satellite data only due to limited resolution, the need for smoothing, and the ocean mass leakage problem. Li et al. (2020) suggest that GIA models may have quite significant uncertainties in particular for the Hudson Bay region. Any rate error in the A et al. 2013 model which we use for correcting GRACE/-FO...
The global land water storage data set release 2 (GLWS2.0) derived via assimilating GRACE and...

TWASA maps might propagate into water storage rate offsets in GLWS2.0; we therefore suggest extra caution in application in this region.

Next, the dominating annual signals in all three data sets are compared in Fig. 2, with the TWSA amplitude in the left column and the phase (month in which the maximum amplitude occurs) in the right column. Since in most river basins the signal variability is dominated by an annual signal, the left column is indeed very similar to the right column in Fig. 1 (this is not true for large reservoirs where long-term water level changes due to resource management dominate, for example, lake Volta). More interesting is the right column, which reveals that phase mismatches of one or two months between GRACE/-FO data and model simulation are not uncommon. While these are (partly) mitigated through assimilation in some regions (West Africa, Central America, Australia, South Asia), one-month mismatches remain between GLWS2.0 and GRACE/-FO phases such as in Siberia or large parts of North America. We note, however, that GLWS inherits the spatial variability of phase (e.g., regional variations in the onset of the rainfall season) from WaterGAP which is probably real and which cannot be represented in GRACE/-FO data due to coarse effective resolution. For the sake of completeness, we note that if we would determine the GLWS annual amplitudes from the ensemble median instead of the ensemble mean, differences would average to insignificant 0.13%.

We now add the temporal dimension to these comparisons and ask to what extent extreme water storage anomalies are represented in GLWS2.0. To this extent, in the right column of Fig. 3, the Hovmöller diagram for TWSA (i.e., latitude-averaged vs. time) as observed by GRACE/-FO is shown. This figure (we borrow the representation from Tapley et al. (2019)) represents storage anomalies with respect to the mean, trend, and seasonal cycle. The anomalies have been attributed to drought events in the Amazon, southern Africa, and Australia in 2005, exceptional flooding in the Amazon basin and in south Central Africa in 2009, ENSO-related precipitation in Australia, Amazon, and southern Africa in 2011, and droughts across the Amazon, southern Africa including Madagascar, and Australia in 2015/2016. Our satellite TWSA diagram differs slightly from Tapley et al. (2019); this might be due to the use of a different GRACE solution, but we cannot exclude that our method for removal of the seasonal cycle differs from theirs.

All major extreme events that had been identified in the GRACE data appear present in GLWS2.0 (as noted earlier, we show the ensemble median here). We notice that, strikingly, the La Nina-related wet period 2007–2012 (Sorí et al. 2021) appears more pronounced and longer-lasting in GLWS2.0 for southern latitudes $-10^\circ$ to $-20^\circ$ (where anomalies in the simulation seem to fit better to GRACE), whereas the unprecedented 2016 drought in South America (Jiménez-Muñoz et al. 2016; Erfanian et al. 2017) appears less pronounced south of about $-10^\circ$ latitude. We find that in general tropical drought events become visible much more spatially localized as compared to GRACE, which is expected and generally corresponds to precipitation-derived drought indicator maps. GLWS2.0 inherits many drought and flood events from the model simulation, with mostly wet conditions more pronounced. As a counter-example, the 2010 drought-related water storage deficit (e.g., Amazon, Horn of Africa, China, U.S.A., Mexico) extends somewhat further to northern latitudes in GLWS as compared to the simulation. However, it is important to remember that anomalies in the maps refer, by definition, to the individual mean, trend, and annual behavior, which is not necessarily identical across the three data sets. The reader should also be reminded that GRACE-derived TWSA extremes are far from...
being Gaussian distributed, with significant skewness in their distribution (Forootan and Kusche 2012).

### 4.2 Comparing GLWS2.0 to GRACE data and to WaterGAP model in the spectral domain

In this part, we seek to compare the assimilation data set with GRACE/-FO and model data in terms of spherical harmonic representation. With what we have seen in the previous section, it is expected that GLWS inherits spatial resolution from WaterGAP while at large spatial scales; there seems somewhat less variability present as compared to GRACE/-FO (e.g., Figure 2). For a consistent and systematic assessment, we thus map global land GLWS, WaterGAP, and also GRACE/-FO total water storage maps to spherical harmonics.

Figure 4 shows the monthly RMS of the three time series per coefficient. For ease of comparison, we remapped all triangle plots such that colors directly signify the contribution to geoid variability, in units of mm. We notice that the DA indeed effectively removes much high-frequency content from WaterGAP, beyond about degree 25, but it retains significantly more variability as compared to GRACE/-FO beyond about degree 35.

The same can be observed when the triangle plots are collapsed to signal degree variance curves expressed in geoid variability, as in Fig. 5. We notice a striking offset between WaterGAP and GRACE/-FO, with WaterGAP displaying a higher spectral power in particular in the very low degrees. We speculate that this could be caused by much larger signals over Siberia and North America in the hydrological model, as can also be seen in the RMS (Fig. 1). The GLWS data are much closer to GRACE/-FO at low degrees, whereas at around degree 30 it is evidently pulled toward exhibiting more power. Here, GLWS seems to adjust to the same power decay law as seen in WaterGAP. We find a similar behavior in the spectral domain when expressing Figs. 4 and 5 in terms of total water storage anomalies (see supplement S1 and S2).

It is instructive to view how single spherical harmonic coefficients get modified through assimilation. We chose $c_{20}$ to be displayed in Fig. 6; again we remind the reader that the ‘GRACE coefficient’ time series shown here is (1) derived from a GRACE/-FO solution where the original $c_{20}$ has been replaced by an SLR solution, and (2) TWSA had been synthesized on the global land and remapped into the spherical harmonic domain. The WaterGAP simulation dynamic flattening contribution appears with much less interannual variability, yet it shows a larger annual amplitude with less within-year variation. GLWS2.0 $c_{20}$ inherits properties from both data sets; it remains to be understood what is closer to...
Fig. 6 $c_{20}$ spherical harmonic coefficient derived by WaterGAP, GRACE/-FO, and GLWS2.0 (GRACE/-FO assimilated into WaterGAP computed for 2003–2019)

A standard deviation of higher than 1 mm/a is mainly derived in cells where large surface water bodies are present, for example, for river grid cells. We notice that trend uncertainties are particularly low in large arid regions, which is likely due to confidence in GRACE/-FO data but is certainly also related to an underestimation of model representation errors in our WaterGAP ensemble. The relation of trend uncertainties to the trends can be found in the supplement (S3 and S4).

Our GRACE/-FO error model does account for the non-diagonal error structure of the spherical harmonic coefficients due to orbital patterns and measurement principle and thus to the dominating north–south dependence of errors. However, a caveat is that after transforming the spherical harmonics to a spatial grid, the resulting spatial correlations at the grid cell level are suppressed. In other words, one should not use the ensemble to represent spatial correlations between GLWS TWSA across grid point locations.

A second concern must be whether the GRACE/-FO uncertainty, e.g., due to insufficient background modeling in the processing can be adequately represented at all through such propagated errors; however, since all published GRACE/-FO level-2 solutions basically utilize the same processing standards and background models, we do not see another way of representing GRACE/-FO errors more realistic. A third issue could be that the limited ensemble error representation in the EnKF may lead to additional artifacts. When referring to Eqs. (3) and (4), it becomes obvious that the estimate in the EnKF is constructed in the ‘observation space,’ i.e., as a weighted linear combination of the observation ensemble $Y$ (which, in our case, consists of ‘clones’ of a single GRACE/-FO solution perturbed with errors sampled from the covariance matrices provided by ITSG, TU Graz). This points to an interesting analogy with current efforts (e.g. within the COST-G project) to find optimal weights for combining different independent GRACE/-FO solutions, e.g., within a variance component estimation procedure.
4.4 Robustness of the assimilation against analysis choices

Our data assimilation framework allows one to choose certain ‘tuning’ parameters. To understand how robust the GLWS2.0 maps are with respect to these choices, we have first rerun the assimilation while varying the inflation factor (cf. Table 1; inflation increases the ensemble spread, to avoid filter ‘saturation’ and a decreasing impact of observations). However, by increasing this factor by 25% we derive a globally averaged RMS change of only 3% when compared to GLWS2.0. The results suggest here that moderate inflation has little impact when being weighted against the 1-sigma observation error. For the next analysis, we return to our previous settings and we replace the 1-sigma TWSA errors by 2-sigma (3-sigma)—this leads to a change of about 9.6% (11.6%) in RMS globally averaged. These differences are still relatively minor. The situation is similar when we modify the spatial resolution of the input GRACE/-FO TWSA grids for the assimilation (data matrix Y). When moving from 4° to 5° grid cells, we find that the assimilation TWSA differs by 11.8% RMS as compared to GLWS2.0. This is not unexpected since with about 50% bigger input cell area, the coarser grid will be less able to represent real spatial variability of hydrological regimes. On the other hand, the coarser grid contains less observation, and thus, the weighting between model and data is additionally shifted toward the model. Last, if we replace the GRACE/-FO solution with solutions from CSR or GFZ, the RMS globally averaged differs by 11.3% or 11.4% (16.5% or 21.7% prior to assimilation), respectively; thus, the choice of solutions has a slightly lower impact than the choice of grid size.

4.5 Comparisons with other global TWSA data sets

In the following, we compare GLWS2.0 TWSA against another global land storage data set, which was developed by assimilating GRACE/-FO mascon data provided by NASA’s Goddard Space Flight Center (GSFC) into the Catchment Land Surface Model (CLSM) (Koster et al. 2000), a widely used global land surface model on 36 km spatial resolution (Girotto et al. 2016; Felsberg et al. 2021).

Overall, we identify a good agreement between both assimilation data sets (Fig. 8). Important, we find linear trends and annual amplitudes (determined via multi-linear regression as in Gerdener et al. (2022)) in the two assimilations closer to the respective GRACE/-FO products than the model simulations (in both cases the ensemble mean); in other words, integrating GRACE/-FO data into the model simulations by assimilation results in trends and seasonal cycle converging. However, when compared to the CLSM-DA, GLWS2.0 exhibits somewhat lower trends and lower annual amplitudes. For example, for CLSM-DA the annual amplitudes in the Zambezi river basin are up to 200 mm, while with GLWS2.0 we find amplitudes reaching about 100 mm in this basin. This difference is mainly confirmed for humid areas, while also for some arid areas like the Sahelian zone or Australia’s interior minor differences in spatial patterns of trends and amplitudes can be observed.

We suggest that differences in the two DA products are mainly due to differences in data and model representation. Whereas GLWS2.0 assimilates gridded GRACE/-FO TWSA maps derived from (DDK3-) filtered spherical harmonic coefficients, the CLSM-DA ingests GRACE/-FO mascons and applies a rescaling of the observations before assimilation. Moreover, CLSM and WaterGAP models differ in concepts and forcing data. For example, the WaterGAP model has an explicit representation of water stored in river networks and adjacent floodplains, and therefore, for large rivers, we find the river routing pattern imprinted in the GLWS2.0, which is not the case for the CLSM-DA. WaterGAP uses a single soil moisture storage and a groundwater storage compartment per grid cell, whereas in CLSM the soil moisture and groundwater are diagnosed from three prognostic variables that have an assumed within-grid cell spatial distribution. Lastly, CLSM-DA uses MERRA2 forcing data while WaterGAP is forced with GSWP3-W5E5. We suggest that future studies should address the effect of assimilation system design and tuning choices across multiple models, with consistent GRACE/-FO input grids and climate forcing data.

5 Validation with GNSS

As was mentioned above, GNSS time series are hypothesized to sense the redistribution of total water storage through elastic crustal loading, provided that all other relevant processes can be modeled and removed. We use 1020 detrended vertical displacement time series that were provided by the IGS within the Repro3 reprocessing (Rebischung, 2021).

Figure 9 shows a direct comparison of predicted monthly vertical motion time series from GRACE/-FO, GLWS2.0, and WaterGAP toward GNSS in terms of correlation. We split all residual time series, i.e., the unexplained variability, into three temporal bands, following a wavelet decomposition approach (Meyer 1990; Bogusz et al. 2013). After partial reconstruction, the short-term band contains all temporal wavelengths from 2 to 5 months, the seasonal band includes the signal between 5 months and 1.4 years, and the long-term band represents the wavelengths beyond 1.4 years. At this point, it is important to mention again that for the validation with GNSS all time-series were detrended, which means that for the validation positive effects due to linear trends are excluded. GLWS2.0 TWSA maps were transformed to the spectral domain, in order to facilitate the loading computa-
The global land water storage data set release 2 (GLWS2.0) derived via assimilating GRACE and...

Fig. 8 Comparison of the two 0.5° GLWS2.0 and CLSM-DA TWSA data sets in terms of trends [mm per year], annual amplitudes [mm], and phases of the annual amplitudes [months]. The data period is 2003–2019

Fig. 9 Comparison of correlation coefficients for short-term, seasonal, and long-term temporal bands, GNSS vs. GRACE/-FO (d/o 96) (left) and GNSS vs. GLWS2.0 degree and order 96 (right), all 2003 to 2019
tion similar to Eq. (2) with the load Love number involving $h'_n$.

Here we truncated the assimilation TWSA first at degree and order 96, i.e., spectrally consistent with GRACE (see supplement S8 and S10 for truncation to degree and order 120 and 180). For GLWS2.0, we note some improvement at shorter timescales in particular in Europe but also in North America, Southern Australia, and South America as compared to GRACE/-FO. At seasonal timescales, improvement in GLWS2.0 w.r.t. GRACE/-FO data is found most notably over Australia and the US and Canada West coast. At longer than seasonal timescales, we find a noticeable improvement over Western and Central Europe and Australia, whereas it appears that for some stations in South America and Northern Canada the original GRACE/-FO data show a better correlation. Compared to WaterGAP, we identify improvement in GLWS2.0, for example, for the seasonal time scale over Europe and Southern Africa.

All in all (Fig. 10), we find a better agreement in all three temporal bands for the GLWS2.0 total water storage grids as compared to GRACE/-FO. We note this result is not related to the original 0.5° spatial resolution in GLWS2.0. It is evident, however, that all statistics are leaning heavily toward Europe, North America, and Australia due to the high station density in these regions and should not be interpreted as being globally representative. Future studies should focus on removing this geographical bias, e.g., by thinning out the GNSS sites for comparison.

The spectral resolution of degree and order 96 corresponds to about 200 kms spatial wavelength. When we truncate GLWS2.0 spherical harmonic synthesis at degree and order 120 or 180, slight but consistent improvements are observed for, e.g., the long-term temporal bands (red and orange bars in Fig. 10); yet these are minor when compared to the remaining variability between GRACE/-FO and GLWS. For example, for the long-term band we derive 385 stations with a correlation higher than 0.5 when truncating to degree and order 180. In contrast, only 380 stations show correlations higher than 0.5 when we truncate to degree and order 96. While this may seem counter-intuitive, we note that even degree and order 180 correspond to about 110 km spatial resolution and will lack many real local hydrological signals due to missing resolution, but foremost since the elastic crustal loading smooths out effectively such effects. This is also consistent with what Springer et al. (2019) observed in experiments with an 0.08° resolution hydrological model. In summary, comparisons of predicted nonlinear hydrological loading to Repro3 GNSS time series suggest that GLWS seems to fit better at all temporal bands than GRACE/-FO. Given that some noise floor must be expected in these comparisons due to the presence of real non-loading, e.g., tectonic displacements and remaining GNSS technique errors (Chanard et al. 2020), this is not necessarily an expected result. We would like to add that our findings appear robust with respect to the metric that we use for comparing loading simulations with GNSS time series; we have added figures that display RMS reduction in percent to the supplement (see S11, S12 and S13).

6 Conclusion

We describe a new 0.5° global land water storage data set, GLWS2.0, which is derived via assimilating maps of GRACE/-FO TWSA into a hydrological model, WaterGAP. Although GLWS2.0 provides all storages that WaterGAP represents, i.e., soil moisture, various surface water body storages, snow, and groundwater storage, our focus here is on the TWSA derived from GLWS2.0.

We find that GLWS2.0 matches GRACE/-FO TWSA in terms of trend regions and variability, with trends being generally less pronounced, likely inherited from the WaterGAP simulations, and possibly too low. GLWS2.0 displays less variability in arid regions as compared to GRACE/-FO. Much more spatial detail is present in particular in trend regions.
Although GRACE/-FO data and the hydrological model differ, GLWS2.0 shows many similarities with the CLSM-DA assimilation. In the spectral domain, GLWS2.0 exhibits more power as compared to GRACE/-FO from about degrees 25 onwards, reaching nearly the same slope of the spectrum as the WaterGAP simulation. This is, to our knowledge, the first assessment of the effect of GRACE/-FO assimilation in the spectral domain. In terms of extreme water storage anomalies, GLWS2.0 reveals much more spatial detail for droughts and flooding events that can be observed but not well resolved spatially, with GRACE/-FO. Beyond this—in fact expected—result, the assimilation suggests that precipitation related water storage surplus was temporally more sustained than in GRACE/-FO solutions, in particular for the southern hemisphere.

We compared GLWS2.0 predicted nonlinear vertical station motions to coordinate time series derived within the Repro3 contribution to ITRF2020. GLWS2.0 seems to fit better to independent GNSS observations at all temporal bands (except the linear trends which need to be removed for this kind of comparison), even when truncated to the same spectral degree as GRACE/-FO.

We suggest ascribing this to the contribution of the hydrological modeling at shorter spatial scales (beyond about degree and order 30 in the spectral domain), but we feel more research is required here. If true it would suggest that merging the GRACE/-FO data with a hydrological model would generally improve TWSA representation. It would also support the idea to integrate GNSS-derived loading time series in assimilation studies. We anticipate that more validation experiments and a generally improved understanding of the loading signals observed in GNSS are required here.

We believe that GLWS2.0 provides a useful data set of total water storage with true 0.5° resolution, in the sense that at shorter spatial scales gradually the information from the hydrological model start to dominate. We suggest that GLWS2.0 could be used for geodetic and geophysical applications like loading predictions for geometric space techniques, hydrological angular momentum computations, or inversions for rheological properties. Companion studies will evaluate its applications in hydrology and climate research.

What is missing so far is an assessment of the daily TWSA as represented in the assimilation. Unlike, e.g., in Girootto et al. (2016), TWSA is updated in our framework at the end of a month, and a meaningful interpretation of daily model states may require that we implement a temporal smoothing of the update (e.g., three times per month as in Zaitchik et al. (2008)). Future versions of GLWS will also integrate other observables such as streamflow or snow cover fraction, primarily with the intention to improve the disaggregation of storages and the realism of the simulation of these variables. It is possible but not guaranteed that this will improve TWSA representation. Forthcoming work will, in addition, address the estimation of model parameters within the EnKF, which is expected to improve model prediction skills, e.g., during GRACE/-FO data gaps, and the skill of the assimilation (and thus future versions of GLWS) when we use different filter algorithms, for example, local filters. We will also work toward better understanding the uncertainty of GLWS2.0 and improving its representation within the ensemble.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00190-023-01763-9.

Acknowledgements We acknowledge Olga Gassen (Engels) for all her effort and contribution to the data assimilation framework and Hannes Müller Schmied and Sebastian Ackermann for providing the WaterGAP model and improving its interfaces for the data assimilation framework. We thank Makan Karegar, Christian Mielke, Bell Kuei-Hua Hsu, Tonie van Dam, Anita Thea Saraswati, Laura Jensen, and Annette Eicker for providing feedback that helped us to improve the GLWS processing. We are grateful to Gabrielle De Lannoy and Manuela Girootto for providing CLSM-DA data and for the discussions. AK is supported by the National Science Centre, Poland, grant no. UMO-2021/41/B/ST10/01458. Finally, we appreciate helpful comments by the associate editor Frank Lemoine and three anonymous reviewers.

Author Contributions JK conceived the presented research with support from PD, HG, JK, and AK planned the experiments. HG and KS carried out the implementation and HG performed the computations. AK contributed the GNSS validations. JK and HG wrote the manuscript with input from all authors.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the German Federal Ministry of Education and Research (BMBF) in the framework of the ‘GlobeDrought’ project (grant number 02WGR1457A) and the German Research Foundation (DFG) in the framework of the Research Unit ‘Understanding the global freshwater system by combining geodetic and remote sensing information with modeling using a calibration/data assimilation approach’ (GlobalCDA, Grant Number KU1207/26-1) and the SFB 1502/1-2022—Grant Number 450058266.

Data availability statement The GLWS2.0 data described in this contribution can be downloaded at PANGAEA: Gerdener, Helena; Schulze, Kerstin; Kusche, Jürgen (2023): GLWS 2.0: A global product that provides total water storage anomalies, groundwater, soil moisture, and surface water with a spatial resolution of 0.5° from 2003 to 2019. PANGAEA, https://doi.org/10.1594/PANGAEA.954742.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence.
References

A G, Wahr J, Zhong S (2013) Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: an application to Glacial Isostatic Adjustment in Antarctica and Canada. Geophys J Int 192(2):557–572. https://doi.org/10.1093/gji/ggs030

A G, Velicogna I, Kimball JS, Du J, Kim Y, Colliander A, Njoku E, Björnsson E, (2017) Satellite-observed changes in vegetation sensitivities to surface soil moisture and total water storage variations since the 2011 Texas drought. Environ Res Lett 12(5):054006. https://doi.org/10.1088/1748-9326/aae965

Altamimi Z, Reuschung P, Métivier L, Collilieux X, KC (2021) Highlights of IGS contribution to ITRF2020. https://files.igs.org/pub/resource/pubs/workshop/2021/03-Altamimi.pdf

Blewitt G, Clarke P (2003) Inversion of Earth’s changing shape to weigh solid Earth. J Geophys Res: Solid Earth. https://doi.org/10.1029/2002JB002290

Bogusz J, Klos A, Kosek W (2013) Wavelet decomposition in the earth’s gravity field investigation. Acta Geodætica et Geophysis 10(1):47–59

Broeke T, Riva R, Simons G, Govers R, Vermeersen B (2015) Postseismic GRACE and GPS observations indicate a rheology contrast above and below the Sumatra slab. J Geophys Res: Solid Earth 120(7):5343–5361. https://doi.org/10.1002/2015JB011951

Carlson G, Shirzaei M, Werth S, Zhai G, Ojha C (2020) Seasonal and long-term groundwater unloading in the Central Valley modifies crustal stress. J Geophys Res: Solid Earth 125(1):e2019JB018490. https://doi.org/10.1029/2019JB018490

Chanard K, Fleitout L, Calais E, Reuschung P, Avouac J-P (2018) Toward a global horizontal and vertical elastic load deformation model derived from GRACE and GNSS station position time series. J Geophys Res: Solid Earth 123(4):3225–3237. https://doi.org/10.1002/2017JB015245

Chanard K, Métois M, Reuschung P, Avouac J-P (2020) A warning against over-interpretation of seasonal signals measured by the Global Navigation Satellite System. Nat Commun 11(1):1–4. https://doi.org/10.1038/s41467-020-15106-7

Chen J, Wilson C, Tapley B, Save H, Cretaux J-P (2017) Long-term and seasonal Caspian Sea level change from satellite gravity and altimeter measurements. J Geophys Res: Solid Earth 122(3):2274–2290

Chen Q, Poropat Z, Zhang L, Zhwun H, Weigelt M, Van Dam T (2018) Validation of the EGSiEM GRACE gravity fields using GNSS coordinate timeseries and in-situ ocean bottom pressure records. Remote Sens. https://doi.org/10.3390/rs10121976

Crossley D, de Linage C, Hinderer J, Boy J-P, Famiglietti J (2012) A comparison of the gravity field over Central Europe from superconducting gravimeters, GRACE and global hydrodynamical models, using EOF analysis. Geophys J Int 189(2):877–897. https://doi.org/10.1111/j.1365-246X.2012.05404.x

Dibi-Anth P, Kone M, Gerdener H, Kusche J, N'Da C (2022) Hydrometeorological extreme events in west Africa: droughts. Surv Geophys. https://doi.org/10.1007/s10712-022-09748-7

Döll P, Hasan HMM, Schulze K, Gerdener H, Börger L, Shadkam S, Ackermann H, Hosseini-Moghari SM, Müller Schmied H, Güntner A, Kusche J (2023) Multi-variable parameter estimation for a global hydrological model: comparison and evaluation of three ensemble-based calibration methods for the Mississippi River Basin. Hydrol Earth Syst Sci Discus [preprint] (in review). https://doi.org/10.5194/hess-2023-18

Eicker A, Schumacher M, Kusche J, Döll P, Müller Schmied H (2014) Calibration/data assimilation approach for integrating GRACE data into the watergap global hydrology model (WGHM) using an ensemble Kalman filter: first results. Surv Geophys 35(6):1285–1309

Erfanian A, Wang G, Fomemko L (2017) Unprecedented drought over tropical South America in 2016: significantly under-predicted by tropical SST. Sci Rep. https://doi.org/10.1038/s41598-017-05373-2

Evensen G (2003) The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn 53(4):343–367. https://doi.org/10.1007/s10236-003-0036-9

Felsberg A, Lannoy GJMD, Girotto M, Poesen J, Reichle RH, Stanley T (2021) Global soil water estimates as landslide predictor: the effectiveness of SMOS, SMAP, and GRACE observations, land surface simulations, and data assimilation. J Hydrometeorol 22(5):1065–1084. https://doi.org/10.1175/JHM-D-20-0228.1

Forootan E, Kusche J (2012) Separation of global time-variable gravity signals into maximally independent components. J Geodesy 86(7):477–497. https://doi.org/10.1007/s00190-011-0532-5

Gardner AS, Moholdt G, Wouters B, Wolken GJ, Burgess DO, Sharp MJ, Cogley JG, Braun C, Labine C (2011) Sharply increased mass loss from glaciers and ice caps in the Canadian archipelago. Nature 473(7347):357–360

Gerdener H, Engels O, Kusche J (2020) A framework for deriving drought indicators from the Gravity Recovery and Climate Experiment (GRACE). Hydrol Earth Syst Sci 24(1):227–248. https://doi.org/10.5194/hess-24-227-2020

Gerdener H, Kusche J, Schulze K, Ghazaryan G, Dubovyk O (2022) Revising precipitation–water storage–vegetation signatures with grace-based data assimilation. J Hydrol 612(A):128096

Getirana A (2016) Extreme water deficit in brazil detected from space. J Hydrometeorol 17(2):591–599

Girotto M, Lannoy GJMD, Reichle RH, Rodell M (2016) Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model. Water Resour Res 52(5):4164–4183. https://doi.org/10.1002/2015WR018417

Han S-C, Gobhadi-Far K, Yeo I-Y, McCullough CM, Lee E, Sauber J (2021) GRACE Follow-On revealed Bangladesh was flooded early in the 2020 monsoon season due to precursor soil saturation. Proc Natl Acad Sci 118(47):e2109086118. https://doi.org/10.1073/pnas.2109086118

Hosseini-Moghari S-M, Araghinejad S, Tourian MJ, Ebrahimi K, Döll P (2020) Quantifying the impacts of human water use and climate variations on recent drying of lake Urmia Basin: the value of different sets of spaceborne and in situ data for calibrating a global hydrological model. Hydrol Earth Syst Sci 24(4):1939–1956

Houtekamer PL, Mitchell HL (1998) Data assimilation using an ensemble Kalman Filter technique. Mon Weather Rev 126(3):796–811. https://doi.org/10.1175/1520-0493(1998)126<0796:DAUES>2.0.CO;2

Humphreys V, Gudmundsson L (2019) GRACE-REC: a reconstruction of climate-driven water storage changes over the last century. Earth Syst Sci Data 11(3):1153–1170. https://doi.org/10.5194/essd-11-1153-2019

Jiménez-Muñoz JC, Mattar C, Barichivich J, Santamaría-Artigas A, Takahashi K, Malhi Y, Sobrino JA, Schrier GVD (2016) Record-breaking warming and extreme drought in the Amazon rainforest during the course of El Niño 2015–2016. Sci Rep 6(1):1–7. https://doi.org/10.1038/srep33130

Jin S, Chambers DP, Tapley BD (2010) Hydrological and oceanic effects on polar motion from GRACE and models. J Geophys Res: Solid Earth. https://doi.org/10.1029/2009JB006635
Khaki M, Hoteit I, Kuhn M, Awange J, Forootan E, Dijk AJMV, Schumacher M, Pattiaratchi C (2017) Assessing sequential data assimilation techniques for integrating GRACE data into a hydrological model. Adv Water Resour 107:301–316. https://doi.org/10.1016/j.adwres.2017.07.001

Koster RD, Suarez MJ, Ducharne A, Stieglitz M, Kumar P (2000) A catchment-based approach to modeling land surface processes in a general circulation model: 1 Model structure. J Geophys Res: Atmos 105(D20):24809–24822. https://doi.org/10.1029/2000JD900327

Kusche J (2007) Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models. J Geodesy 81(1):733–749. https://doi.org/10.1007/s00190-007-0143-3

Kusche J, Schrama EJO (2005) Surface mass redistribution inversion from global GPS deformation and Gravity Recovery and Climate Experiment (GRACE) gravity data. Solid Earth 1:3. https://doi.org/10.5194/se-1-1037-2011

Kvas A, Behzadpour S, Ellmer M, Klinger B, Strasser S, Zehentner N, Mayer-Gürr T, Behzadpour S, Ellmer M, Klinger B, Strasser S, Zehentner N, Mayer-Gürr T (2019) ITSG-Grace2018: overview and evaluation of a new GRACE-only gravity field time series. J Geophys Res: Solid Earth 124(8):9332–9344

Lange S, Mengel M, Treu S, Büchner M (2022) Isimip3a atmospheric climate input data. https://doi.org/10.48364/ISIMIP982724

Li T, Wu P, Wang H, Steffen H, Khan NS, Engelhart SE, Vacchi M, Shaw TA, Peltier WR, Horton BP (2020) Uncertainties of glacial isostatic adjustment model predictions in north America associated with 3d structure. Geophys Res Lett 47(10):e2020GL087944

Li F, Kusche J, Chao N, Wang Z, Löcher A (2021) Long-term (1979-Present) total water storage anomalies over the global land derived by reconstructing GRACE data. Geophys Res Lett 48(8):e2021GL093492. https://doi.org/10.1029/2021GL093492

Li B, Rodell M, Kumar S, Beaudong HK, Getirana A, Zaitchik BF, de Goncalves LG, Cossetin C, Bhanja S, Mukherjee A, Tian S (2019) Global GRACE data assimilation for groundwater and drought monitoring: advances and challenges. Water Resour Res 55(9):7564–7586. https://doi.org/10.1029/2019WR024618

Luthcke SB, Sabaka T, Loomis B, Arendt A, McCarthy J, Camp J (2013) Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution. J Glaciol 59(216):613–631

Mayer-Gürr T, Behzadpour S, Ellmer M, Kvas A, Klinger B, Strasser S, Zehentner N (2018) ITSG-Grace2018 - Monthly, Daily and Static Gravity Field Solutions from GRACE. GFZ Data Services. https://doi.org/10.5880/ICGEM.2018.003

Mémin A, Boy J-P, Santamaría-Gómez A (2020) Correcting GPS measurements for non-tidal loading. GPS Solut 24(2):1–13. https://doi.org/10.1007/s10291-020-0959-3

Meyer Y (1990) Ondelettes et Opérateurs. Hermann, Paris

Müller Schmied H, Cáceres D, Eiserer S, Flörke M, Herbert C, Niemann C, Peiris TA, Popat E, Portmann FT, Reinecke R, Schumacher M, Shadkam S, Telteu C-E, Trautmann T, Döll P, Beek RV, Sandberg Sørensen L (2020) Improving drought simulations within the Murray-darling basin by combined calibration/assimilation of GRACE satellite data. Proc Natl Acad Sci 117(6):E1080–E1089. https://doi.org/10.1073/pnas.1913730117

Schumacher M, Forootan E, van Dijk AJJM, Müller-Schmied H, Crosbie RS, Kusche J, Döll P (2018) Improving drought simulations within the Murray-darling basin by combined calibration/assimilation of GRACE data. Proc Natl Acad Sci 115(6):E1080–E1089. https://doi.org/10.1073/pnas.1704665115

Schumacher M, Forootan E, van Dijk AJJM, Müller-Schmied H, Crosbie RS, Kusche J, Döll P (2016) A systematic impact assessment of GRACE error correlation on data assimilation in hydrological models. J Geodesy 90(6):537–559. https://doi.org/10.1007/s00190-016-0892-y

Schumacher M, Forootan E, van Dijk AI, Schmied HM, Crosbie RS, Kusche J, Döll P (2018) Improving drought simulations within the Murray-darling basin by combined calibration/assimilation of GRACE data into the watergap global hydrology model. Remote Sens Environ 204:212–228

Springer A, Karegar MA, Kusche J, Keune J, Kurtz W, Kollet S (2019) Evidence of daily hydrological loading in GPS time series over Ray RD, Luthcke SB, Dam TV (2013) Monthly crustal loading corrections for satellite altimetry. J Atmos Ocean Technol 30(5):999–1005. https://doi.org/10.1175/JTECH-D-12-00152.1

Razeghi M, Han S-C, McClusky S, Sauber J (2019) A joint analysis of GPS displacement and geosteric data for simultaneous estimation of geocenter motion and gravitational field. J Geophys Res: Solid Earth 124(11):12241–12263

Reager JT, Thomas BF, Famiglietti JS (2014) River basin flood potential inferred using GRACE gravity observations at several months lead time. Nat Geosci 7(8):588–592. https://doi.org/10.1038/NGEO2203

Rebischung P (2021) Terrestrial frame solutions from the IGS third reprocessing. https://doi.org/10.5194/egusphere-egut1-21-414. Conference Name: EGU21

Richey AS, Thomas BF, Lo M-H, Reager JT, Famiglietti JS, Voss K, Swenson S, Rodell M (2015) Quantifying renewable groundwater stress with GRACE. Water Resour Res 51(7):5217–5238. https://doi.org/10.1002/2015WR017349

Rodell M, Famiglietti J, Wiese D, Reager J, Beaudong H, Landerer F, Lo H-M (2018) Emerging trends in global freshwater availability. Nature 557:651–659

Rovira-Navarro M, van der Wal W, Barletta VR, Root BC, Sandberg Sørensen L (2020) GRACE constraints on Earth rheology of the Barents Sea and Fennoscandia. Solid Earth 11(2):379–395. https://doi.org/10.5194/se-11-379-2020

Scalan BR, Zhang Z, Save H, Sun AY, Müller Schmied H, van Beek LPH, Wieser DN, Wada Y, Long D, Reedy RC, Longevanne L, Döll P, Bierkens MFP (2018) Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. Proc Natl Acad Sci 115(6):E1080–E1089. https://doi.org/10.1073/pnas.1704665115

Scalan BR, Zhang Z, Rateb A, Sun A, Wiese D, Save H, Beaudong H, Lo MH, Müller-Schmied H, Döll P, Beek RV, Swenson S, Lawrence D, Croteau M, et al. (2019) Tracking seasonal fluctuations in land water storage using global models and GRACE satellites. Geophys Res Lett 46(10):5254–5264. https://doi.org/10.1029/2018GL081836

Schumacher M, Kusche J, Döll P (2016) A systematic impact assessment of GRACE error correlation on data assimilation in hydrological models. J Geodesy 90(6):537–559. https://doi.org/10.1007/s00190-016-0892-y

Springer A, Karegar MA, Kusche J, Keune J, Kurtz W, Kollet S (2019) Evidence of daily hydrological loading in GPS time series over
