Global high-resolution total water storage anomalies from self-supervised data assimilation using deep learning algorithms

Author(s):
Gou, Junyang; Soja, Benedikt

Publication date:
2024-02

Permanent link:
https://doi.org/10.3929/ethz-b-000665187

Originally published in:
Nature Water 2(2), https://doi.org/10.1038/s44221-024-00194-w

nature water

Article

https://doi.org/10.1038/s44221-024-00194-w

Global high-resolution total water storage anomalies from self-supervised data assimilation using deep learning algorithms

Received: 13 February 2023

Accepted: 5 January 2024

Published online: 12 February 2024

Check for updates

Junyang Gou ✉ & Benedikt Soja

Total water storage anomalies (TWSAs) describe the variations of the terrestrial water cycle, which is essential for understanding our climate system. This study proposes a self-supervised data assimilation model with a new loss function to provide global TWSAs with a spatial resolution of 0.5°. The model combines hydrological simulations as well as measurements from the Gravity Recovery and Climate Experiment (GRACE) and its follow-on (GRACE-FO) satellite missions. The efficiency of the high-resolution information is proved by closing the water balance equation in small basins while preserving large-scale accuracy inherited from the GRACE(-FO) measurements. The product contributes to monitoring natural hazards locally and shows potential for better understanding the impacts of natural and anthropogenic activities on the water cycle. We anticipate our approach to be generally applicable to other TWSA data sources and the resulting products to be valuable for the geoscience community and society.

Monitoring variations in the global water cycle is crucial for understanding the Earth's climate system[1]. The long-term trends in ice-sheet melting[2] and freshwater availability[3] can be investigated by analysing the water cycle over decades, while the short-term variations of the water cycle contain fruitful information for monitoring natural hazards like flood[4] and drought[5] events. To quantify variations in the global water cycle, total water storage (TWS), defined as the storage in all forms of water, has been implemented as an essential climate variable[6]. For decades, TWS has mainly been modelled by simulations from global hydrological models, including global hydrology and water resource models and land surface models[7]. The hydrological models can provide spatial variance and short-term temporal variations but suffer from providing reliable long-term trends, which indicate the climate and human-induced changes in water storage[8]. Since 2002, the Gravity Recovery and Climate Experiment (GRACE) and its follow-on (GRACE-FO) missions (hereafter GRACE) have provided us with a unique opportunity to monitor the changes in global TWS anomalies (TWSAs) by measuring gravity field variations[9–11]. The satellite-measured TWSAs have unprecedented accuracy with global coverage due to the physical measurement principle and provide valuable information about the Earth's climate system from a macro perspective[1,12–14].

Although GRACE products have been widely used, their coarse spatial resolution of about 3° is among the key factors limiting the applications in related fields, such as understanding water storage changes in small catchments[14]. The problem of low spatial resolution mainly comes from two origins as follows. First, the design of the orbit and the accuracy of the instruments inherently limit the possible spatial resolution[15,16]. Second, postprocessing approaches are needed for obtaining meaningful signals[17,18] but they also attenuate the actual geophysical signals, especially the high-frequency signals[19–21]. As a result, reconstructing these high-frequency geophysical signals is vital. To improve the resolution, incorporating additional information with higher spatial resolution is necessary. In the specific task of downscaling GRACE TWSAs, the most promising high-resolution information is contained in hydrological models and measurements. The hydrological models can directly simulate TWSAs, whereas other hydrological parameters such as precipitation provide valuable auxiliary information by considering the water balance[14,22]. Many studies have proven the feasibility of assimilating data to downscale GRACE products over specific regions[23–26] but few of them applied the methods successfully on a global scale. The downscaling algorithms with global generalizability so far can provide GRACE TWSAs of 0.5° using partial least squares

Institute of Geodesy and Photogrammetry, ETH Zurich, Zurich, Switzerland. ✉e-mail: jungou@ethz.ch

https://doi.org/10.1038/s44221-024-00194-w

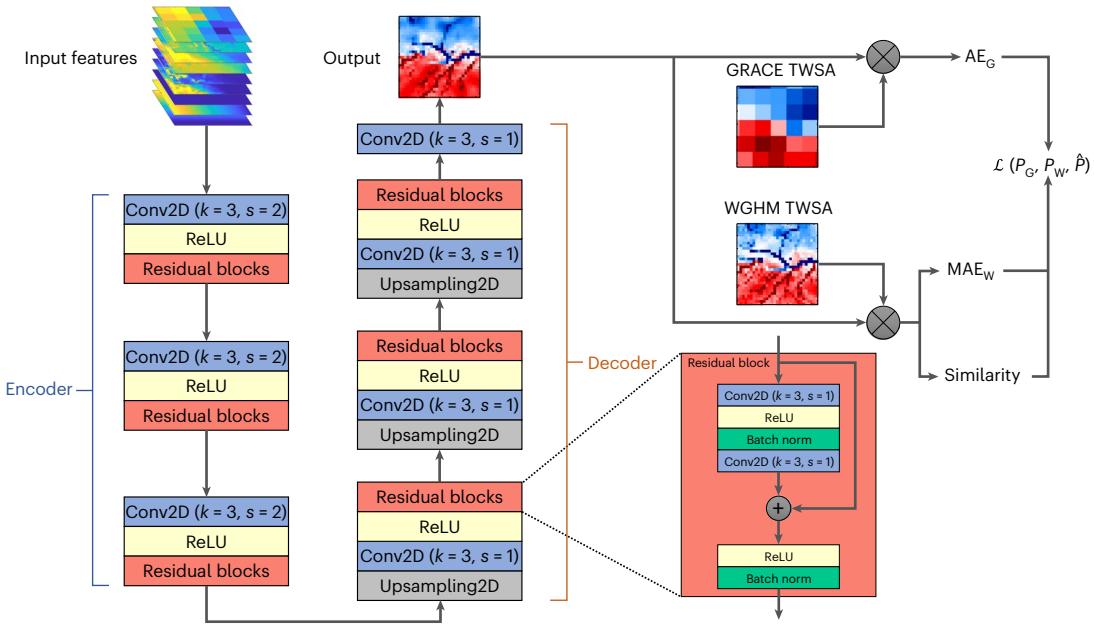

Fig. 1 | The structure of the designed deep learning model in this study with an enlarged structure of the used residual blocks. The 2D convolutional layers and upsampling layers with bilinear interpolation are denoted by Conv2D and Upsampling2D. The kernel size of the 2D convolutional layers is denoted by *k*, whereas the stride is denoted by *s*. The input features go through the encoder–decoder structure to generate the predictions with the same size, which will be compared to the GRACE and WGHM TWSAs to compute the loss function. Therefore, the optimizing process is self-supervised.

regression[27] or an ensemble Kalman filter assimilation pipeline[28] but still have some deficiencies, such as insufficient intrabasin variability preservation or interbasin mass conservation.

In recent years, deep learning has progressed rapidly and shown remarkable potential in modelling the Earth system[29–31]. Many studies investigated the potential of applying deep learning or classical machine learning approaches to downscale GRACE measurements in a supervised learning context[32,33]. The main challenge is the need for high-resolution ground truth of TWSAs, which are inaccessible. Therefore, the studies usually generate the training pairs by downsampling the high-resolution hydrological simulations into the same resolution as GRACE. Under the assumption that the relationship between predictors and targets holds in different resolution domains, they apply the trained model on the original high-resolution hydrological simulations to obtain the downscaled GRACE predictions with the same resolution[34–36]. Another way to deal with the lack of ground truth is generating GRACE-like TWSAs by applying a Gaussian smoother on the hydrological simulations and training the model based on synthetic data pairs[37]. In this context, the necessary assumption is that the captured relationship also holds for the real GRACE measurements. However, all the aforementioned studies only applied their methods locally to continentally, indicating the difficulty of applying proposed deep learning methods globally due to their inherent challenges about generalization[38]. This study contributes to the downscaling problem and solves the two mentioned inadequacies of the existing deep learning algorithms, namely relieving assumptions between different domains and providing global generalizability. First, we developed a loss function based on the average deviations between the outputs and the GRACE TWSAs over a certain area and the similarity between the outputs and WaterGAP Hydrology Model (WGHM)[39] simulations (Fig. 1; Methods). Since the GRACE and WGHM TWSAs are part of the inputs, we do not need extra labels or certain assumptions for generating synthetic training pairs. As a result, the network parameters can be optimized in a self-supervised manner without explicit assumptions bridging the high–low spatial resolution domains or simulation–measurement domains. Second, our downscaled TWSAs inherit the global generalizability from both GRACE and WGHM TWSAs and cover all the land areas, including coastal areas and small islands, except for Greenland and Antarctica due to the deficiency of hydrological models[39,40]. Our model is based on the principle of convolutional neural networks[41], allowing us to consider spatial correlations between individual cells. Our analysis shows the impressive performance of the proposed algorithm in providing high-resolution TWSAs, which represent the high-resolution structures while keeping accurate mass conservation on the basin scale. Therefore, the water balance equation can be better closed in the basins smaller than the GRACE-effective resolution. Ultimately, we also discuss the potential usages of the obtained high-resolution TWSA product for analysing the impacts of climate change and anthropogenic activities on a local scale, as well as for natural hazard monitoring.

Global downscaled TWSA product with uncertainties

We determined a global high-resolution TWSA product from April 2002 to December 2019, covering all global land areas except for Greenland and Antarctica. To provide uncertainty information, we combined probabilistic deep learning principles and Monte Carlo simulations to estimate the uncertainties using deep ensembles[42]. An example of the highly resolved TWSA product is shown in Fig. 2 with six major river basins enlarged. On the global scale, the seasonal changes in water storage are consistent with the ones observed by the GRACE measurements (Supplementary Videos 1 and 2). The regional maps demonstrate the visibility of the main river systems with refined details, which are inherited from the WGHM simulations. Moreover, the results are visually smoother than the WGHM TWSAs, indicating effective noise reduction. The reduction of the outliers is not only beneficial to reducing the amount of abnormal pixels but also helpful for gaining accurate values of the other neighbouring pixels. Once the outliers are reduced, the magnitudes of the neighbouring pixels with actual signals are calibrated by considering the constraint of agreements with GRACE TWSAs over an area larger than the GRACE-effective resolution[16]. Overall, the downscaled TWSAs have a global median uncertainty of 7.3 mm. The regions where water storage changes rapidly are usually accompanied by relatively large uncertainties due to their higher TWSA values, such as the mainstreams shown in Fig. 2.

https://doi.org/10.1038/s44221-024-00194-w

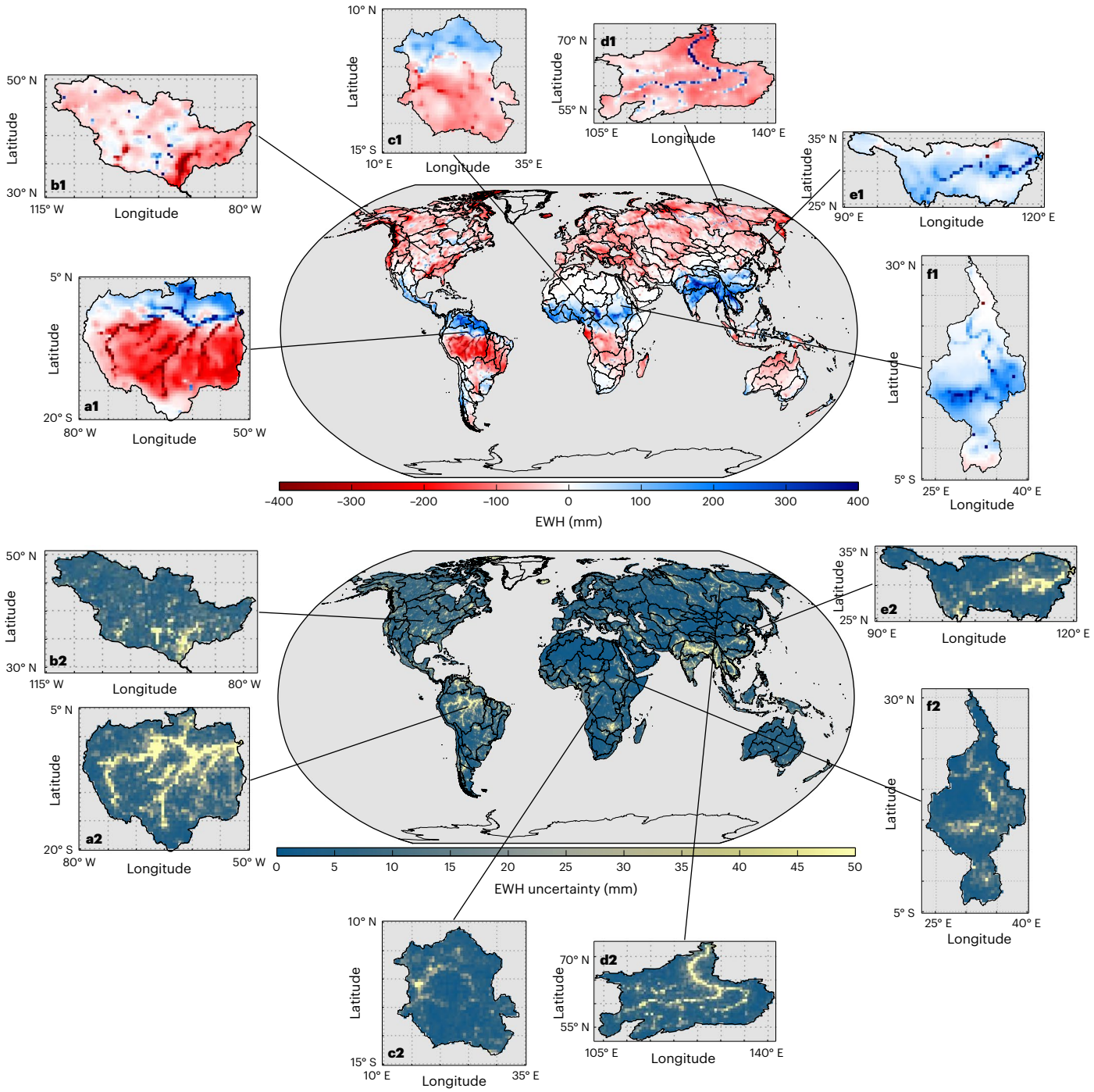

Fig. 2 | The downscaled TWSAs (top) and their uncertainties (bottom) in August 2008. The data are provided in the format of equivalent water height (EWH) with a spatial resolution of 0.5°. **a**–**f**, Six major river basins are shown with enlarged details: Amazon (**a**), Mississippi (**b**), Congo (**c**), Lena (**d**), Yangtze (**e**) and Nile (**f**). The regions without valid information are shaded. Note the different spatial scales for the enlarged images for better visualization.

High-resolution details with large-scale mass conservation

Since accurate high-resolution TWSA measurements on a global scale are inaccessible, we cannot directly evaluate the quality of the downscaled product. Therefore, we rely on the GRACE measurements and WGHM simulations in different contexts. First, we compared the downscaled TWSAs with the WGHM simulations to study whether the high-resolution structures are sufficiently reconstructed. To achieve this goal, we considered each pixel over the whole time span as a time series and computed the pixel-wise Pearson correlations between downscaled and WGHM TWSAs. The impact of the inaccurate magnitudes in the WGHM TWSAs is reduced since the Pearson correlation is invariant under changes in scale. As shown in Fig. 3a, the overall correlation is high with a median value of 0.80, meaning an improvement of 51% compared to GRACE TWSAs (0.53). The relatively low correlations are mainly found in arid regions such as the north of Africa, the Middle East and the middle of Asia. These low correlations are understandable due to weak hydrological signals in these arid regions since GRACE and WGHM are not sufficiently sensitive to accurately measure or simulate them. Furthermore, comparisons with independent satellite altimetry measurements show that the high-resolution information of the downscaled product is beneficial (Supplementary Table 1 and Supplementary Fig. 5).

https://doi.org/10.1038/s44221-024-00194-w

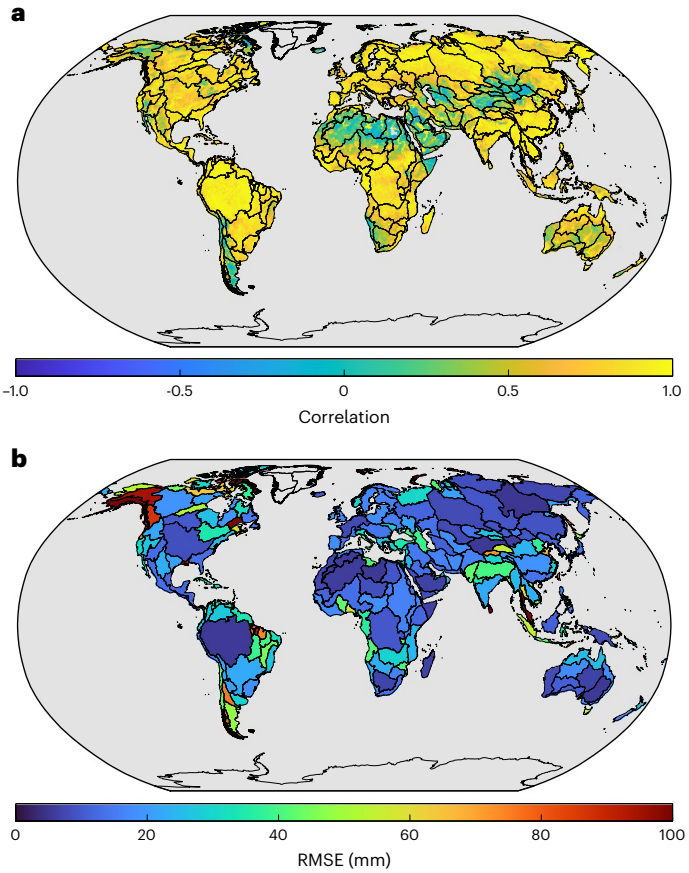

Fig. 3 | Evaluation of the downscaled TWSAs from 2002 to the end of 2019. **a**, The global pixel-wise Pearson correlation with WGHM simulations. **b**, The basin-wise RMSE compared to GRACE measurements. The regions without valid information are shaded.

To evaluate the basin-wise quality of the downscaled TWSAs, we rely on the GRACE measurements since the GRACE TWSAs are considered to be accurate over the their effective resolution[16,43]. We first averaged GRACE and downscaled TWSAs of each individual basin to generate basin-wise time series and compute the root mean square errors (RMSEs) between these two types of time series. The results are shown in Fig. 3b for 288 basins globally. The RMSEs are lower than 30 mm in most of the land areas, resulting in a global average RMSE of 21.9 mm weighted by basin areas. This value demonstrates the quality of the downscaled TWSAs since the typical GRACE uncertainties are 20–30 mm (ref. 15). Compared to WGHM simulations (weighted RMSE of 49.2 mm), our method provides an improvement of around 56%. Further analysis reveals that certain basins exhibit relatively high RMSEs, such as the basins in the glaciated areas of Alaska. Insufficiently modelling of glaciers and ice sheets is a known issue of the hydrological models[8,39]. Since we trained one neural network for the whole globe, the network cannot handle the substantial differences between the hydrological simulations and GRACE measurements in these specific regions because the issue is inessential in other regions, constituting the major part of the samples. Therefore, we consider this issue as a trade-off between generalizability and performance in specific areas.

Long-term trends and seasonal variations

To understand the performance of the proposed deep learning model on the basin scale in more detail, we plot the time series of the average TWSAs over the six selected major basins (Fig. 4). In the Amazon basin, dominant seasonal signals are relatively homogeneous. Both downscaled and WGHM TWSAs agree well with GRACE measurements, although the WGHM simulations tend to underestimate the amplitudes. For the basins where average TWSAs are less stationary and with more inhomogeneous variations (Mississippi, Congo, Lena and Nile), the downscaled TWSAs agree better with the GRACE observations than the WGHM simulations. These phenomena may be related to the known limitation of hydrological simulations, as they cannot fully capture the trends in water changes[8]. For example, the downscaled TWSAs successfully model the steady decrease in the Lena basin after 2007 and the increase in the Nile basin since 2010, whereas the WGHM time series do not represent these trends properly. However, both downscaled and WGHM TWSAs underestimate the increasing trend in the Yangtze River basin after 2010. A potential reason is the active anthropogenic impacts in this basin, such as construction of dams[44]. Although the WGHM model considers human intervention, it may perform unsatisfactorily in specific regions and result in a relatively big disparity from GRACE measurements[39]. As a result, the typical relationship captured by our model based on global data may not be the best solution in these regions, which is again a trade-off between generalizability and performance in specific areas. Conversely, the good agreements between the downscaled TWSAs and the GRACE measurements in the Mississippi and Congo basins demonstrate the performance of our model.

To quantify the performance of our method in retaining long-term trend, annual and semi-annual signals, we estimated these signals from the three TWSA types over 160 basins larger than 200,000 km^2 and show the results in Fig. 5a. Here, we set the spatial threshold of 200,000 km^2 to obtain more reliable GRACE estimations as reference[16,39,43]. The correlation between WGHM- and GRACE-derived trends is only 0.47, showing the major limitation of the hydrological models in capturing long-term trends. Our method substantially improves this situation and reaches a correlation of 0.94. This improvement reveals the effectiveness of the proposed algorithm for data assimilation. The network has learned to rely on the GRACE measurements to calibrate the average magnitudes over an area larger than the GRACE-effective resolution. Therefore, the trends contained in the GRACE measurements have been successfully inherited to the downscaled TWSAs. The largest trend differences are found in the glaciated basins of Alaska and their neighbouring basins (Fig. 5b), which related to the problem of insufficient modelling of glaciers and the resulting leakage errors again. The performance of WGHM on annual and semi-annual signal estimations is clearly better than estimating trends with correlations of 0.83 for both. Nevertheless, our method still outperforms it and reaches correlations of 0.97 and 0.95, respectively. The results prove the realistic temporal changes in our downscaled TWSAs on the basin scale, even though we treated every month separately and did not explicitly feed any temporal information to the model. Moreover, the phases of the annual signals of our downscaled product agree well with those measured by GRACE (Fig. 5c), with the only major differences occurring in the north of Africa where the hydrological signals are weak and the phases are therefore ambiguously defined. It is a remarkable improvement compared to WGHM simulations, which suffer from phase shifts of 2 or 3 months for many of the basins globally (Supplementary Figs. 3 and 4).

Closing water balance equation beyond the GRACE resolution

The downscaled TWSAs are beneficial for better closing the water balance equation in regions smaller than the GRACE-effective resolution. Figure 6 depicts the agreements between water changes inferred from the downscaled TWSAs and those computed from ERA5-Land water budget components[45] in level-4 basins[46]. The downscaled TWSAs show a reasonable ability to close the water balance equation globally with positive Nash–Sutcliffe efficiency (NSE)[47] of 83% studied area, whereas GRACE and WGHM have positive NSE in 77% and 75%. Most of the negative values come from the arid regions where the hydrological signals are weak. The improvements of downscaled TWSAs compared

https://doi.org/10.1038/s44221-024-00194-w

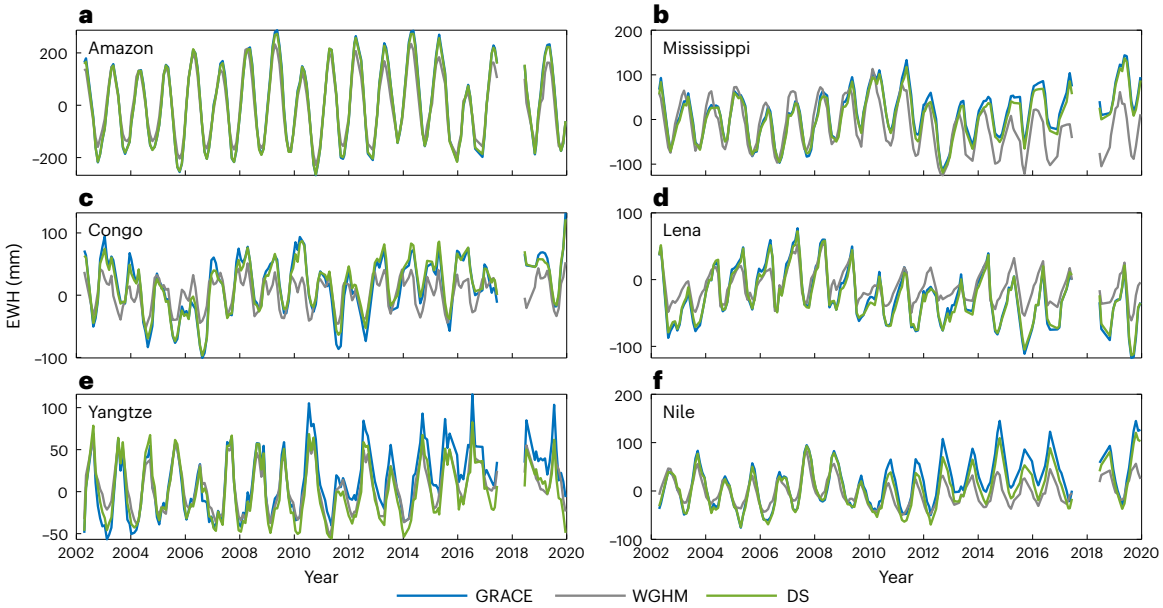

Fig. 4 | The basin-wise average TWSA time series (mm) of six selected major river basins for the whole studied time span from 2002 to the end of 2019. a–**f**, Amazon (**a**), Mississippi (**b**), Congo (**c**), Lena (**d**), Yangtze (**e**) and Nile (**f**). Our downscaled product is abbreviated as DS. Note the different *y* axis ranges for different basins.

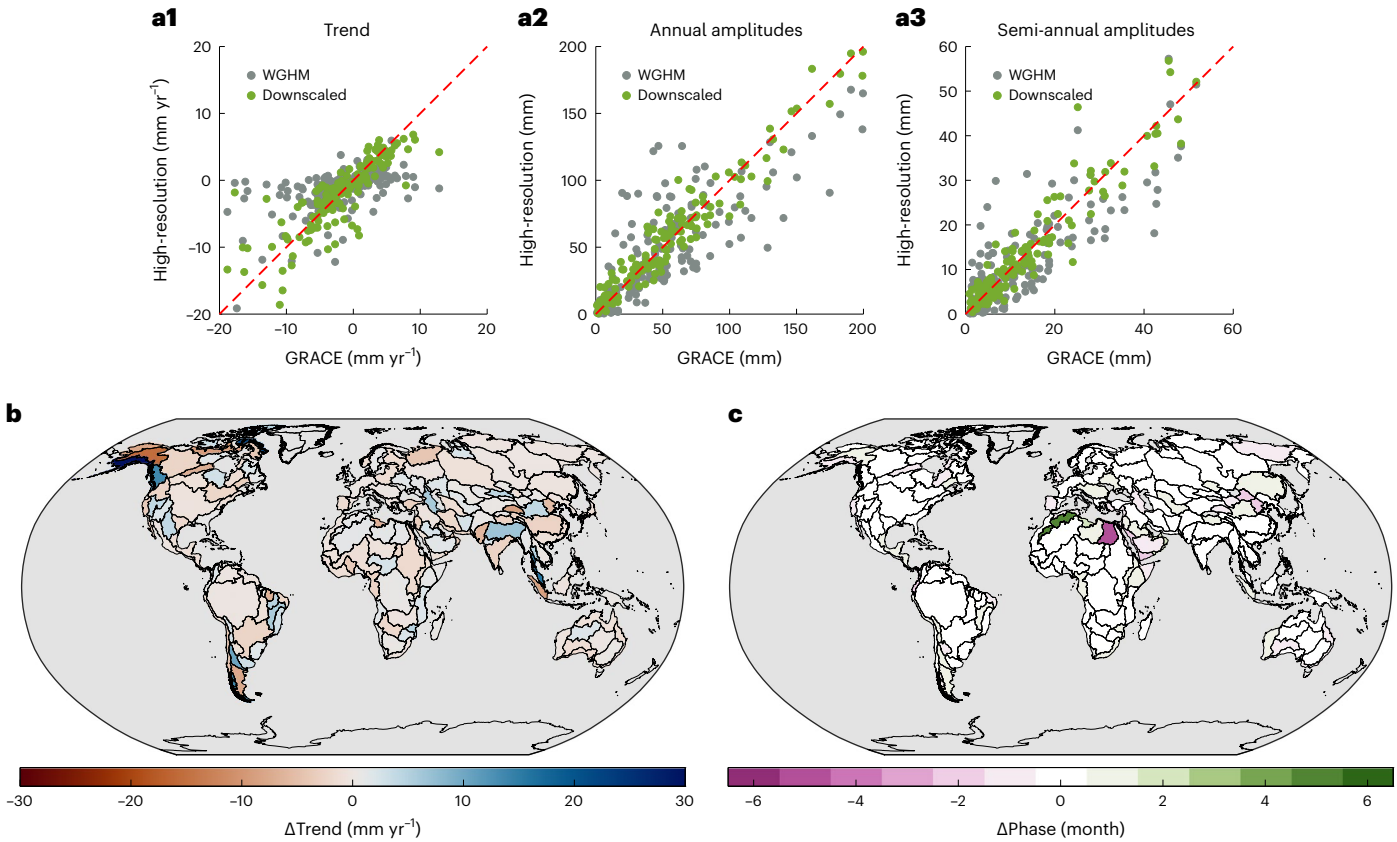

Fig. 5 | Temporal decomposition of the basin-wise signals for the 160 basins larger than 200,000 km². a, The long-term trend, annual and semi-annual amplitudes estimated from the high-resolution products (*y* axis) versus GRACE measurements (*x* axis). **b**, Map of basin-wise trend differences (downscaled GRACE). **c**, Map of basin-wise phase shifts (downscaled GRACE). The regions without valid information are shaded.

to the original GRACE product strongly correlate with the basin sizes (Fig. 6b). The average improvement in NSE given by the downscaled TWSAs is 0.13 for basins larger than GRACE-effective resolution (200,000 km²)[39], 0.21 for basins between the effective resolution and limiting resolution (63,000 km²)[16] and 1.21 for basins smaller than 63,000 km². Conversely, the benefits compared to WGHM TWSAs do not strongly correlate with basin sizes but probably come from the more accurate values obtained from data assimilation.

https://doi.org/10.1038/s44221-024-00194-w

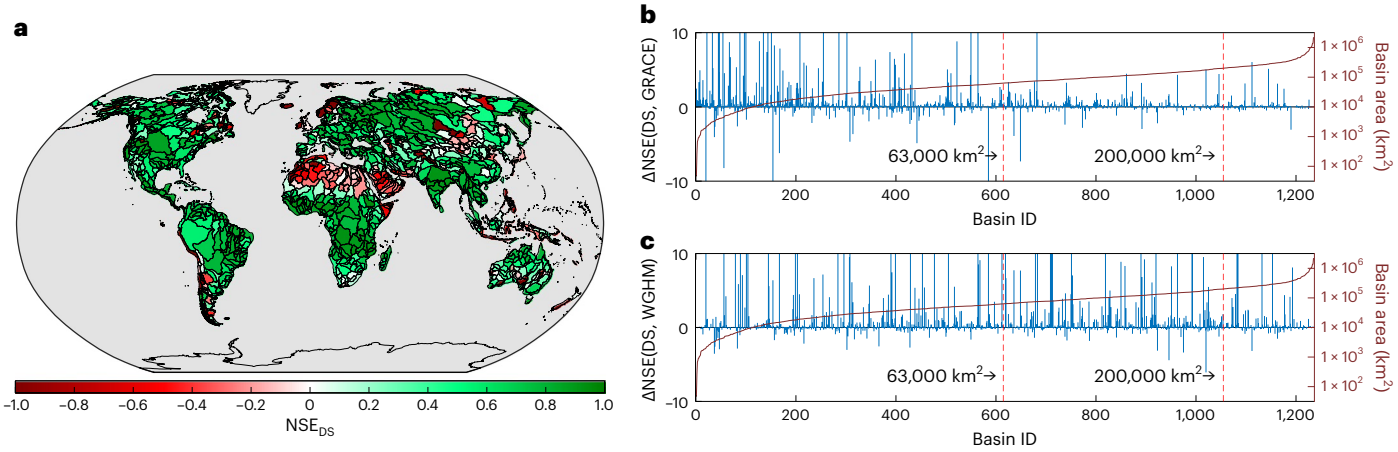

Fig. 6 | Agreements between total water storage changes inferred from downscaled TWSAs and those computed from water budget components. **a**, Basin-wise NSE of downscaled TWSAs in level-4 basins. The regions without valid information are shaded. **b**, Differences in NSE (left axis) between downscaled and GRACE TWSAs sorted by basin area (right axis). Positive values indicate a better performance of the downscaled TWSAs. **c**, The same as **b** but compared with WGHM TWSAs.

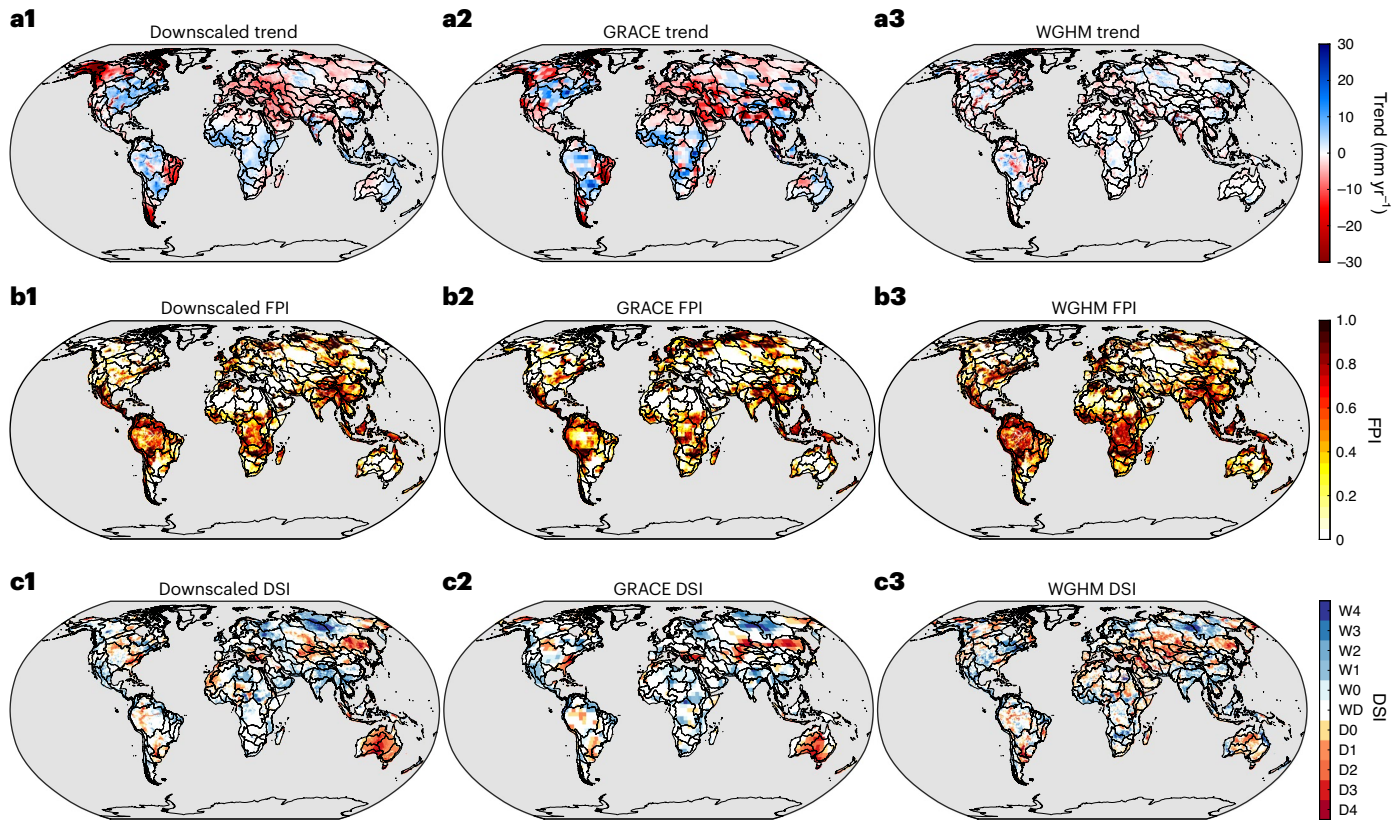

Fig. 7 | Comparison of three high-level products derived from three different sources of TWSAs. The columns from left to right show the indices derived from downscaled, GRACE and WGHM TWSAs. The top row shows the pixel-wise TWSA trends from 2002 to 2019 in the form of EWH (mm yr^{-1}) after removing the annual and seasonal variations. The middle row shows the pixel-wise maximum FPI of 2008. The numbers indicate the risk of potential flooding events, with a large number indicating relatively high risk and vice versa. The bottom row shows the monthly DSI for August 2008. The 11 categories from red to blue indicate exceptional drought, extreme drought, severe drought, moderate drought, abnormally dry, near normal, slightly wet, moderately wet, very wet, extremely wet and exceptionally wet.

Discussion

The downscaled TWSAs provide special insight for studying the climate and anthropogenic impacts locally, which enables the development of targeted strategies for sustainable management of water resources. Figure 7a shows the comparison of the derived trends from the three sources of TWSAs. The high-resolution signals in the downscaled trends are inherited from WGHM simulations and their values are calibrated by considering the agreement with GRACE measurements on the effective resolution. For example, the downscaled trends clearly show the three hot spots of groundwater depletion in the United States (the High Plains aquifer, the Mississippi embayment and the Central Valley of California)[48,49]. Among them, the notable negative trends in the High Plains aquifer are not observable in GRACE TWSAs since the positive trends caused by progress from dry

to wet periods[3] on the neighbouring pixels average them out. GRACE TWSAs indicate negligible to positive trends mainly caused by increasing precipitation in central and southern India[3,50], which do not fully represent the remarkable water storage declines in the regions with high population density and groundwater irrigation[51]. Caveat: we note that the downscaled product generally depends on the quality of used hydrological simulations. Prominent trends in the simulations will impact downscaled trends to a certain extent. We should interpret these signals with care since relatively larger uncertainties are expected as a result of imperfections in the hydrological simulations.

The downscaled TWSAs enable multiple downstream applications, including flood and drought monitoring on a local scale. To study the potential benefits of our downscaled TWSAs, we computed two well-known TWSA-based indices: the flooding potential index (FPI)[4] and the drought severity index (DSI)[5]. We report the maximum FPI of the year 2008 in Fig. 7b to show the most notable signals in a single plot, whereas Fig. 7c depicts the monthly DSI of August 2008. FPI is sensitive to the value range of TWSAs since it relies on the relative relationship between the storage deficit and accumulated precipitation. The outliers present in WGHM simulations can ruin the relationship and cause unrealistic flood potential, resulting in an FPI map with far more high-risk regions than GRACE-derived FPI. Our proposed method clearly ameliorates this issue by suppressing the outliers, which allows us to obtain realistic high-resolution FPI with a reasonable visual agreement with GRACE-derived FPI, such as in the Congo basin and along the eastern coastline of North America. Similarly, the abnormal values in the WGHM simulations may cause opposite categories in DSI, resulting in noisy patterns (north of Africa) or abnormally underestimated severity (Australia). Again, the DSI derived from the downscaled TWSAs agree better with the GRACE-derived ones and open the window to monitor extreme environmental events with higher spatial resolution. However, the environmental monitoring indices derived from downscaled TWSAs inherited the same limitations that GRACE TWSAs have[52], since the downscaled TWSAs only provide higher spatial resolution but do not provide longer observations.

Our current approach still has some limitations, which reveal the potential for further improving high-resolution TWSAs. First, more effort should be put into modelling the glaciers by considering additional measurements or specific models. Second, human intervention modelling is sophisticated and may need specific modifications. Including population, farming area and water usage statistics in the deep learning model may provide better results. Last, deep learning models have the potential to consider constraints based on the interactions between different forms of water, such as the interaction between terrestrial water and ocean or free water and glaciers. Nevertheless, with the preliminary study on the potential use of the downscaled TWSAs for monitoring water change and natural hazards, we demonstrate the significance of the proposed method. In practice, timeliness is a key factor. The training process of the proposed algorithm can be finished in around 3 days for the global model using consumer-level platforms (NVIDIA RTX3080TI), which is efficient considering the typical delay of GRACE monthly products. Therefore, the major limitations for rapidly delivering high-resolution TWSAs are the processing time demands of GRACE measurements and hydrological simulations. For applications that need higher temporal frequencies, such as daily to weekly solutions, we can benefit from the principle of online machine learning and the proposed model can be updated within 1 hour. The operational delivery of the downscaled product should be beneficial for the geoscience community and society, especially in the fields of hydrology, climate science, sustainable water management and hazard prediction.

Methods

GRACE mascon solution

The analysis centres of GRACE provide a variety of products regularly. One of the most user-friendly products is the mass concentration (mascon) solution, where the mass variations are directly estimated by explicitly relating the intersatellite range–rate measurements to the mascon formulation[53–55]. Compared to the spherical harmonic solutions, the mascon solutions suffer less from leakage errors and can better separate the land and ocean signals[56]. Therefore, the mascon solutions typically have a finer resolution for small regions[57]. In this study, we used the mascon solutions provided by NASA Jet Propulsion Laboratory (JPL)[58]. JPL has applied many data processing steps, including the replacement of $C_{2,0}$ coefficients with the solutions from satellite laser ranging[59], applying a Glacier Isostatic Adjustment (GIA) model[60], removal of the impacts of ocean, atmosphere and land ice masses. In the end, the remaining monthly gravity changes can provide a precise measure of mass redistribution in the Earth's water cycle[57]. Besides, the mean values from 2004.0 to 2009.999 are removed from the products to produce the TWSA in the form of EWH. The product used in this study is without the land-grid-scaling gain factors[61] so that the data can provide us with information that is entirely independent of hydrological models.

Hydrological models and basin boundaries

WaterGAP is a global hydrological model that thoroughly describes water storage, usage and resources in all land areas except for Antarctica. WaterGAP v.2.2d, including WGHM, was published in 2021[39]. WGHM comprehensively models daily water flows and water storage since it includes various forms of water, such as groundwater, rivers and snow. As one of the standard WGHM outputs, monthly TWS is provided at a high spatial resolution of 0.5° × 0.5°. This TWS product is the sum of the water storage in the canopy, snow, soil, groundwater, wetland, lake, reservoir and river storage[39]. Owing to the modelling approach and high spatial resolution, we can observe the principal rivers from the simulated TWS. The direct comparison between the obtained TWSAs and the GRACE TWSAs is possible[62]. However, we must remove the average values from 2004.0 to 2009.999 to generate the TWSAs with the same temporal baseline. Since the WGHM-modelled TWS does not include assimilation of GRACE measurements, the obtained WGHM TWSAs and GRACE TWSAs are entirely independent. Although the global structures of the WGHM TWSAs are noticeably more finely resolved than GRACE products, the WGHM TWSAs also have two non-negligible limitations. First, the WGHM simulations are much noisier due to errors in the simulation procedure. Second, the values of the WGHM TWSAs are less accurate compared to the GRACE TWSAs since they are not based on real observations. WGHM clearly underestimates the mean annual TWSA amplitudes in more than half of 147 investigated river basins by more than 10%, which may relate to the fact that WaterGAP does not simulate glaciers[39].

In addition, we included hydrological information from the Global Land Data Assimilation System (GLDAS), which aims to assimilate satellite- and ground-based observational products to provide fields of land surface states and fluxes[40]. Within this platform, many land surface models (LSM) are integrated. In this study, we used the data from the GLDAS Noah Land Surface Model L4 monthly 0.25° × 0.25° v.2.1 (refs. 40,63). GLDAS v.2.1 is forced with a combination of model and observation data from 2000 to the present without assimilating GRACE measurements. Therefore, the products provided by GLDAS v.2.1 are additional data sources independent of the GRACE TWSAs. Considering the water balance equation, which describes the relationship between the changes of TWSA (TWSC) and precipitation (P), evapotranspiration (ET) and runoff (R):

$$\mathrm{TWSC} = P - ET - R, \tag{1}$$

we exclusively focused on the three mentioned parameters. The data are downsampled into the resolution of 0.5° × 0.5° by computing the average of four neighbouring pixels to obtain the same resolution as the WGHM TWSAs.

The hydrological boundaries are obtained from HydroBasins[46], which represent a series of vectorized polygon layers that depict sub-basin boundaries at a global scale. All continents, except Antarctica, are included. The HydroBASINS product follows the Pfafstetter concept[64] and provides levels 1 to 12 globally. In this study, we focused on the HydroBasins level 1 (nine continents), 3 (292 sub-basins) and 4 (1,342 sub-basins) products.

Feature selection and preprocessing

One of the essential prerequisites for the success of the deep learning model is to determine a set of meaningful features that can represent the changes in TWSAs to a sufficient degree. First, the TWSAs from GRACE and WGHM are the most important features because the GRACE TWSAs have relatively accurate values over a larger region and those from WGHM provide information about high-resolution structures. Furthermore, we included precipitation, evapotranspiration and runoff as features inspired by equation (1). We should note that the GLDAS provides runoff split into three components: storm surface runoff, baseflow-groundwater runoff and snow melt. Therefore, we have five additional features in total. In the end, since multiple studies pointed out the correlation between the changes of TWS and geocoordinates[1] and the positive contribution of geocoordinates to global deep learning models[65], we also considered latitudes and longitudes as additional features. We normalized the features on the basis of their 0.01th percentiles and 99.99th percentiles to reduce the impacts of outliers.

The final step before we fed the data into our model is splitting the global area into small patches of the same size. We first found all the pixels on land by relying on the basin boundaries. Then, we considered them as the central pixels and generated a 16° × 16° patch around each central pixel, which means each patch has a size of 32 × 32 with a resolution of 0.5°. The patch size is a multiple of the coarse resolution of GRACE TWSAs (3° × 3°). A 16° × 16° patch contains more than 25 effective GRACE pixels, making the patch-wise average of GRACE TWSAs meaningful. Some patches near the coastlines unavoidably contain areas over oceans where the hydrological models do not provide any values. We filled these with the average values of valid pixels in the same patch. The patches were considered as images in our deep learning model and the nine features were put into nine channels.

Self-supervised data assimilation model

The lack of high-resolution ground truth impedes the application of supervised learning approaches to provide high-resolution TWSAs. All the mentioned deep learning-based studies artificially generated input–output pairs, which capacitates a supervised regression but always under the constraints of certain assumptions. To avoid strict assumptions about the input–output pairs and exploit the information from the available data sources as much as possible, we proposed a data assimilation model with a specifically designed loss function that allows self-supervised optimization. Our model is based on the principle of the convolutional neural networks[41], which is a specific type of neural networks that uses convolution in place of general matrix multiplication in some layers. The convolutional operators, usually known as kernels, can extract high-level feature maps by considering the relative positional relationship between pixels or low-level features. In our specific case, the model can extract information about hydrological phenomena, such as water storage changes in waterbodies, from the values of individual pixels. We also applied the concept of residual learning[66]. In this context, the network explicitly approximates the residual function $\mathcal{F}(x) = \mathcal{H}(x) - x$, which is the difference between the original target function $\mathcal{H}(x)$ and the input x. The fitting of the residual function should not be more difficult than fitting the original target function itself due to the existence of the skip connection. As a result, a deeper model should have a training error no greater than its shallower counterpart[66]. Batch normalization[67] is also included to reduce the sensitivity to the initialization and thereby improve the optimizing process.

Our model aims to assimilate the satellite observations and hydrological simulation by balancing the accurate values from GRACE observations over an area larger than their effective resolution and the high-resolution structures from the WGHM simulations. Therefore, the loss function is designed in the way that the outputs of our model are compared with both inputs, GRACE TWSAs and WGHM TWSAs. The first goal of our optimizing process is to let the values of the outputs be as close as the GRACE TWSAs over each patch. Since the GRACE measurements of individual 0.5° pixels are not representative, we computed the absolute error (AE) between the averaged GRACE TWSAs and the averaged predicted TWSAs over each patch:

$$\mathrm{AE_G}\left(P_\mathrm{G}, \hat{P}\right) = \left| \frac{1}{N}\sum_{n=1}^{N} p_{\mathrm{G},n} - \frac{1}{N}\sum_{n=1}^{N} \hat{p}_n \right|, \tag{2}$$

where P_G and $\hat{P}$ denote the GRACE patches and predicted patches, respectively. Each patch includes N pixels with values denoted by p_G and $\hat{p}$ for GRACE and predicted patches. The second goal of our optimizing process is to learn the high-resolution structures from the WGHM TWSAs. For this purpose, we introduced the Pearson correlation coefficient (R) between the outputs and WGHM TWSAs to describe the similarity since it proves superior to other similarity metrics like structure similarity index (SSIM) as argued in another study[68] and confirmed in our tests. We introduced a second metric to enhance the measurements of structural similarity, namely the mean absolute error (MAE) between WGHM TWSAs and predicted TWSAs. Equations (3) and (4) show the corresponding definitions with p_W denoting the pixels of the WGHM patch P_W and $\hat{p}$ denoting the pixels of the output patch $\hat{P}$, respectively:

$$\mathrm{R}(P_\mathrm{W}, \hat{P}) = \frac{\sum_m \sum_n \left(P_{\mathrm{W},mn} - \bar{P}_\mathrm{W}\right)\left(\hat{p}_{mn} - \bar{\hat{P}}\right)}{\sqrt{\left(\sum_m \sum_n \left(P_{\mathrm{W},mn} - \bar{P}_\mathrm{W}\right)^2\right)\left(\sum_m \sum_n \left(\hat{p}_{mn} - \bar{\hat{P}}\right)^2\right)}}, \tag{3}$$

$$\mathrm{MAE_W}\left(P_\mathrm{W}, \hat{P}\right) = \frac{1}{N}\sum_{n=1}^{N} \left| p_{\mathrm{W},n} - \hat{p}_n \right|. \tag{4}$$

Finally, we combined the proposed terms to achieve the two goals within the same optimizing process, leading to the following formulation of the loss function:

$$\mathcal{L}\left(P_\mathrm{G}, P_\mathrm{W}, \hat{P}\right) = \frac{1}{B}\sum_{b=1}^{B} \left\{ \mathrm{AE_G}(P_\mathrm{G}, \hat{P}) + \left[1 - \mathrm{R}\left(P_\mathrm{W}, \hat{P}\right)\right] \times \mathrm{MAE_W}\left(P_\mathrm{W}, \hat{P}\right) \right\}, \tag{5}$$

where B is the batch size. The reason for using both AE and MAE rather than the L2 metrics is that the L1 metrics are usually more robust against outliers. During our experiments, we observed that the GRACE-term and WGHM-term converge with similar magnitudes, indicating the ability of our model to balance the two terms without over-relying on any of them. Thus, there is no need to add more hyperparameters to weigh these two terms explicitly.

The detailed realization of the global model is shown in Fig. 1, including details of the applied residual blocks. The encoder contains three two-dimensional (2D) convolutional layers followed by a ReLU activation function ($\mathrm{ReLU}(x) = \max(0, x)$) and three residual blocks. The encoding process is realized by the fact that each convolutional layer has a stride of 2 to reduce the size of the outputs and increase the receptive field. The convolutional layers have increasing numbers of kernels (16, 32 and 64) with a size of 3 to increase the latent information, namely the feature dimension. In the decoder, the feature maps have to be upsampled first. Here, we use 2D-bilinear upsampling layers to pre-upsample the feature maps and feed them into the 2D convolutional layers, followed by ReLU and residual blocks. Unlike the encoder, the convolutional layers have stride 1 to keep the size of outputs the same. They have 64, 32 and 16 kernels to reproduce the spatial

information from the latent information. At the end of the architecture, another convolutional layer with kernel size 1 without activation function is designed so that it can project the final feature maps to the actual TWSA values. The resulting output size is the same as the input size (32 × 32). Once the outputs are generated, they are compared to the original GRACE and WGHM TWSAs to compute the loss using equation (5). To understand the benefits of our proposed model structure, we can analyse the two parts of the optimizing process separately. If we only consider the parts of the loss function related to the WGHM terms (bottom part in Fig. 1), our model is similar to an autoencoder[69,70], which aims to reconstruct the WGHM TWSAs while reducing noise. Therefore, the number of outliers is remarkably reduced. Then, optimizing the part of the loss function related to the GRACE terms is like a regression problem (top part in Fig. 1), aiming to calibrate the values of the reconstructed high-resolution TWSAs on the patch scale. Our proposed loss function combines these two principles so that they are optimized jointly.

To tune the model structure and hyperparameters, we started with experiments over four river basins (Amazon, Congo, Mississippi and Lena). At this stage, we randomly split the data of 158 months until the end of 2016 into training (70%, 110 months), validation (15%, 24 months) and test sets (15%, 24 months). By monitoring the training process, we observed that the models usually converged between training epochs 120 and 150. The models did not suffer from overfitting issues and had similar performance on data from 2017 to the end of 2019. Therefore, we decided to train our models using all 180 months of data to obtain the same quality for the whole time interval (April 2002 to December 2019). We used TensorFlow v.2.6.0 (ref. 71) to implement all the networks as well as the training, validation and test processes. The optimizer is Adam[72] with default settings and a batch size of 512. The training process is efficient and can be finished in around 3 days for the global model using consumer-level platforms (NVIDIA RTX 3080TI).

Estimating uncertainties based on deep ensemble

To quantify the uncertainties of downscaled TWSAs, we followed the principle of deep ensemble learning[42] and trained five independent deep learning models from scratch with different random initial states. As a relatively small number of independent models is sufficient for modelling the predictive uncertainty[73], we computed our ensemble results (μ_*) and uncertainties (σ_*) as:

$$\mu_* = \frac{1}{M}\sum_{m=1}^{M}\mu_{\theta_m}, \tag{6}$$

$$\sigma_* = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\mu_{\theta_m}^2 + \sigma_{\theta_m}^2\right) - \mu_*^2}, \tag{7}$$

where μ_{θ_m} and σ_{θ_m} are the predicted TWSAs and associated uncertainties of model m and M is the total number of models, namely five in this study. However, due to the specific design of the loss function, σ_{θ_m} cannot be directly estimated by deep learning models. To overcome this issue, we used Monte Carlo simulations. We sampled 20 sets of GRACE inputs randomly on the basis of their uncertainties for each model m to estimate σ_{θ_m}. Ultimately, the ensemble uncertainty σ_* was estimated from five independent deep learning models, with 100 Monte Carlo simulation runs in total. We note that the uncertainties of WGHM simulations and GLDAS inputs are unavailable and not considered. Therefore, the uncertainties reported in this study may be underestimated.

Closure of water balance equation

The closure of the water balance equation is realized by comparing the left side of equation (1) computed from the derivatives of TWSA products and the right side of equation (1) computed from water budget components. We chose the precipitation, evapotranspiration and runoff products from ERA5-Land[45] to provide independent external evaluation since they are not considered in the deep learning model. Moreover, they are proven to agree well with GRACE measurements[22]. To obtain homogeneous time steps for water changes, we interpolated the GRACE and our downscaled TWSAs to the middle of each month using PCHIP[74] interpolation. The GRACE and GRACE-FO eras were dealt with separately to avoid a biased interpolation caused by the gap of 1 year. The TWSCs were obtained by centred finite differences of TWSAs:

$$\mathrm{TWSC}(t) = \frac{\mathrm{TWSA}(t+1) - \mathrm{TWSA}(t-1)}{2\Delta t}, \tag{8}$$

where Δt indicates 1 month. The P, ET and R time series were further smoothed to reduce potential high-frequency artefacts introduced by the finite differences[75]:

$$\tilde{X}(t) = \frac{1}{4}X(t-1) + \frac{1}{2}X(t) + \frac{1}{4}X(t+1), \tag{9}$$

where X denotes P, ET or R. To this end, we received two types of TWSCs: $\mathrm{TWSC}_{\mathrm{GRACE}}$ estimated from GRACE measurements or other TWSA product and $\mathrm{TWSC}_{\mathrm{budget}}$ estimated from P, ET and R. The agreement of these two TWSCs was quantified by computing the NSE[47]:

$$\mathrm{NSE} = 1 - \frac{\frac{1}{T}\sum_{t=1}^{T}\left(\mathrm{TWSC}_{\mathrm{budget}}(t) - \mathrm{TWSC}_{\mathrm{GRACE}}(t)\right)^2}{\frac{1}{T}\sum_{t=1}^{T}\left(\mathrm{TWSC}_{\mathrm{GRACE}}(t) - \overline{\mathrm{TWSC}}_{\mathrm{GRACE}}(t)\right)^2}. \tag{10}$$

TWSA-derived environmental monitoring indices

To demonstrate the applicability of the highly resolved TWSAs, we relied on the concepts of environmental monitoring indices introduced in previous studies, including FPI[4] and DSI[5]. The motivation for FPI is the different capacities of storing water in each cell. We obtained the water deficit $S_{\mathrm{Def}}(t) = S_{\mathrm{Max}} - S(t-1)$ by computing the difference between the water storage of the previous month ($S(t-1)$) and historic maxima S_{Max} for each individual cell. Under the assumption that the high flooding risks are caused by extreme precipitation, we computed the flood potential index $F(t) = P_{\mathrm{Mon}} - S_{\mathrm{Def}}(t)$, where $P_{\mathrm{Mon}}(t) = P(t) \times \mathrm{d}t$ denotes the monthly precipitation (ERA5L in this study). In the end, the FPI was normalized to the range 0 to 1: $\mathrm{FPI} = F(t)/\max[F(t)]$. During this approach, the gaps within the TWSA records may cause incorrect water deficit due to the integral over time for the monthly precipitation. DSI is a standardized metric and also considers the different characteristics of each cell. First, we computed the average values ($\overline{\mathrm{TWSA}}_m$) and standard deviations (σ_m) of each month m over all the years with valid TWSA records. Then, we computed standardized anomalies for the month m and year y as $(\mathrm{TWSA}_{m,y} - \overline{\mathrm{TWSA}}_m)/\sigma_m$. In the end, the standardized anomalies were place in 11 categories on the basis of the thresholds set by ranking percentiles (30%, 20%, 10%, 5% and 2%, two-sided)[76].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The downscaled TWSA product generated in this study is publicly available at https://doi.org/10.3929/ethz-b-000648738 (ref. 77). The raw data used in this study are publicly available. The JPL GRACE mason product is available at https://grace.jpl.nasa.gov/data/get-data/. The WGHM TWSA simulations are available at https://doi.org/10.1594/PANGAEA.948461 (ref. 78). The GLDAS Noah products are available at https://disc.gsfc.nasa.gov/. The ERA5-Land products are available at

https://cds.climate.copernicus.eu. The HydroBasin data are available at https://www.hydrosheds.org/products/hydrobasins. Source data are provided with this paper.

Code availability

The core codes for the deep learning model used including the trained model and weights are available at https://gitlab.ethz.ch/spacegeodesy_public/grace_seda. Further codes for analysing and producing other results are available from the corresponding authors upon request.

References

1. Tapley, B. D. et al. Contributions of GRACE to understanding climate change. *Nat. Clim. Change* **9**, 358–369 (2019).
2. Kim, B.-H. et al. Antarctic ice mass change (2003–2016) jointly estimated by satellite gravimetry and altimetry. *J. Geophys. Res.* **127**, e2021JB023297 (2022).
3. Rodell, M. et al. Emerging trends in global freshwater availability. *Nature* **557**, 651–659 (2018).
4. Reager, J. T. & Famiglietti, J. S. Global terrestrial water storage capacity and flood potential using GRACE. *Geophys. Res. Lett.* https://doi.org/10.1029/2009GL040826 (2009).
5. Zhao, M., Velicogna, I. & Kimball, J. S. Satellite observations of regional drought severity in the continental United States using GRACE-based terrestrial water storage changes. *J. Climate* **30**, 6297–6308 (2017).
6. Zemp, M. et al. *The GCOS 2022 Implementation Plan (GCOS-244)* (WMO, 2022).
7. Bierkens, M. F. Global hydrology 2015: state, trends and directions. *Water Resour. Res.* **51**, 4923–4947 (2015).
8. Scanlon, B. R. et al. Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proc. Natl Acad. Sci. USA* **115**, E1080–E1089 (2018).
9. Tapley, B. D., Bettadpur, S., Watkins, M. & Reigber, C. The gravity recovery and climate experiment: mission overview and early results. *Geophys. Res. Lett.* https://doi.org/10.1029/2004GL019920 (2004).
10. Wahr, J., Swenson, S., Zlotnicki, V. & Velicogna, I. Time-variable gravity from GRACE: first results. *Geophys. Res. Lett.* https://doi.org/10.1029/2004GL019779 (2004).
11. Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F. & Watkins, M. M. GRACE measurements of mass variability in the Earth system. *Science* **305**, 503–505 (2004).
12. Rodell, M. & Reager, J. T. Water cycle science enabled by the GRACE and GRACE-FO satellite missions. *Nat. Water* **1**, 47–59 (2023).
13. Reager, J. et al. A decade of sea level rise slowed by climate-driven hydrology. *Science* **351**, 699–703 (2016).
14. Chen, J. et al. Applications and challenges of GRACE and GRACE follow-on satellite gravimetry. *Surv. Geophys.* https://doi.org/10.1007/s10712-021-09685-x (2022).
15. Wahr, J., Swenson, S. & Velicogna, I. Accuracy of GRACE mass estimates. *Geophys. Res. Lett.* https://doi.org/10.1029/2005GL025305 (2006).
16. Vishwakarma, B. D., Devaraju, B. & Sneeuw, N. What is the spatial resolution of GRACE satellite products for hydrology? *Remote Sens.* **10**, 852 (2018).
17. Wahr, J., Molenaar, M. & Bryan, F. Time variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE. *J. Geophys. Res.* **103**, 30205–30229 (1998).
18. Swenson, S. & Wahr, J. Post-processing removal of correlated errors in GRACE data. *Geophys. Res. Lett.* https://doi.org/10.1029/2005GL025285 (2006).
19. Klees, R., Zapreeva, E., Winsemius, H. & Savenije, H. The bias in GRACE estimates of continental water storage variations. *Hydrol. Earth Syst. Sci.* **11**, 1227–1241 (2007).
20. Velicogna, I. & Wahr, J. Time-variable gravity observations of ice sheet mass balance: precision and limitations of the GRACE satellite data. *Geophys. Res. Lett.* **40**, 3055–3063 (2013).
21. Vishwakarma, B. D., Devaraju, B. & Sneeuw, N. Minimizing the effects of filtering on catchment scale GRACE solutions. *Water Resour. Res.* **52**, 5868–5890 (2016).
22. Lehmann, F., Vishwakarma, B. D. & Bamber, J. How well are we able to close the water budget at the global scale? *Hydrol. Earth Syst. Sci.* **26**, 35–54 (2022).
23. Houborg, R., Rodell, M., Li, B., Reichle, R. & Zaitchik, B. F. Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resour. Res.* https://doi.org/10.1029/2011WR011291 (2012).
24. Eicker, A., Schumacher, M., Kusche, J., Döll, P. & Schmied, H. M. 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**, 1285–1309 (2014).
25. Gemitzi, A., Koutsias, N. & Lakshmi, V. A spatial downscaling methodology for GRACE Total water storage anomalies using GPM IMERG precipitation estimates. *Remote Sens.* **13**, 5149 (2021).
26. Tourian, M. J. et al. A copula-supported Bayesian framework for spatial downscaling of GRACE-derived terrestrial water storage flux. *Remote Sens. Environ.* **295**, 113685 (2023).
27. Vishwakarma, B. D., Zhang, J. & Sneeuw, N. Downscaling GRACE total water storage change using partial least squares regression. *Sci. Data* **8**, 95 (2021).
28. Gerdener, H., Kusche, J., Schulze, K., Döll, P. & Klos, A. The global land water storage data set release 2 (GLWS2.0) derived via assimilating GRACE and GRACE-FO data into a global hydrological model. *J. Geodesy* **97**, 73 (2023).
29. Reichstein, M. et al. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
30. Irrgang, C. et al. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nat. Mach. Intell.* **3**, 667–674 (2021).
31. Yu, S. & Ma, J. Deep learning for geophysics: current and future trends. *Rev. Geophys.* **59**, e2021RG000742 (2021).
32. Sabzehee, F., Amiri-Simkooei, A., Iran-Pour, S., Vishwakarma, B. & Kerachian, R. Enhancing spatial resolution of GRACE-derived groundwater storage anomalies in Urmia catchment using machine learning downscaling methods. *J. Environ. Manag.* **330**, 117180 (2023).
33. Seyoum, W. M., Kwon, D. & Milewski, A. M. Downscaling GRACE TWSA data into high-resolution groundwater level anomaly using machine learning-based models in a glacial aquifer system. *Remote Sens.* **11**, 824 (2019).
34. He, H. et al. Deep learning approaches to spatial downscaling of GRACE terrestrial water storage products using EALCO model over Canada. *Can. J. Remote Sens.* **47**, 657–675 (2021).
35. Ali, S. et al. Constructing high-resolution groundwater drought at spatio-temporal scale using GRACE satellite data based on machine learning in the Indus Basin. *J. Hydrol.* **612**, 128295 (2022).
36. Yin, W. et al. Improving the spatial resolution of GRACE-based groundwater storage estimates using a machine learning algorithm and hydrological model. *Hydrogeol. J.* **30**, 947–963 (2022).
37. Irrgang, C., Saynisch-Wagner, J., Dill, R., Boergens, E. & Thomas, M. Self-validating deep learning for recovering terrestrial water storage from gravity and altimetry measurements. *Geophys. Res. Lett.* **47**, e2020GL089258 (2020).

38. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).
39. Müller Schmied, H. et al. The global water resources and use model WaterGAP v2.2d: model description and evaluation. *Geosci. Model Dev.* **14**, 1037–1079 (2021).
40. Rodell, M. et al. The global land data assimilation system. *Bull. Am. Meteorol. Soc.* **85**, 381–394 (2004).
41. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
42. Lakshminarayanan, B., Pritzel, A. & Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. in *Advances in Neural Information Processing Systems 30* (eds Guyon, I. et al.) 6402–6413 (NIPS, 2017).
43. Boergens, E. et al. Uncertainties of GRACE-based terrestrial water storage anomalies for arbitrary averaging regions. *J. Geophys. Res.* **127**, e2021JB022081 (2022).
44. Zhang, A. T. & Gu, V. X. Global Dam Tracker: a database of more than 35,000 dams with location, catchment and attribute information. *Sci. Data* **10**, 111 (2023).
45. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
46. Lehner, B. & Grill, G. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrol. Proc.* **27**, 2171–2186 (2013).
47. Nash, J. E. & Sutcliffe, J. V. River flow forecasting through conceptual models part I—a discussion of principles. *J. Hydrol.* **10**, 282–290 (1970).
48. Konikow, L. F. Long-term groundwater depletion in the United States. *Groundwater* **53**, 2–9 (2015).
49. Scanlon, B. R. et al. Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. *Proc. Natl Acad. Sci. USA* **109**, 9320–9325 (2012).
50. Asoka, A., Gleeson, T., Wada, Y. & Mishra, V. Relative contribution of monsoon precipitation and pumping to changes in groundwater storage in India. *Nat. Geosci.* **10**, 109–117 (2017).
51. Dangar, S., Asoka, A. & Mishra, V. Causes and implications of groundwater depletion in India: a review. *J. Hydrol.* **596**, 126103 (2021).
52. Vishwakarma, B. D. Monitoring droughts from GRACE. *Front. Environ. Sci.* **8**, 584690 (2020).
53. Watkins, M. M., Wiese, D. N., Yuan, D.-N., Boening, C. & Landerer, F. W. Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *J. Geophys. Res.* **120**, 2648–2671 (2015).
54. Save, H., Bettadpur, S. & Tapley, B. D. High-resolution CSR GRACE RL05 mascons. *J. Geophys. Res.* **121**, 7547–7569 (2016).
55. Luthcke, S. B. et al. Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution. *J. Glaciol.* **59**, 613–631 (2013).
56. Scanlon, B. R. et al. Global evaluation of new GRACE mascon products for hydrologic applications. *Water Resour. Res.* **52**, 9412–9429 (2016).
57. Landerer, F. W. & Cooley, S. S.*Gravity Recovery and Climate Experiment Follow-on (GRACE-FO) Level-3 Data Product User Handbook* (Jet Propulsion Laboratory, California Institute of Technology, 2021).
58. Wiese, D. N., Dah-Ning, Boening, C., Landerer, F. W. & Watkins, M. M. JPL GRACE Mascon Ocean, Ice and Hydrology Equivalent Water Height RL06 CRI Filtered Version 02 (JPL, accessed 27 February 2022); https://doi.org/10.5067/TEMSC-3JC62
59. Cheng, M., Ries, J. C. & Tapley, B. D. Variations of the Earth's figure axis from satellite laser ranging and GRACE. *J. Geophys. Res.* https://doi.org/10.1029/2010JB000850 (2011).
60. Richard Peltier, W., Argus, D. F. & Drummond, R. Comment on "An assessment of the ICE-6G_C (VM5a) glacial isostatic adjustment model" by Purcell et al. *J. Geophys. Res.* **123**, 2019–2028 (2018).
61. Wiese, D. N., Landerer, F. W. & Watkins, M. M. Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resour. Res.* **52**, 7490–7502 (2016).
62. Angermann, D., Pail, R., Seitz, F. & Hugentobler, U. *Mission Earth* (Springer, 2022).
63. Beaudoing, H. & Rodell, M. GLDAS Noah Land Surface Model L4 monthly 0.25×0.25 degree V2.1 (GES DISC, accessed 17 March 2022); https://doi.org/10.5067/SXAVCZFAQLNO
64. Verdin, K. & Verdin, J. A topological system for delineation and codification of the Earth's river basins. *J. Hydrol.* **218**, 1–12 (1999).
65. Lang, N., Jetz, W., Schindler, K. & Wegner, J. D. A high-resolution canopy height model of the Earth. *Nat. Ecol. Evol.* https://doi.org/10.1038/s41559-023-02206-6 (2023).
66. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in *Proc. IEEE Conf. Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016).
67. Ioffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *Proc. 32nd Int. Conf. Machine Learning* Vol. 37 (eds Bach, F. & Blei, D.) 448–456 (PMLR, 2015).
68. Starovoytov, V., Eldarova, E. & Iskakov, K. T. Comparative analysis of the SSIM index and the pearson coefficient as a criterion for image similarity. *Eur. J. Math. Comput. Appl.* **8**, 76–90 (2020).
69. Bourlard, H. & Kamp, Y. Auto-association by multilayer perceptrons and singular value decomposition. *Biol. Cybernet.* **59**, 291–294 (1988).
70. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006).
71. Abadi, M. et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems Software* (tensorflow.org, 2015).
72. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 (2014).
73. Ovadia, Y. et al. Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. *in Advances in Neural Information Processing Systems* **32**, 13991–14002 (2019).
74. Fritsch, F. N. & Carlson, R. E. Monotone piecewise cubic interpolation. *SIAM J. Numeric. Anal.* **17**, 238–246 (1980).
75. Landerer, F. W., Dickey, J. O. & Güntner, A. Terrestrial water budget of the Eurasian pan-Arctic from GRACE satellite measurements during 2003–2009. *J. Geophys. Res.* https://doi.org/10.1029/2010JD014584 (2010).
76. Svoboda, M. et al. The drought monitor. *Bull. Am. Meteorolog. Soc.* **83**, 1181–1190 (2002).
77. Gou, J. & Soja, B. *GRACE-SeDA: A Global Total Water Storage Anomaly Product with a Spatial Resolution of 0.5 Degrees From Self-Supervised Data Assimilation* (ETH Research Collection, 2023); https://doi.org/10.3929/ethz-b-000648738
78. Müller Schmied, H. et al. The global water resources and use model WaterGAP v2.2d—alternative model output driven with gswp3-w5e5. *PANGAEA* https://doi.org/10.1594/PANGAEA.948461 (2022).

Acknowledgements

We thank the Jet Propulsion Laboratory, Goddard Space Flight Center, the WaterGap research group and the HydroSHEDS research group, who provide the valuable data publicly available. Special thanks to H. Müller Schmied, who kindly provided us with support in handling the WGHM data with clear explanations.

Author contributions

J.G. implemented the experiments, did the analysis and prepared the original paper. B.S. proposed the topic, supervised the study

and revised the paper. Both of the authors worked on the theoretical considerations, discussed the results and contributed to the final paper.

Funding
Open access funding provided by Swiss Federal Institute of Technology Zurich.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44221-024-00194-w.

Correspondence and requests for materials should be addressed to Junyang Gou.

Peer review information *Nature Water* thanks the anonymous reviewers for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Junyang Gou

Last updated by author(s): Jan 1, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All the used raw data in this study are publicly available. The links to the data are listed in the Data availability section. No specific software was used.
Data analysis	The deep learning part of this study is conducted using Tensorflow V2.6.0. The data preprocessing and analysis were conducted using Matlab R2022a.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The downscaled TWSA product generated in this study is publicly available at https://doi.org/10.3929/ethz-b-000648738. The raw data used in this study are publicly available. The JPL GRACE mason product is available at https://grace.jpl.nasa.gov/data/get-data/. The WGHM TWSA simulations are available a https://

doi.org/10.1594/PANGAEA.948461. The GLDAS-Noah products are available at https://disc.gsfc.nasa.gov/. The HydroBasin data are available at https://www.hydrosheds.org/products/hydrobasins.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N.A.
Population characteristics	N.A.
Recruitment	N.A.
Ethics oversight	N.A.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	In this study, we proposed a self-supervised deep learning algorithm to assimilate total water storage anomalies (TWSAs) from satellite measurements and hydrological simulations. The achieved model has remarkable generalizability to provide accurate high-resolution TWSAs of 0.5 degrees globally. With the evaluation against other TWSA products and discussion about downstream applications, we proved the quality of our products and the significance of the proposed approach.
Research sample	We focus on the intersecting period of the GRACE mission and WGHM simulations (2002 to 2019) containing 180 months data in total. All the pixels on land are considered as the central pixels to generate a 16 degrees times 16 degrees patch around each central pixel.
Sampling strategy	We use all available data from 2002 to 2019
Data collection	The raw data used in this study are publicly available. The JPL GRACE mason product is available at https://grace.jpl.nasa.gov/data/get-data/. The WGHM TWSA simulations are available at https://doi.org/10.1594/PANGAEA.918447. The GLDAS-Noah products are available at https://disc.gsfc.nasa.gov/. The HydroBasin data are available at https://www.hydrosheds.org/products/hydrobasins. The datasets and models generated and/or analysed during the current study are provided to the editor(s) and reviewers during the review process. The trained model and the generated downscaled TWSAs will be made publicly available upon publication.
Timing and spatial scale	The temporal resolution is approximately monthly, affected by some missing values of GRACE measurements. The spatial resolution is 0.5 degrees times 0.5 degrees.
Data exclusions	The data over Greenland were excluded due to the deficiency of hydrological models.
Reproducibility	The trained model and weights are provided to the editor(s) and reviewers during the review process and will be made publicly available upon publication.
Randomization	Multiple models were trained with different random initialisation to prove the robustness. Moreover, the uncertainties of the ensemble results are reported.
Blinding	Blinding is not relevant for our study as none of such experiments was involved.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging