https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

Open Access Earth System Science Data Discussions

High-resolution global groundwater storage anomalies: A 1 km downscaled dataset

Chen Zhang [1, 3, 5, ★], Yifei Fan [2, 4, 5, ★], Wenxia Han [2, 4, 5, *], Yingchun Ge [1, 3, 5], Jiabei Zhu [1, 3, 5], Tao Che [1, 3, 5, *]

1 State Key Laboratory of Cryospheric Science and Frozen Soil Engineering, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China
2 Key Laboratory of Geology and Environment of Salt Lake, Qinghai Institute of Salt Lakes, Chinese Academy of Sciences, Xining, China
3 Heihe Remote Sensing Experimental Research Station, Key Laboratory of Remote Sensing of Gansu Province, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China
4 Key Laboratory of Green and High-End Utilization of Salt Lake Resources and Qinghai Provincial
5 University of Chinese Academy of Sciences, Beijing 100049, China
★ These authors contributed equally to this work.

Correspondence to: Wenxia Han (wenxia_han@163.com) and Tao Che (chetao@lzb.ac.cn)

Abstract. The global depletion of groundwater poses a significant challenge to water security. However, the coarse spatial resolution of GRACE satellite observations obscures fine-scale hydrological dynamics and limits the separation of localized anthropogenic extraction signals from large-scale climatic forcing. To address this limitation, we present a high-resolution (1 km), continuous monthly dataset of global groundwater storage anomalies (GWSA) covering the period from 2002 to 2020. The dataset is generated using a production framework that integrates Singular Spectrum Analysis (SSA) for temporal gap filling and an aquifer-stratified machine learning approach driven by 19 high-resolution hydroclimatic and geophysical predictors. To ensure robustness and spatial consistency, multiple predictive approaches were evaluated as part of a quality control procedure, and the most stable model was selected for final production based on multi-scale validation. Cross-scale evaluation shows that the downscaled dataset preserves the large-scale spatial patterns of the original GRACE observations with high agreement (R^2 = 0.972, RMSE = 2.10 cm). Independent validation using 1,518 in situ monitoring wells, combined with a geographically stratified specific yield matrix for dimensional conversion, further demonstrates the ability of the dataset to capture long-term groundwater variability across diverse hydrogeological conditions (R^2 = 0.44, $p < 0.01$). The resulting 1

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

km dataset provides enhanced spatial detail and enables the identification of sharp nonlinear boundaries associated with intensive human pumping, as well as spatial polarization patterns in groundwater storage changes. This dataset offers a reliable, observation-constrained resource for water resource assessment, hydrological modeling, and studies of coupled climate and human influences on groundwater systems.

Keywords: Groundwater storage anomaly; GRACE downscaling; Machine learning; High spatiotemporal resolution; Human pumping characteristics

1 Introduction

Global groundwater depletion presents a significant threat to water security. In the context of the accelerating global water cycle, deep aquifer systems are experiencing substantial dynamic evolution (Famiglietti, 2014; Taylor et al., 2013). As the world's largest accessible freshwater resource, groundwater plays a crucial role in mitigating the impacts of climate variability on ecosystems (Rodell et al., 2018; Wada et al., 2010; Weber et al., 2016). Over the past four decades, anthropogenic overextraction has led to the accelerated decline of groundwater levels in multiple regional aquifers globally, resulting in seawater intrusion, land subsidence, water quality deterioration, ecosystem degradation and reduced river discharge. Additionally, global climate change, especially the intensification of drought events, has decreased effective precipitation infiltration and groundwater recharge, while increasing surface evapotranspiration and the reliance of agricultural irrigation on groundwater, thus further exacerbating the depletion risk of groundwater systems. Several large-scale aquifer systems, such as those in the North China Plain, the U.S. High Plains (Ogallala) and northwestern India, are approaching critical depletion thresholds (Rodell et al., 2018; Scanlon et al., 2012). However, conventional groundwater monitoring mainly relies on observation-well networks, which are constrained by limited spatial coverage, time-series gaps caused by insufficient maintenance and barriers to data sharing, making spatiotemporally continuous monitoring challenging.

The Gravity Recovery and Climate Experiment (GRACE), launched in 2002, and its follow-on mission GRACE-FO have supported the continuous retrieval of groundwater storage anomalies (GWSA), emerging as an important approach for monitoring groundwater changes. Nevertheless, due

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

to their coarse spatial resolution, with the native resolution being only 3° × 3°, existing data cannot resolve the significant spatial heterogeneity caused by complex topography, lithological heterogeneity and differences in aquifer structures. Moreover, local scale studies face substantial limitations, particularly the difficulty of establishing reliable connections between GRACE estimates and local ground-based observation sites, which further restricts the accuracy of groundwater-change assessments. Therefore, there is an urgent need to downscale GRACE-derived GWSA.

In Comparison to physically based downscaling, statistical downscaling has attracted increasing attention because of its simple computational framework and strong cross-regional transferability, especially in dealing with nonlinear relationships. Its fundamental assumption is that the statistical relationships between predictor variables, such as hydrological, vegetation and climatic factors, and the target variable remain approximately consistent between the original resolution and the high resolution. On this basis, existing studies have developed a series of GWSA downscaling models, including partial least squares regression (PLS), artificial neural networks (ANN), ground-based scaling factors (SF), and the most widely used tree-based machine learning models, such as random forests (RF) and gradient-boosted trees (XGBoost). However, existing studies are often confined to regional or basin scales (Miro and Famiglietti, 2018; Seyoum et al., 2019), which limits global observation. Moreover, most downscaled products still rely heavily on global hydrological models with coarse spatial resolutions for model training and lack constraints from independent in situ observational benchmarks (Scanlon et al., 2018). This dependence often causes the resulting products to represent only smoothed interpolations of low-frequency background signals, with limited ability to capture the localized characteristics induced by human groundwater pumping.

In this study, we developed a spatiotemporally seamless global 1 km monthly Groundwater Storage Anomaly (GWSA) dataset covering the period from 2002 to 2020. Our framework integrates Singular Spectrum Analysis (SSA) for temporal gap filling (Yi and Sneeuw, 2021) and utilizes multisource 1 km environmental covariates. To ensure the physical consistency and spatial robustness of the dataset, we systematically evaluated representative machine learning (ML) architectures, including deep neural networks (DNN) and XGBoost, as a rigorous quality-control measure (Reichstein et al., 2019; Shen, 2018). We validated our dataset using a comprehensive global

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

observation network of 1,518 in situ monitoring wells (Jasechko et al., 2024) and cross-compared it with GHMs. The results indicate that the downscaled GWSA can provide a spatial resolution capable of effectively identifying local agricultural boundaries, thereby bridging the spatial-scale gap in global assessments and providing a robust, observation-constrained benchmark dataset for global water resource management and the development of regional groundwater digital twins.

2 Data and Methodology

To generate the 1 km global groundwater storage anomaly (GWSA) dataset, a systematic framework composed of five inter-related stages was devised (Fig. 1). Initially, multi-source data integration was implemented. This entailed filling the temporal gaps in 0.25° GRACE/GRACE-FO observations via Singular Spectrum Analysis (SSA) and constructing a 1 km environmental covariate database. Subsequently, coarse-resolution GWSA was inverted by means of a vertical water balance method and refined through water body and glacier masking. To tackle the spatial scale disparity, an aquifer-stratified downscaling model was employed. This model utilized tree-based (XGBoost) and continuous (DNN) architectures, followed by a residual correction mechanism based on cubic spline interpolation. The reconstructed high-resolution product was then subjected to a stringent multi-scale accuracy assessment, including 0.25° internal consistency checks and independent validation against 1,518 in-situ groundwater monitoring stations through a regional specific yield (S_y) matrix. The finalized dataset was utilized to conduct spatiotemporal analysis, characterizing global groundwater dynamics through seasonal contribution analysis and long-term trend detection. The specific theoretical and operational details of each stage are elaborated in the following subsections.

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

Open Access Earth System Science Data Discussions

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

Figure 1. Methodological framework for 1 km GWSA inversion and downscaling.

The framework consists of five main modules: (1) Data Acquisition and Pre-processing of 0.25° input data and 1 km environmental variables; (2) GWSA Inversion based on vertical water balance; (3) Downscaling Modeling & Reconstruction via machine learning (XGBoost and DNN) and residual correction; (4) Accuracy Evaluation through model prediction metrics, spatial comparison, and measured station validation; and (5) Spatiotemporal analysis evaluating both seasonal variations and long-term change trends

2.1 Data Sources

All data supporting the findings of this study are publicly available. This research utilized the GRACE/GRACE-FO mass concentration (mascon) solutions (RL0603) provided by the University of Texas Space Research Center (CSR) and the GLDAS-2.1 product from the NASA Global Land Data Assimilation System Version 2 (GLDAS-2) as the primary target variables for the downscaling modeling process. Additionally, a spatiotemporal unified database of environmental variables with a global 1 km resolution was established in this study. This database encompasses meteorological, hydrological, topographic and soil, and biophysical variables, along with the cumulative fluxes of certain variables.

Meteorological variables include temperature (TAS) and precipitation (PRE); hydrological variables consist of soil moisture (SM) and actual surface evapotranspiration (ET); topographic and soil attribute variables are composed of global mean elevation (DEM), proportion of clay particles (Clay), proportion of silt particles (Silt), proportion of sand particles (Sand), coarse fragments (CFVO), cation exchange capacity (CEC), soil pH, soil organic carbon content (SOC), and total nitrogen (TN); biophysical variables involve normalized vegetation index (NDVI) and land surface temperature (LST). These variables serve as crucial predictive factors for facilitating the refined mapping of global groundwater reserve anomalies. All data employed in model construction are listed, and detailed information on datasets is presented in Table 1.

Although GRACE/GRACE-FO mascon solutions are available through November 2025, the temporal coverage of the downscaled GWSA dataset was limited to the period from April 2002 to

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

December 2020. This temporal restriction mainly reflects the availability of the high-resolution environmental covariates used for model training, particularly the meteorological variables.

Table 1. Summary of datasets and sources for downscaling model.

Variable	Spatial resolution	Temporal resolution	Temporal coverage	Data source
CSR GRACE/ GRACE-FO	0.25°	Monthly	Apr 2002–Nov 2025	https://www2.csr.utexas.edu/grace
GLDAS-2.1	0.25°	Monthly	Apr 2002–Dec 2020	https://disc.gsfc.nasa.gov/datasets
Climate data (TAS, PRE)	1000m	Monthly	Apr 2002–Dec 2020	(Karger et al., 2017)
ET	1000m	8-day	Apr 2002–Dec 2020	(Zhang et al., 2019)
SM	1000m	Monthly	Apr 2002–Dec 2020	(Zheng et al., 2023)
DEM	1000m	–	2010	https://www.usgs.gov/coastal-changes-and-impacts/gmted2010
Soil properties (Clay, Silt, Sand, CFVO, pH, CEC, SOC, TN)	250m	–	2020	(Poggio et al., 2021)
NDVI	500m	16-day	Apr 2002–Dec 2020	https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD13A1
LST	1000m	8-day	Apr 2002–Dec 2020	https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD11A2

2.1.1 GRACE Terrestrial Water Storage

Our research employed the GRACE Mascon solutions provided by CSR, encompassing monthly Terrestrial Water Storage Anomaly (TWSA) sequences from April 2002 to November 2025. The reflected TWS anomalies were measured relative to the average baseline spanning from 2004 to 2009. Prior studies have indicated that Mascon solutions present substantial enhancements over spherical harmonics in terms of leakage errors and post-processing (Palazzoli et al., 2025; Scanlon et al., 2016). They do not necessitate additional smoothing, empirical de-striping, filtering, or scaling factors, and are more user-friendly (Save et al., 2016; Wang et al., 2024). The newly released RL06 Mascon solutions adopt a novel grid structure founded on equidistant geodetic grids with a spatial resolution of 1°, which is resampled to a resolution of 0.25° for publication. This effectively mitigates the leakage between land and ocean signals. In this investigation, we utilized Singular Spectrum Analysis (SSA) to fill the 35-month data gaps. In combination with the Global Land Data Assimilation System (GLDAS), we constructed the Groundwater Storage Anomaly (GWSA) sequence. Moreover, we applied machine learning and deep learning methods to downscale the spatial resolution to 0.008333° (approximately 1

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

152 km at the equator), with the aim of attaining globally seamless high-resolution GWSA mapping with
153 spatiotemporal resolution.

154 **2.1.2 GLDAS Noah V2.1**

155 The National Aeronautics and Space Administration (NASA) Global Land Data Assimilation
156 System (GLDAS) is designed to integrate satellite and ground-based observational data. By leveraging
157 land surface modeling and data assimilation techniques, it aims to attain optimal simulations of land
158 surface conditions and flux fields (Rodell et al., 2004). It encompasses three land surface models: the
159 Noah model, the Catchment (CLSM) model, and the Variable Infiltration Capacity (VIC) model.

160 The Noah land surface model includes four soil temperature and moisture layers, along with
161 models for vegetation canopy, snow cover prediction, evapotranspiration, and soil drainage runoff.
162 Owing to its precise representation of surface energy and water budgets, it has been extensively
163 employed in global hydrological and water resource change research (Campbell et al., 2019; Ek et al.,
164 2003; Niu et al., 2011).

165 In this study, the annual time-series data of soil water storage (SWS), snow water equivalent
166 (SWE), and vegetation canopy water storage (CWS) provided by Noah V2.1 were utilized to calculate
167 the groundwater storage anomaly (GWSA). All input data were transformed into anomalies relative to
168 the average baseline from 2004 to 2009 (Wang et al., 2024; Zhang et al., 2021) to ensure consistency
169 with the definitions of the Gravity Recovery and Climate Experiment (GRACE).

170 **2.1.3 Predictor Variables**

171 The surface hydrological cycle demonstrates intricate dynamic variations governed by water,
172 energy, and biogeochemical fluxes. For example, the equilibrium of terrestrial water storage (TWS) is
173 affected by anomalous changes in constituent components across diverse climatic zones (Girotto &
174 Rodell, 2019). Furthermore, terrestrial heat and TWSA themselves can influence the predictive
175 precision of GLDAS (Campbell et al., 2023). To comprehensively consider various influencing factors,
176 19 variables closely associated with climate, hydrology, and vegetation dynamics were selected to
177 construct a global environmental variable database at a 1 km scale as modeling predictors. This dataset
178 encompasses 8 monthly-scale dynamic variables, 9 multi-year average variables, and two additional

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

temporal coding variables (derived from sine and cosine transformations of monthly values to capture seasonal characteristics).

Monthly average near-surface air temperature (TAS) and monthly cumulative precipitation (PRE) data are obtained from Climatologies at High resolution for the Earth's Land Surface Areas (CHELSA). This dataset, based on ERA climatic reanalysis and terrain correction algorithms, offers global meteorological forcing data with a spatial resolution of 1 km at the monthly scale, effectively supporting high-precision model simulations at local scales (Karger et al., 2017).

The evapotranspiration (ET) data are sourced from the PML_V2 high-resolution dataset (Zhang et al., 2019), which provides a spatial resolution of 500 m and a temporal resolution of 8 days for ET measurements along with three key independent components (i.e., vegetation transpiration, soil direct evaporation, and canopy interception evaporation). These data have undergone cross-validation in the Global Flux Network (FLUXNET) and exhibit robust representation of surface water consumption patterns. In this study, monthly synthetic ET was generated by rescaling the 8-day resolution data according to temporal scales, followed by cumulative ET calculations to obtain monthly cumulative evapotranspiration (ETC) values since April 2002.

Soil moisture (SM) was derived from the Global Daily Surface Soil Moisture Dataset (Zheng et al., 2023), which was jointly inverted using ERA5 reanalysis data and ESA-CCI SSM products. This dataset provides seamless global surface soil moisture data with a daily resolution of 1 km for the period 2000–2021. The monthly-scale product was selected.

The DEM is derived from the Global Multi-resolution Topographic Elevation Dataset (GMTED2010) provided by the United States Geological Survey (USGS). This study specifically extracts the spatial mean (Mean) product from this dataset, aiming to precisely characterize the average elevation variation of the underlying surface at the target grid scale.

The soil attributes are derived from the High Precision Global Soil Information System (SoilGrids) published by the International Soil Reference Information Centre (ISRIC). This dataset is generated through machine learning algorithms coupled with multi-source environmental covariate physical coupling inversion, providing global-scale products at a resolution of 250 m. The study extracts variables including Proportion of clay particles (Clay), Proportion of silt particles (Silt), Proportion of

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

sand particles (Sand), Coarse fragments (CFVO), Cation Exchange Capacity (CEC), Soil pH, Soil organic carbon content (SOC), and Total nitrogen (TN) from depths of 0–30 cm to reflect surface soil (or topsoil) characteristics.

NDVI and LST were derived from MOD13A1 and MOD11A2 products, respectively. This study conducted quality validation based on the GEE platform and synthesized monthly average data for the period 2002-2020.

2.2 Singular Spectrum Analysis for Data Gap Filling

Singular Spectrum Analysis (SSA) is a technique for processing nonlinear time series (Vautard et al., 1992). Through the decomposition, sorting, and reconstruction of the trajectory matrix of the target time series, SSA can extract trends and noise from the data and has been extensively utilized for missing value imputation (Kondrashov and Ghil, 2006; Wang et al., 2021). In the present study, SSA was utilized to impute 22 discrete short gaps that occurred prior to June 2017 during the GRACE mission period, which were caused by periodic power system outages (defined as SSA-filling-a). Additionally, it was used to impute 11 consecutive months of missing data after July 2017 between the GRACE/GRACE-FO mission periods and two months of missing data within the GRACE-FO mission period (collectively defined as SSA-filling-b). This approach aims to minimize the influence of missing data on subsequent downscaling modeling and temporal analysis. To ensure the accuracy of SSA reconstruction, the full available GRACE/GRACE-FO monthly record from April 2002 to November 2025 was used for the SSA calculation. The SSA calculation method is as follows:

First, define the time series for uniform $\boldsymbol{X}$ sampling:

$$\boldsymbol{X}=[x_1, x_2, \ldots, x_N] \quad (1)$$

where x_i $(i = 1, 2, \ldots, N)$ represents the monthly GRACE observation values, where missing months were initialized with zero as starting values for the SSA iteration. N represents the sample size from April 2002 to November 2025. For the periods of October 2011 and April 2015, repeated observations were averaged, hence $N = 284$.

Define the trajectory matrix $\mathbf{Y}$ for time series $\boldsymbol{X}$:

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

Open Access Earth System Science Data Discussions

$$\mathbf{Y} = \begin{bmatrix} x_1 & x_2 & \cdots & x_L \\ x_2 & x_3 & \cdots & x_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_M & x_{M+1} & \cdots & x_N \end{bmatrix} \tag{2}$$

$$L = N - M + 1 \tag{3}$$

$$K = \min(M, L) \tag{4}$$

where L represents the number of columns in the trajectory matrix; M represents the window width; and K represents the maximum modal number decomposable by the trajectory matrix (Eq. (4)). This research employed the optimization parameter search scheme proposed by Yi and Sneeuw (Yi & Sneeuw, 2021), with $M = 24$ and $K = 12$ for SSA-filling-a, and $M = 48$ with K ranging from 1 to 12 for SSA-filling-b. These parameters were ascertained through iterative procedures and cross-validation. Conduct singular value decomposition on the trajectory matrix $\mathbf{Y}$:

$$\mathbf{Y} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^{\mathrm{T}} \tag{5}$$

$$\boldsymbol{e}_i = \mathbf{U}_{(:,i)}\sigma_i \tag{6}$$

$$\boldsymbol{p}_i = \mathbf{V}_{(:,i)} \tag{7}$$

$$\mathbf{Z}_i = \boldsymbol{e}_i \times \boldsymbol{p}_i^{\mathrm{T}} \tag{8}$$

$$\mathbf{Y} = \sum_{i=1}^{K} \mathbf{Z}_i \tag{9}$$

where $\mathbf{\Sigma}$ represents the singular value diagonal matrix; σ_i represents the ith singular value; $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices; $\boldsymbol{e}_i$ represents the ith empirical orthogonal function; $\boldsymbol{p}_i$ represents the ith principal component; and $\mathbf{Z}_i$ represents the ith modal matrix. The trajectory matrix $\mathbf{Y}$ can be expressed as the sum of the K modal matrices $\mathbf{Z}_i$.

The reconstructed component sequence $\boldsymbol{\vartheta}_i$ was obtained by diagonal averaging of each modal matrix $\mathbf{Z}_i$, and the reconstructed time series $\boldsymbol{X}'$ was calculated as follows:

$$\vartheta_i^p = \mathrm{mean}(Z_i^{j,k}),\ (j + k = p + 1) \tag{10}$$

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

Open Access Earth System Science Data Discussions

$$\boldsymbol{X}' = \sum_{i=1}^{K'} \boldsymbol{\vartheta_i} \tag{11}$$

where ϑ_i^p represents the value of the ith reconstructed component at the pth position of the time series; $\boldsymbol{\vartheta_i}$ represents the complete reconstructed component sequence composed of ϑ_i^p, calculated as the average of the skew-diagonal elements of each modal matrix $\mathbf{Z}_i$; p represents the position index of the time series ($p = 1, 2, \ldots, N$); j and k represent the row and column numbers of $\mathbf{Z}_i$, respectively; K' represents the actual number of retained components; and $\boldsymbol{X}'$ represents the reconstructed time series. Finally, a two-layer iterative approach is employed for $\boldsymbol{X}$ interpolation. In the outer iteration, K' is incrementally increased from 1 to determine the optimal number of modes. When $K' = K$, the sum of all reconstructed components equals $\boldsymbol{X}$. In the inner iteration, $\boldsymbol{X}'$ is continuously calculated for each fixed K' until the difference between consecutive reconstruction results ($\boldsymbol{X}'$) falls below a predefined threshold, indicating stabilization of the SSA results. This stabilized state is then adopted as the optimal reconstruction outcome for the current K' value. Additionally, further details on this methodology can be referred to the paper by Yi and Sneeuw (Yi & Sneeuw, 2021).

The uncertainty of the SSA reconstruction was evaluated using two error metrics, ε_1 from the SSA-filling-a phase and ε_2 from the SSA-filling-b phase:

$$\varepsilon_1 = \mathrm{Std}(\boldsymbol{X}_{\mathrm{obs}} - \widehat{\boldsymbol{X}}) \tag{12}$$

$$\varepsilon_2 = \min \sqrt{\frac{1}{12}\sum \varepsilon^2} \tag{13}$$

$$\varepsilon = \sqrt{\frac{1}{N}\sum (\boldsymbol{X}_{\mathrm{obs}} - \widehat{\boldsymbol{X}})^2} \tag{14}$$

Both SSA filling and error evaluation were carried out on a pixel-by-pixel basis for the original Gravity Recovery and Climate Experiment (GRACE) data grid. In the SSA-filling-a phase, the reconstruction error was defined as the standard deviation of the discrepancy between the observed sequences from non-missing months and the SSA-reconstructed sequences. Here, $\boldsymbol{X}_{\mathrm{obs}}$ and $\widehat{\boldsymbol{X}}$ represent the original observation sequences for all non-missing months within the GRACE mission period and the reconstructed sequences obtained by applying the SSA-filling-a algorithm to the

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

corresponding months, respectively; ε_1 represents the reconstruction uncertainty in this phase, with larger values indicating lower reconstruction accuracy.

In the SSA-filling-b phase, annual masking was applied to the original observation series using 2003–2016 as the cross-validation background, and each year from 2004 to 2015 was sequentially masked to form 12 annual gap-filling experiments. For each masked year, the root mean square error (*RMSE*), denoted as ε, was calculated between the original observations and the reconstructed values using Eq. (14). The 12 annual *RMSE* values were then aggregated using Eq. (13). Subsequently, a sequential search was conducted within the range of the maximum modal number K (ranging from 1 to 12), and the minimum aggregated *RMSE* was defined as the fitting error for this phase. The corresponding combination of the window width M and the maximum modal number K was identified as the optimal fitting parameters for this phase. Here, $\boldsymbol{X}_{\text{obs}}$ and $\widehat{\boldsymbol{X}}$ represent the original observation series of a masked year and the reconstructed series obtained by applying the SSA-filling-b algorithm to that year, respectively; N represents the number of non-missing months in the original GRACE data for that year, signifying the actual sequence length; ε_2 represents the minimum aggregated *RMSE* from the 12 annual masking experiments, with larger values indicating lower fitting accuracy.

2.3 GWSA Inversion and Masking Strategy

Terrestrial Water Storage (TWS) represents the aggregate water reserves both above and beneath the land surface, encompassing surface water, soil moisture, groundwater, snow cover, ice, and vegetation water storage (Girotto & Rodell, 2019). Through the utilization of a vertical water model in conjunction with GRACE and GLDAS data, anomalies in groundwater storage (GWSA) can be segregated from TWSA anomalies via the subsequent calculation approach:

$$\Delta GWS = \Delta TWS - (\Delta SMS + \Delta SWE + \Delta CWS) \quad (15)$$

The data of soil moisture storage (SMS), snow water equivalent (SWE), and canopy water storage (CWS) were obtained from Noah V2.1, where Δ represents the anomalies relative to the baseline period (2004-2009). The anomalies in terrestrial water storage (ΔTWS) were derived from the Gravity Recovery and Climate Experiment (GRACE) data, while ΔGWS represents the anomalies in the isolated groundwater storage (GWSA). We calculated the observed GWSA sequence ($GWSA_{obs}$;

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

290 $N = 249$) from the original GRACE data, and the reconstructed GWSA sequence ($GWSA_{SSA}$; $N = 284$)
291 from the SSA-reconstructed GRACE data. The two sequences were fully spatially aligned and directly
292 comparable at coincident time points. To validate the reliability of the reconstruction, the observed
293 GWSA sequence was kept intact, and only for the months with missing GRACE data, the $GWSA_{SSA}$
294 data ($N = 35$) were used to replace it. This alternative sequence was then used as the benchmark for
295 subsequent reconstruction and precision evaluation.

296 It is worth noting that this formula still has certain limitations. For example, in polar and alpine
297 regions, TWSA is primarily governed by mass changes from intense iceberg discharge, submarine
298 melting, and melt-water runoff. In these areas, TWSA dynamics are not solely driven by snowfall,
299 precipitation, or evapotranspiration (Chen et al., 2007; Mankoff et al., 2020; Rodell & Famiglietti,
300 2001). In tropical regions, the anomalies in surface water reserves, including those in rivers, inland
301 water bodies, reservoirs, wetlands, and floodplains, play a dominant role in the changes of TWSA,
302 which may introduce uncertainties into the inversion of the Groundwater Storage Anomaly (GWSA).
303 Therefore, this study employed rigorous glacier and water-body masking strategies to minimize the
304 errors in the global GWSA estimation.

305 **2.3.1 Permanent Water Body Mask**

306 To minimize signal leakage errors induced by permanent water bodies and maximize the
307 preservation of coastal terrestrial information, this study employed the Global Surface Water Explorer
308 (GSWE) dataset furnished by the European Commission Joint Research Centre (JRC) (Pekel et al.,
309 2016). This dataset documents the spatiotemporal dynamics of global surface water bodies at a spatial
310 resolution of 30 m. Regions with pixel values of 3 (Permanent water) were extracted from its Yearly
311 Classification product as reliable water body benchmarks. Pixel-wise set union processing was
312 conducted on the temporal classification results spanning from 2002 to 2020, and resampling to a 1 km
313 resolution was carried out, ultimately generating a unified global permanent water body range mask.

314 **2.3.2 Glacier and Ice Sheet Mask**

315 The global mountain glacier mask was established by utilizing data from The Randolph Glacier
316 Inventory (RGI) supplied by the National Snow and Ice Data Center (NSIDC)

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

(https://nsidc.org/data/nsidc-0770/versions/7#anchor-documentation). This dataset encompasses single-contour vector maps of each glacier dating back to approximately 2000. These maps were integrated within ArcGIS Pro 3.5.4 to demarcate the global mountain glacier extent.

The Greenland land ice extent was obtained from the MEaSUREs Greenland Ice Mapping Project (GIMP) Land Ice and Ocean Classification Mask version 1 data provided by NSIDC (https://nsidc.org/data/nsidc-0714/versions/1), making use of the 30m spatial resolution classification data from the GimpIceMask_30m_2015 product.

Ultimately, the two masks were combined into a unified global glacier mask, which was then merged with the water body mask to generate the baseline land ice extent.

2.4 Machine Learning Downscaling Framework

The fundamental premise of the downscaling model posits that the statistical correlation between the target variable (Groundwater Storage Anomaly, GWSA) and the independent variables remains invariant across diverse spatial scales. Building upon this premise, we constructed the model utilizing lower-resolution (0.25°) data spanning from 2002 to 2020, while leveraging higher-resolution (1 km) data for prediction purposes. The specific methodology is delineated as follows:

(1) Data pre-processing: Initially, all 19 input variables with a 1 km resolution were subjected to masking processing to ensure complete congruence with the reference mask. Scattered missing values were imputed using inverse distance weighting, ultimately generating a spatiotemporally continuous 1 km environmental variable database. Subsequently, the 19 variables were resampled to a 0.25° resolution and aligned with the GWSA grid, a process accomplished through mean aggregation via the gdal.Warp function.

(2) Model development: Given the substantial spatial heterogeneity of hydrogeological conditions and Groundwater Storage Anomalies (GWSA) at the global scale, it is challenging to formulate a unified model that reflects the overall scenario. Therefore, separate models were developed for each aquifer type. After rasterizing the aquifer data, each pixel was assigned a specific aquifer category. First, 200,000 samples were randomly selected from each aquifer to optimize the model parameters, and then the models were trained using all samples from each aquifer category. Random Forest (RF) and Deep

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

Neural Network (DNN) models were constructed, yielding lower-resolution training outcomes to assess model accuracy.

(3) Prediction and correction: The 1 km environmental variable database was inputted to obtain high-resolution global GWSA reconstruction results. These results were resampled to a lower resolution, and the residuals between them and the $GWSA_{obs}$ or $GWSA_{SSA}$ sequences were calculated. The low-resolution residuals were then interpolated to a 1 km resolution using cubic spline interpolation and added to the reconstructed results to implement residual correction.

(4) Validation: This mainly encompasses two aspects. First, the reconstructed results were compared with the GWSA inversion results at the original spatial resolution (0.25°). Second, the high-resolution reconstructed results (0.008333°) were compared with the extracted in-situ observational data to systematically evaluate the overall accuracy of the downscaling products.

2.4.1 Machine Learning Architectures and Model Configuration

To capture the intricate nonlinear relationships between groundwater storage anomaly (GWSA) and multi-source hydro-climatic variables, two representative machine-learning architectures were employed: Extreme Gradient Boosting (XGBoost) and Deep Neural Networks (DNN). XGBoost, a scalable ensemble grounded in the gradient-boosting decision tree algorithm (Chen & Guestrin, 2016), demonstrates superiority in establishing statistical mappings for GWSA downscaling (Ali et al., 2023; Zhang et al., 2021). It effectively mitigates multicollinearity and overfitting through the randomization of samples and variables (Chen et al., 2020; Khan et al., 2024). Complementarily, DNNs leverage interconnected hierarchical layers to automatically extract deep-level variable interactions from extensive datasets (LeCun et al., 2015; Shen, 2018). The high model capacity of DNNs effectively compensates for the potential limitations of tree-based decision spaces in smoothing nonlinear surfaces, especially under extreme hydrological forcing conditions (Reichstein et al., 2019; Shen, 2018).

To account for global hydrogeological heterogeneity, both models were independently trained for each aquifer category. For each aquifer, 200,000 pixels were randomly sampled. Seventy percent of the data was allocated for training and hyperparameter optimization, while the remaining 30% was reserved for validation (e.g., calculating the root-mean-square error (RMSE)). For XGBoost, the optimal hyperparameters, including *n_estimators, learning_rate, max_depth*, and *min_child_weight*,

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

were identified via the *RandomizedSearchCV* method. For the DNN, a Multi-Layer Perceptron (MLP) architecture was constructed, and a Robust Scaler was applied to both input and target variables to eliminate interference from extreme climate outliers (Shen, 2018). The optimal network configuration (3–6 layers and 128–768 units) was autonomously determined using the Optuna optimization framework (Akiba et al., 2019). The DNN was trained using the Adam optimizer with a mean squared error (MSE) loss function, incorporating an Early Stopping mechanism to prevent over-fitting on the large-scale spatiotemporal dataset.

2.5 Model Accuracy Evaluation

The model evaluation procedure consists of three crucial elements. Firstly, by utilizing the 30% test samples set aside during the training phase, we measured the prediction accuracy of the model across diverse aquifer zones via evaluation metrics and carried out an analysis of the relative significance of feature variables. Secondly, cross-validation was carried out between the reconstructed outcomes and the original Groundwater Storage Anomaly (GWSA) data at the original spatial resolution (0.25°) to quantify the deviations from the baseline data. Finally, the high-resolution reconstruction results (0.008333°) were verified using global long-term groundwater monitoring well data (covering the period from 2006 to 2020) to systematically evaluate the overall accuracy of the downscaling products.

2.5.1 Statistical Evaluation Indicators

We assess the downscaled model through three metrics: the coefficient of determination (R^2), the correlation coefficient (CC), and the root mean square error ($RMSE$). The coefficient of determination (R^2) denotes the model's explanatory capacity for the variance of the dependent variable, signifying the proportion of the total variance that can be ascribed to the model. The correlation coefficient (CC) gauges the intensity of the linear correlation between the predicted and observed values, whereas the root mean square error ($RMSE$) quantifies the overall deviation between the predicted and observed results. The calculation methods are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - o_i)^2}{\sum_{i=1}^{n}(o_i - \bar{o})^2} \tag{16}$$

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

Open Access Earth System Science Data Discussions

$$CC = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(o_i - \bar{o})^2}} \quad (17)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(o_i - y_i)^2}{n}} \quad (18)$$

where o represents the observed value; y represents the predicted value; n represents the sample size; and $\bar{o}$ and $\bar{y}$ are the means of o and y respectively.

2.5.2 In Situ Cross-Validation and Dimensional Conversion

To objectively evaluate the physical reliability of the 1 km downscaled global groundwater storage anomaly (GWSA) product, we collected and integrated in situ groundwater records from 1,518 monitoring wells with long-term continuous observations worldwide for the period 2006–2020. This in situ dataset integrates multiple high-quality observation networks, including national and regional monitoring stations reported in the China Groundwater Yearbook and the latest global groundwater-level database compiled by Jasechko et al. (2024). To suppress random high-frequency noise at the individual-well scale caused by local geological heterogeneity or irregular pumping, all valid well time series within the same 0.25° grid cell were arithmetically averaged to represent the regional groundwater-level anomaly dynamics at that grid scale (groundwater level anomaly, GWLA). However, the equivalent water height (EWH) anomalies derived from satellite gravimetry and the in-situ groundwater-level anomalies (GWLA) differ in physical dimension, and their relationship can be expressed as:

$$\Delta GWS = \Delta GWL \times S_y \quad (19)$$

Where ΔGWS and ΔGWL represent the anomaly change in groundwater storage and the anomaly change in observed groundwater level, respectively, both expressed in cm; S_y is the specific yield, which is used to convert between the two quantities and ensure unit consistency. Notably, given the strong heterogeneity of deep aquifer systems at the global macroscale, directly introducing high-resolution spatial grids of S_y with substantial uncertainty would likely introduce spurious local-variation noise during the validation process. Therefore, following widely adopted standards and robust evaluation practices in large-scale GRACE-based groundwater studies (Famiglietti et al., 2011;

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

Rodell et al., 2009; Scanlon et al., 2012), we constructed a geographically zoned dynamic specific-yield parameter matrix based on the hydrogeological settings of the major continental regions and representative aquifer systems worldwide (Table 2). Representative regional empirical mean S_y values ranging from 0.08 to 0.18 were then assigned to different latitude–longitude domains to linearly scale the observed groundwater-level trends, thereby enabling an independent validation of the multi-year trends in the reconstructed GWSA.

Table 2. Regional Specific Yield (S_y) parameter matrix.

Region (Inferred)	Latitude Range	Longitude Range	Specific Yield (S_y)
Asia	10°N to 80°N	60°E to 150°E	0.11
Europe & Middle East	35°N to 75°N	25°W to 60°E	0.14
North America	15°N to 75°N	170°W to 50°W	0.16
South America	60°S to 15°N	90°W to 30°W	0.18
Africa	35°S to 35°N	20°W to 55°E	0.08
Australia / Oceania	50°S to 0°	110°E to 180°E	0.09
Global Default (Other regions)	-	-	0.15

2.6 Seasonal Analysis

Hydrological time-series are generally affected by robust periodic seasonal rhythms, and seasonality molds the dynamics of the hydrosphere on a global scale. Seasonal-induced extreme hydrological anomalies can even notably counterbalance or intensify the effects of long-term trends (Cooley et al., 2021; Li et al., 2025). Nevertheless, research in this domain frequently neglects these aspects. Consequently, we employed the methodology put forward by Li et al. (2025), leveraging reconstructed groundwater storage anomaly (GWSA) sequences with high spatiotemporal resolution to quantify the seasonal contributions to global groundwater dynamics at the pixel scale.

$$C_{seasonality} = \frac{\bar{\sigma}}{\bar{\sigma}+\delta} \tag{20}$$

where $\bar{\sigma}$ and δ represent the annual standard deviation of GWSA and the standard deviation of the multi-year average of GWSA for the period 2002–2020, respectively. $C_{\text{seasonality}}$ represents the relative seasonal contribution, with a value range of 0% to 100%. A value above 50% signifies that seasonality dominates the groundwater dynamics at that pixel.

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

2.7 Long-term Trend Analysis

To precisely extract the long-term evolution characteristics of Groundwater Storage Anomaly (GWSA), this study utilized a seasonal-trend decomposition approach based on local weighted regression (STL, Seasonal-Trend decomposition using LOESS) to conduct de-seasonalization of the original time series. STL is a commonly employed non-parametric estimation method. Its fundamental premise is that time series are composed of seasonal components, long-term trends, and residuals (Cleveland et al., 1990). By separating the seasonal term $\boldsymbol{S_t}$ from the original time series $\boldsymbol{Y_t}$, the de-seasonalized sequence $\boldsymbol{X_t}$ is acquired. In comparison with traditional linear regression based on the least-squares method, its non-parametric slope calculation exhibits greater robustness towards missing data and outliers in the time series (Sen, 1968; Theil, 1950).

$$\boldsymbol{X_t = Y_t - S_t} \tag{21}$$

where the seasonally adjusted $\boldsymbol{X_t}$ series comprehensively preserves both the trend component signal of the groundwater storage anomaly (GWSA) and the natural residual component signal, which encompasses extreme hydrological events. On this basis, the Theil-Sen Median Trend method is utilized to ascertain the long-term variation trend. The calculation method is as follows:

$$\beta = \text{Median}\left(\frac{X_j - X_i}{j - i}\right), \quad \forall j > i \tag{22}$$

where X_j and X_i represent the deseasonalized GWSA values at months j and i, respectively; β is the median slope of all pairwise combinations in the time series; $\beta > 0$ indicates an increasing trend in groundwater storage, whereas $\beta < 0$ indicates a decreasing trend.

The Mann-Kendall Significance Test serves as a crucial complement to Theil-Sen trend estimation and is a non-parametric test method recommended by the World Meteorological Organization (WMO). Its advantage lies in not requiring sample data to follow a specific distribution, making it widely applicable for trend significance testing in hydro-meteorological series (Kendall, 1975; Mann, 1945). The calculation methodology can be referenced in (Wang et al., 2022). This study conducted trend tests for GWSA changes at significance levels of $p < 0.05$ and $p < 0.01$.

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

3 Results

3.1 Model Performance and Multi-scale Validation

3.1.1 Temporal Continuity and SSA-based Reconstruction Quality

To ensure the spatiotemporal continuity of the downscaling baseline without relying on artificial annual averaging, Singular Spectrum Analysis (SSA) was employed to reconstruct missing GRACE observations. The reconstruction quality was evaluated using two error metrics corresponding to the SSA-filling-a and SSA-filling-b phases (Fig. 2). Global evaluation showed that the SSA framework successfully maintained low reconstruction errors while preserving the integrity of the original temporal dynamics across two distinct gap-filling phases.

During the SSA-filling-a phase, the reconstruction error was quantified as the standard deviation of residuals between observed and reconstructed values over all non-missing months. Globally, the error ranged from 0.232 to 62.458 cm, with a tightly constrained median of 1.156 cm. The 1% and 99% percentiles were 0.475 and 7.550 cm, respectively, indicating robust reconstruction precision during this stage.

During the SSA-filling-b phase, which was designed to evaluate reconstruction performance under long-term continuous data gaps, the global median error was 1.679 cm, with the 1% and 99% percentiles of 0.732 and 10.204 cm, respectively. Over terrestrial regions excluding Antarctica, the median error increased to 2.708 cm, with the 1% and 99% percentiles of 0.862 and 15.295 cm, respectively. Although reconstruction errors over land were slightly higher than the global average, they remained low and within an acceptable range.

Spatially, the high-error centers across both phases exhibited high consistency, primarily clustering in the high-latitude cryosphere of the Northern Hemisphere, the Amazon Basin, southern South America, East Africa, and Southeast Asia. Although uncertainties in these specific regions are relatively higher, these variations are primarily driven by large natural signal amplitudes, significant seasonal oscillations, and complex local processes rather than algorithmic instability. Consequently, the SSA reconstruction provides a physically reliable, continuous temporal baseline for the subsequent high-resolution downscaling.

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

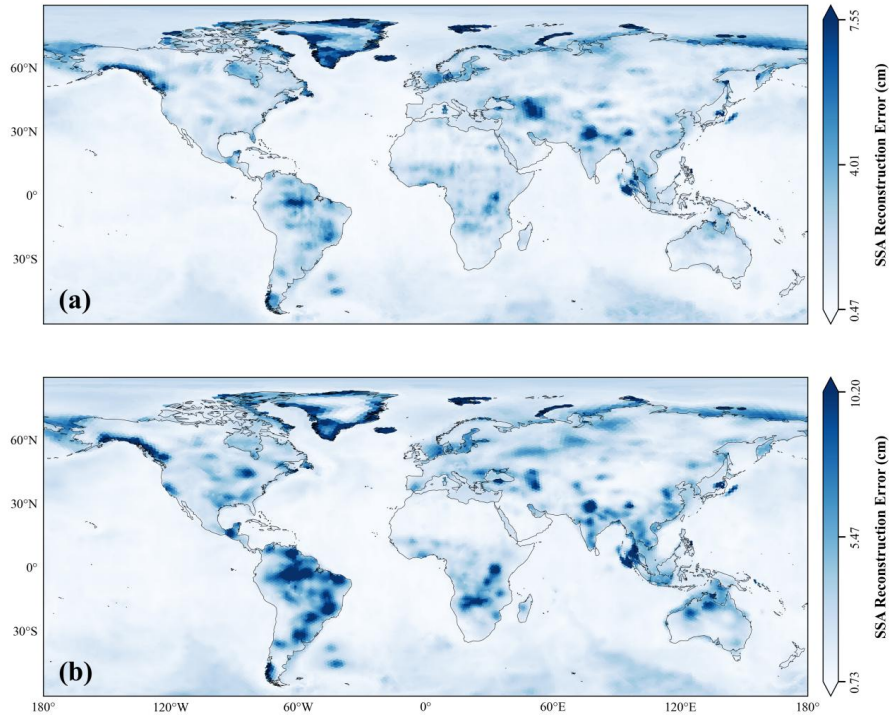


Figure 2. Evaluation of the reconstruction accuracy of Singular Spectrum Analysis (SSA). (a) Reconstruction error in the SSA-filling-a phase (cm), quantified as the standard deviation of residuals between observed and reconstructed values. (b) Reconstruction error in the SSA-filling-b phase (cm), quantified as the minimum aggregated root mean square error (*RMSE*) from annual masking experiments.

3.1.2 Internal Consistency Across Global Aquifer Systems

To guarantee the robustness and spatial consistency of the final dataset across various hydro-climatic zones, an assessment based on the WHYMAP Global Mega Aquifer Systems was carried out. In structurally stable or naturally climate-driven porous aquifer systems (e.g., AQU 11 and AQU 12), the downscaled outcomes exhibited high accuracy with negligible absolute errors. For example, in AQU 11, the selected XGBoost architecture attained a Root Mean Square Error (*RMSE*) of 1.15 mm. Nevertheless, in regions exposed to intensive agricultural irrigation and over-exploitation of fossil groundwater (e.g., AQU 22 and AQU 23), the necessity of our architectural selection became

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

evident. The adopted tree-based framework (XGBoost) achieved highly compact error convergence (e.g., R^2 = 0.90, *RMSE* = 2.56 mm in the Indus Basin), demonstrating significantly superior performance compared to the continuous DNN baseline (R^2 = 0.85, *RMSE* = 3.23 mm). This quality control procedure validates that the tree-based ensemble has an inherent structural advantage in capturing the sharp boundaries caused by localized human pumping, thus justifying its function as the core engine for our dataset production.

Table 3. Cross-validation performance by aquifer system and mask.

Aquifer ID	Mega-Aquifer System / Specific Mask	XGBoost		DNN	
		RMSE	R^2	*RMSE*	R^2
11	Cuvelai-Upper Zambezi Basin	1.15	0.921	1.61	0.845
12	Stampriet-Kalahari Basin	1.72	0.929	2.34	0.868
13	Karoo Basin	2.37	0.94	2.93	0.908
14	Northern Great Plains Aquifer	2.8	0.902	3.25	0.867
15	Cambrian-Ordovician Aquifer	3.88	0.942	4.34	0.928
22	Arabian Aquifer System	1.95	0.9	2.42	0.846
23	Indus Basin	2.56	0.903	3.23	0.846
24	Ganges-Brahmaputra Basin	3.17	0.916	3.96	0.87
25	West Siberian Basin	4.42	0.926	5.91	0.867
33	Russian Platform Basins	4.12	0.916	4.9	0.881
34	North Caucasus Basin	5.4	0.943	5.97	0.93

3.1.3 Cross-scale Consistency and Independent In-situ Validation

To precisely evaluate the mass conservation and spatial fidelity of the resolution enhancement process, the 1 km downscaled GWSA results were spatially aggregated to a 0.25° resolution and cross-validated against the original 2002–2020 mean observations (Fig. 3). The results demonstrate high consistency across both modeling architectures ($R^2 > 0.8$), effectively reproducing the long-term spatial patterns of macroscopic groundwater storage anomalies.

At the global scale, the Extreme Gradient Boosting (XGBoost) model outperformed the Deep Neural Network (DNN) model. As illustrated in the global panels of Fig. 3, XGBoost achieved an R^2 of 0.972, a correlation coefficient (*CC*) of 0.986, and a root-mean-square error (*RMSE*) of 2.097 cm. In comparison, the DNN model achieved an R^2 of 0.961, a *CC* of 0.981, and an *RMSE* of 2.661 cm. Across the 11 evaluated aquifer categories, XGBoost exhibited higher R^2 and *CC* values, alongside lower *RMSE*, in 10 categories, indicating a more accurate characterization of spatial distributions with reduced systemic bias. Further spatial breakdown reveals that major groundwater basins (AQUs 11–15) and regions with complex hydrogeological structures (AQUs 22–25) showed superior fitting accuracy,

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

maintaining an average R^2 of 0.961 and an *RMSE* of 0.730 cm. Although localized shallow aquifers (AQUs 33–34) exhibited slightly greater uncertainty with an average *RMSE* increasing to 2.091 cm, this variance is physically plausible. Such regions are heavily influenced by localized micro-topography, intensive land-use changes, and complex near-surface recharge processes, which inherently introduce high-frequency spatial variations. Consequently, the highly robust XGBoost architecture was selected to generate the final dataset for subsequent in-depth analysis.

Beyond internal cross-scale consistency, rigorous independent validation was conducted utilizing long-term physical evolutionary trends (2006–2020) from 1,518 global groundwater monitoring stations (Fig. 4). To systematically resolve the dimensional discrepancy between satellite-derived Equivalent Water Height (EWH) and in-situ Groundwater Level (GWL), a geographically zoned Specific Yield (*Sy*) matrix was introduced. Furthermore, quantile binning was applied to the grid trends to mitigate high-frequency local noise caused by heterogeneous geological conditions or isolated pumping events, thereby extracting representative macroscopic evolution patterns.

As shown in Fig. 4, after eliminating high-frequency spatial noise, both downscaling architectures demonstrated robust linear relationships with the ground-measured trends ($p < 0.01$). However, they exhibited physically significant differences in overall fitting accuracy and error magnitude. The XGBoost model (Fig. 4a) demonstrated a superior capacity to explain the variance of observed trends, yielding an R^2 of 0.44 and a lower *RMSE* of 0.51 cm yr^{-1}. Notably, its scattered data points were densely and symmetrically distributed around the 1:1 reference line. In the context of dynamic range analysis, XGBoost effectively captured deep groundwater depletion signals driven by intensive anthropogenic activities (< -1.0 cm yr^{-1}) and successfully reproduced reservoir recovery characteristics in the positive ranges. In contrast, although the DNN model (Fig. 4b) maintained a comparable linear correlation (*CC* = 0.67), it yielded a lower explained variance (R^2 = 0.39) and a higher prediction error (*RMSE* = 0.57 cm yr^{-1}). Moreover, the DNN predictions exhibited greater deviations from the reference line at extreme depletion and recharge values.

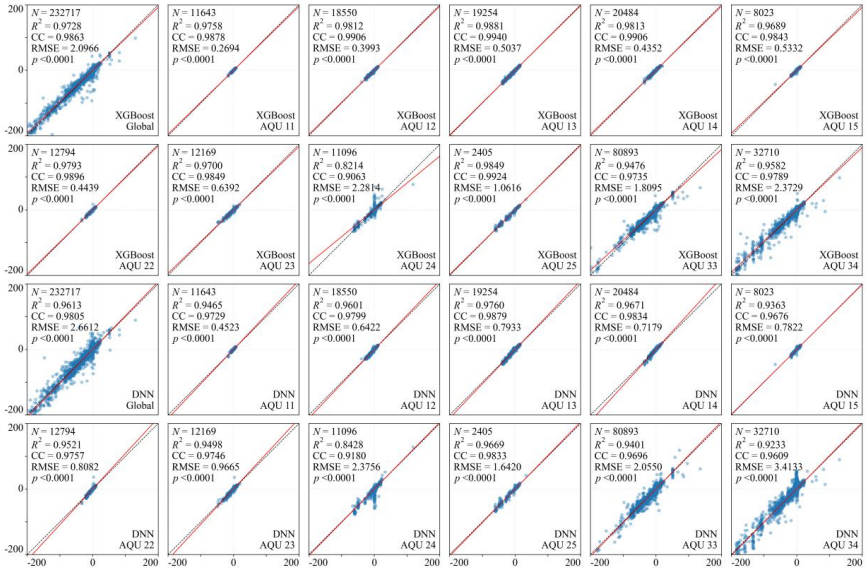
When interpreting these validation metrics, it is imperative to acknowledge the inherent point-to-pixel scale mismatch. In-situ wells record micro-scale groundwater fluctuations influenced by highly localized fractured media or individual pumping events, whereas the downscaled GWSA

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

Open Access Earth System Science Data Discussions

555 represents spatially averaged storage dynamics over a 1 km grid cell. Given these significant
556 dimensional and physical scaling challenges, an R^2 of 0.44 confirms the strong efficacy of the
557 tree-based architecture in reconstructing core macroscopic evolutionary trends, rather than attempting
558 precise point-scale replication.

559



560
561 **Figure 3. Cross-validation of the downscaled groundwater storage anomalies (GWSA) and the**
562 **original-resolution GWSA multi-year mean for 2002–2020.**

563 The first and second rows show the XGBoost results, and the third and fourth rows show the DNN
564 results. Panels are arranged in the following order: global, Major groundwater basins (AQUs 11–15),
565 Complex hydrogeological structures (AQUs 22–25), and Local and shallow aquifers (AQUs 33–34). In
566 each panel, the x axis represents the original GWSA, and the y axis represents the downscaled GWSA
567 aggregated to the same resolution. The red line indicates the fitted line. The corresponding sample size
568 (*N*), coefficient of determination (R^2), correlation coefficient (*CC*), root-mean-square error (*RMSE*),
569 and significance level (*p*) are shown in the upper-left corner.

570

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

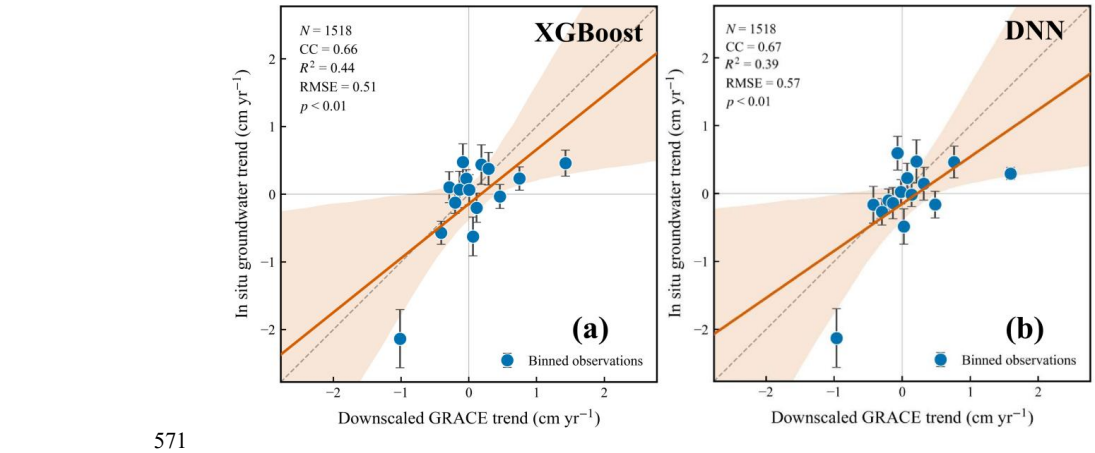


Figure 4. Cross-validation of downscaled groundwater storage anomalies (GWSA) and long-term evolutionary trends of ground-measured groundwater levels (2006–2020).

Figures (a) and (b) respectively present the independent validation results for the XGBoost and DNN models. The x-axis represents the long-term trends of GWSA predicted by the downscaling models, while the y-axis displays the trends of measured groundwater equivalent water height (EWH) converted through geographic Specific Yield (S_y) matrices, with the unit being cm yr^{-1}. To mitigate the heterogeneity resulting from individual well observations and high-frequency noise, 1,518 validation grids with continuous observational records were globally aggregated into 15 bins (Binned observations, blue dots). The blue dots indicate the arithmetic means of grid trends within each bin, and the vertical error bars represent the standard errors of the mean (SEM) for the measured EWH trends. The gray dashed line serves as the 1:1 reference line, and the dark orange solid line and its light translucent shaded area respectively represent the best linear fit and the 95% confidence interval. Comprehensive statistical evaluations reveal that, despite having comparable linear correlations, XGBoost (CC = 0.66, R^2 = 0.44, $RMSE$ = 0.51 cm yr^{-1}, $p < 0.01$) significantly outperforms DNN (CC = 0.67, R^2 = 0.39, $RMSE$ = 0.57 cm yr^{-1}, $p < 0.01$) in minimizing prediction errors and explaining trend variances. This validates its superior robustness and accuracy in reproducing the true physical evolutionary dynamics of groundwater storage at the global macro-scale.

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

3.2 Spatial Fidelity and Resolution Enhancement of Downscaled GWSA

To systematically evaluate the capability of the downscaling framework in characterizing localized hydrological features, a detailed comparative analysis was conducted across four globally representative regions (Fig. 5). Original GRACE observations, restricted by the truncation of spherical harmonic coefficients, exhibit significant spatial smoothing and signal leakage effects. This manifests as coarse macroscopic anomalies (Fig. 5, Column a) that obscure fine-scale groundwater dynamics. In contrast, the 1 km downscaled products effectively mitigate these observational artifacts while demonstrating high spatial fidelity.

In the Amazon Basin (Area 1), driven by natural hydro-climatic forcing, the downscaling framework successfully reconstructs complex dendritic drainage textures that are entirely absent in the coarse-resolution GRACE data. Furthermore, in the mountainous region of southern Tibet (Area 2), characterized by extreme topographical complexity, the finalized dataset (XGBoost, Fig. 5, panel 2c) captures fine-scale geomorphological signals, resolving spatial textures that align closely with the micro-topography of mountain gorges. To demonstrate the necessity of our architectural selection, the DNN baseline (Fig. 5, panel 2b) is shown to exhibit a tendency to fit dominant forcing signals, producing more generalized regional patches that fail to delineate the sharp gradients of the deep valley systems. This disparity confirms that the selected tree-based ensemble possesses superior capacity in processing high-frequency spatial information derived from static geophysical covariates such as digital elevation models.

The advantages of this high-resolution reconstruction are particularly pronounced in regions dominated by intensive anthropogenic intervention. In the agricultural zones surrounding New Delhi, India (Area 3) and the Zhangye agricultural oasis within the hyper-arid Heihe River Basin (Area 4), the finalized dataset accurately resolves localized depletion hotspots. It clearly delineates fine-scale groundwater depletion patterns (Fig. 5, panels 3c and 4c) that correspond highly with actual agricultural boundaries and exploitation extents.

The contrast with the continuous baseline architecture highlights the robustness of the finalized product in these heavily pumped areas. The DNN baseline tends to produce spatially smoothed anomaly zones, such as broad depletion cones (Fig. 5, panel 3b), which obscure abrupt transitions

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

between irrigated oases and the surrounding arid background. In contrast, the adopted XGBoost architecture exhibits an inherent edge-preserving property, successfully capturing the sharp, nonlinear boundaries induced by localized groundwater pumping. These findings demonstrate that the integration of multi-source 1 km covariates is not merely a mathematical downscaling procedure but a necessary reconstruction of the underlying hydrological heterogeneity, providing a robust digital benchmark for localized water resource management.

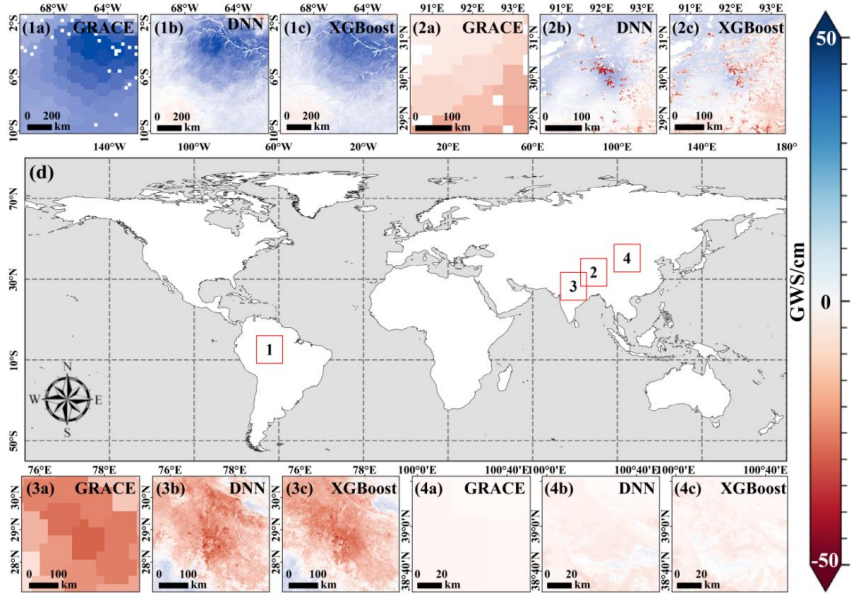


Figure 5. Spatial enhancement and feature reconstruction of 1 km resolution groundwater storage anomalies (GWSA) across four globally representative regions.

Columns present (a) the original 0.25° GRACE observations, alongside downscaled outputs generated by (b) Deep Neural Network (DNN) and (c) Extreme Gradient Boosting (XGBoost) models at the 1 km scale. The four selected regions (indicated by red boxes) encompass: (1) the Amazon Basin, driven by natural hydro-climatic forcing; (2) the mountainous region of southern Tibet, characterized by extreme topographical complexity; (3) the agricultural zones surrounding New Delhi, India, driven by intensive anthropogenic groundwater extraction; and (4) the Zhangye agricultural oasis within the hyper-arid Heihe River Basin. A comparative assessment highlights that original GRACE observations (Column a)

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

display coarse macroscopic anomalies due to spatial leakage. Following downscaling, the XGBoost model (Column c) captures fine-scale geomorphological and anthropogenic signals, resolving spatial textures aligned with canyon micro-topography in southern Tibet (2c) and clearly delineating localized groundwater depletion patterns corresponding to agricultural boundaries in New Delhi (3c) and Zhangye (4c). Conversely, the DNN model (Column b) exhibits a tendency to fit dominant forcing signals, resulting in more generalized, spatially smoothed anomaly zones (e.g., broad depletion cones) that obscure fine-scale hydrological details in both topographically complex (2b) and heavily pumped regions (3b).

3.3 Global Climatology and Spatiotemporal Variability Patterns

3.3.1 Baseline Imbalances and Meridional Profiles

The spatial pattern of the global mean groundwater storage anomalies (Mean GWSA) was reconstructed at a 1 km resolution for the period from 2002 to 2020, along with the corresponding meridional mean profiles (Fig. 6). This reconstruction clarifies the baseline state of mass balance in the global groundwater system during the study period and offers deeper insights into the current imbalanced distribution patterns of the terrestrial water cycle on Earth. From the perspective of the global spatial pattern (Fig. 6b), the terrestrial water storage exhibits systematic mass imbalances, manifested primarily as a net surplus in tropical and mid-to-high-latitude humid zones, whereas cumulative depletion occurs in areas of intensive human intervention and arid regions.

Significant negative anomalies are not randomly distributed but are highly concentrated in several typical large alluvial plain aquifers and arid regions globally. These localized depletion centers include the Indus-Ganges Basin, the North China Plain, the Arabian Peninsula, the Caspian Sea region, and the High Plains of the United States. In sharp contrast, significant positive anomalies, functioning as global groundwater accumulation zones, are stably distributed in tropical rainforest areas influenced by the Intertropical Convergence Zone (ITCZ) and certain mid-to-high-latitude regions affected by persistent moist air currents.

The meridional mean profile (Fig. 6a) further condenses the two-dimensional heterogeneity into distinct zonal distribution characteristics along the latitudinal gradient. A core feature of this profile is

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

the global groundwater depletion zone situated between 25°N and 45°N, with its trough spatially corresponding to the most densely distributed anthropogenic depletion hotspots on the map. From the perspective of the dynamic mechanism, this phenomenon results from the high spatial coupling between climatic drought stress under natural geographic conditions and human activities. This latitude belt not only lies within the long-term recharge-deficient region caused by the descending branch of the Hadley Cell but also represents the most densely populated and agriculturally irrigated area globally.

The reduction in the depth of the mean profile indicates that the groundwater system in this zone has fundamentally deviated from its climate-driven natural equilibrium state, leading to persistent mass loss. The profile exhibits stable positive peaks near the equator (Eq), consistent with the high concentration of reserves in tropical rainforest regions, while a pronounced negative shift occurs at extremely high latitudes (around 80°N). This shift effectively captures land mass loss signals driven by cryosphere retreat or related dynamic processes at the polar margins. Ultimately, this spatial distribution of multi-year averages serves as a physical framework for understanding the evolution of the global groundwater system, providing essential contextual support for the subsequent analysis of dynamic evolution rates.

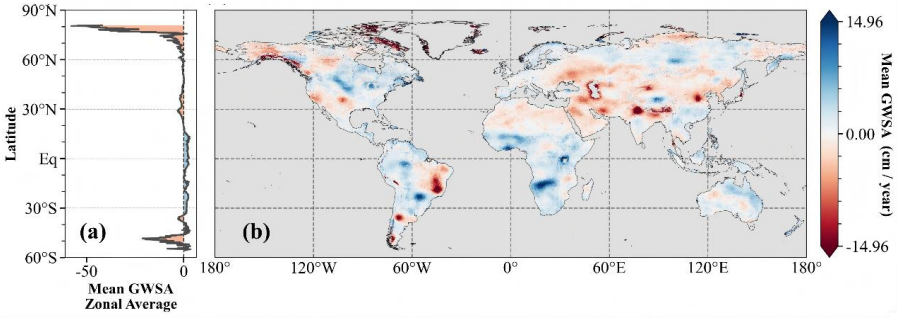


Figure 6. Spatial pattern and latitudinal profile of global mean groundwater storage anomalies (Mean GWSA) during the period from 2002 to 2020.

(a) Latitudinal mean profile, presenting area-weighted averages of inland land pixels across diverse latitudes. The black solid line denotes the latitudinal mean, and the red and blue filled areas signify groundwater depletion and surplus states, respectively. (b) Global 1 km spatial pattern map depicting the high-resolution mean groundwater storage anomaly distribution reconstructed via downscaling

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

techniques. The color bands exhibit a smooth transition from dark red (depletion) to dark blue (surplus), and the white areas represent regions in a near-equilibrium state.

3.3.2 Seasonal Contributions and Global Hydrological Rhythms

To further evaluate the influence of cyclic factors on the dynamics of global groundwater storage, this study quantified the relative contributions of seasonal and non-seasonal components using the downscaled monthly GWSA time series (Fig. 7). In this framework, seasonality is characterized by the average intra-annual standard deviation for each year from 2002 to 2020, whereas non-seasonality is measured by the standard deviation of the annual mean values over the same 19-year period. The relative seasonal contribution is determined by the ratio of the seasonal component to the sum of both components, with a threshold of 50% used to identify regions predominantly governed by seasonal factors.

The results indicate that global GWSA dynamics generally exhibit distinct seasonally-regulated characteristics. Based on equal-area projection statistics, the seasonal contribution exceeds 50% across 52.30% of the global land area. This suggests that over half of the terrestrial domains are predominantly governed by stable and recurrent intra-annual cycles, rather than non-seasonal anomalies at interannual scales. Furthermore, in 26.09% of the land area, the seasonal contribution is greater than 66%, meaning the seasonal component is more than 1.5 times the magnitude of non-seasonal fluctuations, indicating a pronounced dominance of seasonality.

Spatially, regions with high seasonal contributions (60% to 90%) are primarily concentrated in areas with regular hydrological rhythms, including the Amazon and Congo Basins, the South and East Asian monsoon regions, northern Europe to the southern margin of Siberia, and high-latitude zones in North America. These patterns reflect a robust response of groundwater storage to periodic variations in precipitation, evapotranspiration, surface water-groundwater interactions, and freeze-thaw processes.

In contrast, regions with low seasonal contributions are predominantly distributed in inland arid zones, such as North Africa, the Arabian Peninsula, Central Asia, inland Australia, and western North America. In these areas, non-seasonal fluctuations dominate, manifesting primarily as persistent GWSA deficits driven by long-term trends. The frequent occurrence of extreme heat, compound droughts, and heatwave events in these regions exacerbates long-term soil moisture imbalances. These

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

groundwater systems are heavily driven by persistent drought, intensive overexploitation, and complex topography-climate coupling. Consequently, they have transitioned from structured, seasonal recharge-depletion cycles to a state dominated by non-periodic disturbances.

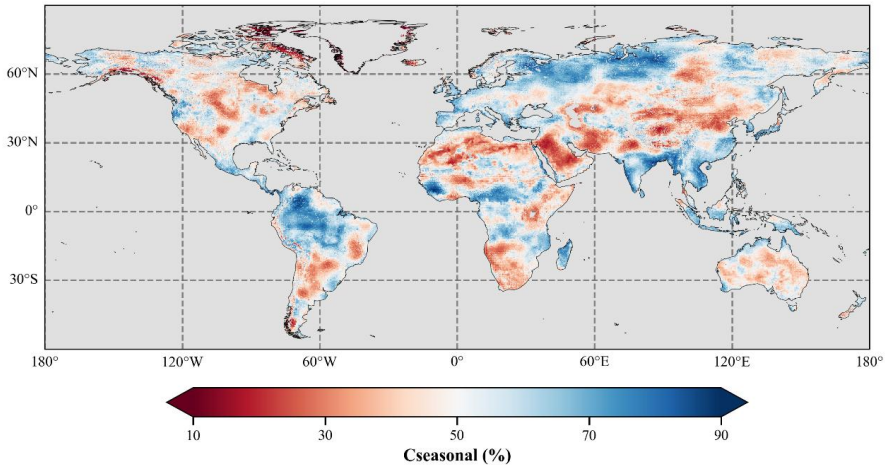


Figure 7. Seasonal contribution to global 1 km groundwater storage anomalies (GWSA).

Color bar shows the contribution of seasonality to GWSA variability during 2002–2020 (%). Red denotes a higher seasonal contribution, whereas blue denotes a lower seasonal contribution. Regions with a seasonal contribution greater than 50% were defined as seasonality-dominated areas.

3.4 Spatiotemporal Evolution and Trend Polarization

3.4.1 Spatial Heterogeneity of Global GWSA Trends

As illustrated in Fig. 8, the evolution of global groundwater storage over the past two decades has exhibited notable regional clustering characteristics and spatial differentiation trends. With a spatial resolution of 1 km, the downscaled trend maps mitigate the spatial smoothing effects inherent in traditional coarse-resolution GRACE observations, thereby precisely revealing complex and diverse local hydrological evolution gradients.

Globally, significant groundwater depletion signals ($p < 0.01$ and trend rate < -0.10 cm/month) are mainly manifested in two typical spatial patterns.

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

The first pattern consists of relatively clustered, patchy, and linear depletion centers, the spatial distribution of which shows a high degree of overlap with global high-intensity agricultural irrigation basins. For example, the 1 km dataset effectively captures steep depletion gradients along specific pre-mountain alluvial fans or typical farmland boundaries in regions such as the Central Valley of California, the High Plains Aquifer (Ogallala), northwestern India, and the North China Plain of China.

The second pattern involves large-scale, contiguous regional depletion, primarily concentrated in arid and semi-arid climate zones and major fossil groundwater extraction areas, such as the Middle East, the Arabian Peninsula, the hinterland around the Caspian Sea, and the La Plata Plain at the border of Argentina and Brazil in South America. Most of the Middle East region is characterized by a tropical desert climate, while the hinterland around the Caspian Sea is located deep within the continent, featuring arid conditions and intense evaporation, with inherently scarce water resources. The dense distribution of significant depletion points in these areas (indicated by black stippling in Fig. 8) indicates that their groundwater systems remained in a state of sustained and significant depletion during the study period.

Spatially corresponding to the aforementioned deficit patterns, significant groundwater accumulation regions worldwide ($p < 0.01$ with trend rates > 0.10 cm/month) predominantly display diffuse spatial distribution characteristics in accordance with major surface water systems and natural topography. In equatorial and tropical monsoon climate zones, including the transverse extension areas of Africa's Sahel region, the Victoria Lake Basin, and the core Amazon Basin, the Groundwater Storage Anomalies (GWSA) demonstrate extensive and spatially continuous significant surplus. Additionally, high-latitude regions in North America, such as northern Canada with a polar long-winter climate, and localized areas in northern Eurasia, such as the southeastern Siberian Plain in Russia where permafrost is prevalent, show groundwater accumulation signals in trend maps. These signals potentially reflect deep water infiltration driven by accelerated hydrological cycles or cryosphere degradation at high latitudes. These surplus zones, dominated by large-scale climatic or natural hydrological processes, exhibit relatively smooth spatial transitions at 1 km scales, forming distinct morphological contrasts with the sharply defined deficit boundaries caused by localized high-intensity human activities.

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

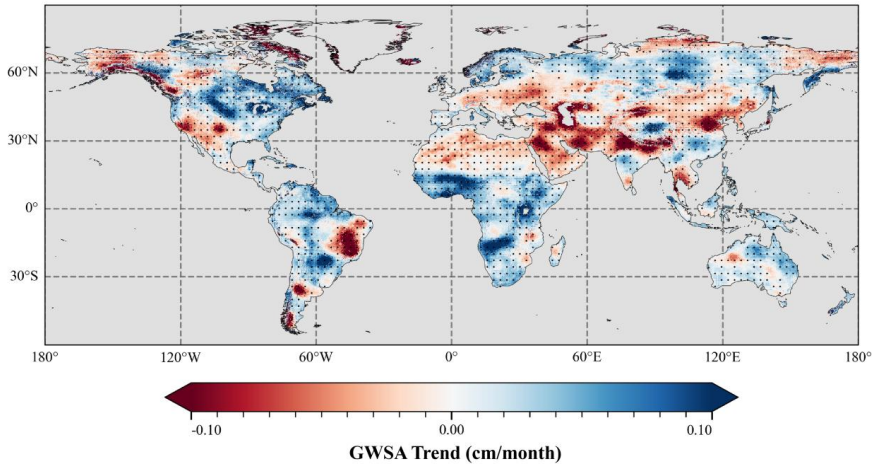


Figure 8. Long-term evolutionary trends and significant spatial distributions of global 1 km resolution groundwater storage anomalies (GWSA).

The color scale of the base map denotes the long-term change rate of GWSA (unit: cm/month) calculated on a pixel-by-pixel basis using the Theil-Sen median method. The red color series indicates continuous groundwater depletion, whereas the blue color series represents cumulative groundwater accumulation. Areas covered by dense black stippling points signify that the evolutionary trend of the pixel has passed the non-parametric Mann-Kendall significance test ($p < 0.01$). Light gray regions correspond to oceans, polar permafrost areas, or extremely arid deserts where valid data are lacking.

3.4.2 Statistical Classification and Graded Evaluation of Groundwater Evolution

To systematically quantify the statistical robustness and physical significance of global groundwater evolution trends, this study integrated the absolute magnitude of Theil-Sen trend rates with non-parametric Mann-Kendall (MK) significance test thresholds to establish a spatially stratified evaluation system composed of seven evolutionary stages (Fig. 9). This dual-indicator classification framework not only identifies the rate of reservoir variations but also determines whether groundwater systems exhibit statistically significant responses to external forcing factors, such as climate oscillations and high-intensity extraction, through significance thresholds.

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

778 Pixel-level statistical results indicate that groundwater reserves in most global regions have
779 significantly deviated from long-term dynamic equilibrium states from 2002 to 2020. Data reveal that
780 approximately 38.84% of the effective land area is experiencing significant or extremely significant
781 groundwater depletion ($p < 0.05$ or $p < 0.01$), whereas about 49.16% of regions show significant or
782 extremely significant recharge. In contrast, only 12.00% of land areas maintained a relatively stable
783 dynamic equilibrium with non-significant trends during the study period.

784 Based on the quantitative description provided by the MK test, the study further reveals a
785 representative spatial polarization pattern in global groundwater evolution (Fig. 9). Analysis shows that
786 in all regions displaying significant change signals, the evolutionary gradients are extremely steep,
787 lacking smooth transitional zones across space. Specifically, among the total deficit areas accounting
788 for 38.84%, the proportion of areas showing extremely significant reduction ($p < 0.01$) reaches 37.13%,
789 while only 1.70% exhibit significant reduction ($p < 0.05$). Accumulation areas present similar
790 characteristics, with the proportion of extremely significantly increased areas (47.01%) far exceeding
791 that of significantly increased areas (2.15%).

792 This distribution pattern, characterized by the absence of a broad buffer zone between
793 non-significant and extremely significant changes, is manifested as clearly defined spatial physical
794 boundaries on the 1 km high-resolution classification base map. From the perspective of physical
795 mechanisms, this phenomenon implies that the global groundwater system responds to intense external
796 pressures, such as structural overextraction or extreme precipitation deficits, with abrupt, unidirectional
797 physical evolution rather than gradual processes.

798

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

Open Access Earth System Science Data Discussions

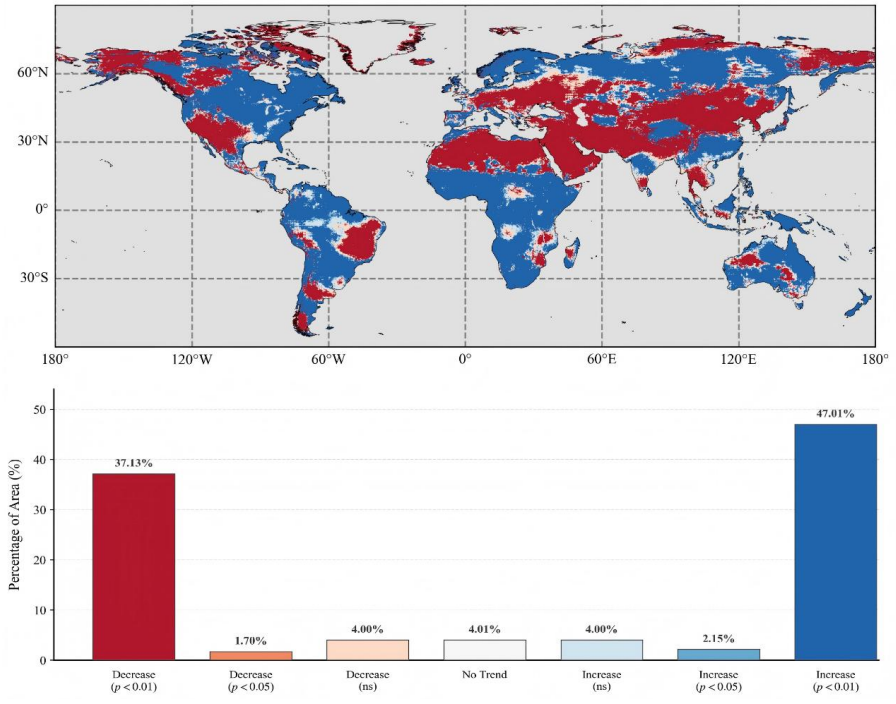


Figure 9. Spatial Classification and Statistical Analysis of the Evolutionary Status and Area Proportion of Global 1 km Groundwater Reserve Anomalies (GWSA). The upper figure depicts the global spatial distribution of groundwater evolutionary status (from 2002 to 2020, which is constructed by integrating Theil-Sen trend rates and Mann-Kendall significance tests. Different evolutionary stages are represented by RdBu divergence color bands. Dark red and dark blue respectively signify highly significant reduction and highly significant increase (p < 0.01), while lighter colors correspond to dynamically balanced states with insignificant changes (ns) or no trend. The lower figure presents a bar chart that shows the global proportion of effective land area, maintaining a strict color correspondence with the spatial distribution map. The statistical results disclose a prominent "spatial polarization" in global groundwater evolution: regions with highly significant reduction (37.13%) and highly significant increase (47.01%) account for the majority of the global land area, whereas intermediate transitional zones with significant changes (p < 0.05) constitute a minimal proportion (less than 2.5%). This distribution pattern implies that the global groundwater system demonstrates abrupt,

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

unidirectional physical evolutionary responses under intense external pressures rather than gradual processes.

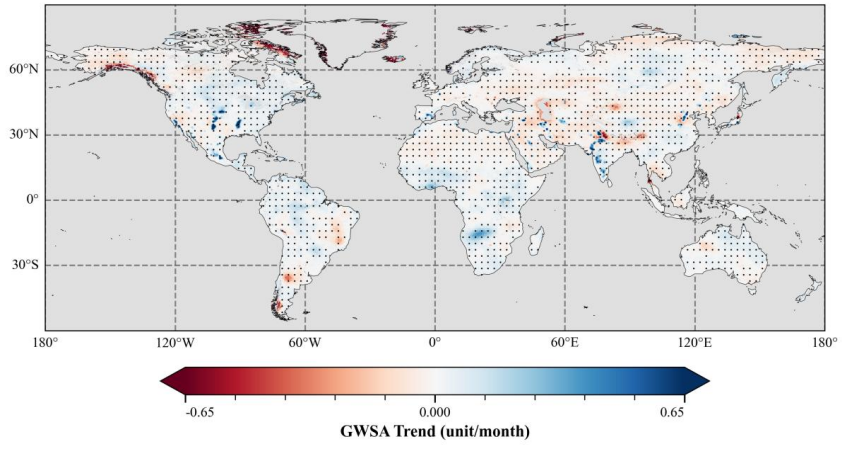
3.5 Anthropogenic Fingerprint Identification via Residual Attribution

In the long-term evolutionary panorama of global 1 km resolution groundwater storage anomalies (GWSA), regional depletion and accumulation hotspots not only manifest significant spatial polarization but also exhibit distinct spatial dynamics at the morphological level. To effectively isolate the influence of human intervention from composite signals, this study establishes a diagnostic framework based on residual attribution. By comparing high-resolution reconstructed sequences (obtained by filling missing values via Singular Spectrum Analysis) with simulation benchmarks from global hydrological models (GHMs) founded on physical mechanisms, we quantitatively determined the estimated magnitude of intensive anthropogenic extraction. Specifically, the 0.5° natural background signals simulated by the GHMs were bilinearly interpolated to a 1 km resolution to achieve spatial matching with the downscaled GWSA before subtraction. Maintaining the 1 km resolution provides precise spatial morphological criteria for distinguishing anthropogenic anomalies from natural gradients.

Among the global deficit hotspots identified in the previous sections, the residual sequences following STL decomposition for the elimination of periodic components exhibit remarkable spatial heterogeneity (Fig. 10). In high-intensity irrigation regions, such as the Indus Basin (AQU 23) and the North China Plain, the residual trend maps feature distinct, steep-gradient deep red patches. In the North China Plain, long-term intensive agricultural practices have led to severe and continuous groundwater deficits, such as those persisting from 2019 to 2020. Significantly, these abrupt changes in spatial morphology demonstrate high physical consistency with actual administrative irrigation boundaries or farmland exploitation extents. Conversely, regions mainly influenced by natural factors, such as the core Amazon Basin, display residual values approaching zero after subtracting the model background, suggesting a stable distribution. This morphological contrast verifies that the significant deficits observed at the 1 km scale ($p < 0.01$) are predominantly driven by human exploitation surpassing natural recharge thresholds.

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

840 The residual attribution framework further reveals the structural limitations of physical
841 hydrological models in simulating complex human interventions. In the Arab Aquifer (AQU 22), where
842 the over-extraction of fossil groundwater is severe, traditional GHMs typically generate simulation
843 trajectories characterized by gradual natural attenuation, failing to accurately replicate the observed
844 sharp nonlinear depletion processes. Using the reconstructed time-series data with missing values filled
845 through SSA and temporal coherence preserved, our study identifies significant negative deviations
846 between observed trends and model-simulated trajectories. The continuous accumulation of such
847 time-series residuals offers a direct quantification of groundwater imbalance driven by human pumping
848 activities, effectively overcoming the limitations of physical models in characterizing anthropogenic
849 extraction processes. By maintaining the 1 km resolution throughout the attribution process, this
850 residual distribution map precisely identifies groundwater depletion hotspots driven by localized
851 high-intensity irrigation activities, effectively eliminating the spatial smoothing effects from the
852 original GRACE observations.



853
854 **Figure 10. Human Pumping Fingerprint Identification Based on Observation and Model Residual**
855 **Attribution ($p < 0.01$).**

856 This visualization presents the distribution characteristics of physical signals subsequent to the
857 attribution of groundwater trends. The computational approach entails bilinearly interpolating the
858 natural background signals simulated by global hydrological models (GHMs) to a 1 km resolution, and
859 subsequently subtracting these natural baselines from the generated 1 km downscaled groundwater

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

storage anomaly (GWSA) reconstruction sequence. Trend computations utilize the Theil-Sen median method, followed by seasonal time series decomposition (STL) for de-seasonalization. The red regions signify observed depletion rates that significantly surpass the explanations provided by natural climate variability, representing an approximation of localized depletion anomalies likely driven by intensive anthropogenic extraction, following the removal of modeled natural fluctuations. The blue areas may indicate artificial recharge or model overestimation of natural deficits. Dense black dots denote residual trends that pass the Mann-Kendall significance test ($p < 0.01$). By maintaining the 1 km resolution throughout the attribution process, this residual distribution map precisely identifies groundwater depletion hotspots driven by localized high-intensity irrigation activities, effectively eliminating the spatial smoothing effects from the original Gravity Recovery and Climate Experiment (GRACE) observations.

4. Discussion

4.1 Comparison with Existing Macro-scale Groundwater Datasets

To objectively evaluate the scientific positioning of the 1 km global groundwater storage anomaly (GWSA) dataset, it is essential to systematically compare it with existing mainstream monitoring products. Native GRACE satellite observations are fundamentally constrained by their coarse spatial resolution, which frequently precludes the capture of fine-scale groundwater dynamics and restricts their utility in sub-basin scale management. In contrast, the 1 km dataset presented in this study functions as a high-fidelity bridge across this cross-scale observation gap, providing a refined spatial benchmark that is difficult to achieve through macro-scale gravity inversion alone.

Compared with traditional Global Hydrological Models (GHMs), such as GLDAS, PCR-GLOBWB, and WaterGAP, this dataset offers significant improvements in characterizing localized anthropogenic interventions. Although the latest generation of GHMs exhibits high physical rigor in simulating natural climatic cycles under natural forcings (Scanlon et al., 2018; Müller Schmied et al., 2021), these models often exhibit limitations in simulating the highly nonlinear trajectories of water resource depletion in intensive agricultural regions. This discrepancy primarily stems from their reliance on coarse spatial grids and the widespread absence of localized groundwater extraction

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

887 modules. As demonstrated in Section 3.5, while traditional GHMs predominantly reflect gradual
888 natural attenuation backgrounds in the Indus Basin and the North China Plain, our 1 km downscaled
889 dataset successfully captures sharp depletion troughs that exceed natural resilience thresholds.
890 Consequently, in the absence of large-scale pumping records, this dataset functions as a robust
891 observational reconstruction and provides an objective benchmark for quantifying the localized impacts
892 of unsustainable anthropogenic extraction.

893 Furthermore, our framework provides a distinct methodological alternative to previous statistical
894 downscaling efforts. Most existing studies remain confined to specific regional basins (Miro and
895 Famiglietti, 2018) or exhibit a spatial resolution plateau at the 10-km scale (Yin et al., 2022). The
896 current 1 km global reconstruction is achieved not through simple mathematical interpolation, but by
897 leveraging the Extreme Gradient Boosting (XGBoost) architecture to extract localized spatial
898 heterogeneity from high-resolution auxiliary environmental covariates. This approach accurately
899 reconstructs natural hydro-geomorphological patterns, such as canyon-aligned textures in southern
900 Tibet, while simultaneously isolating agricultural water-use hotspots in hyper-arid regions like
901 Zhangye. Such fine-scale resolution is critical for resolving the spatial polarization of groundwater
902 evolution identified in this study.

903 Finally, this dataset serves as a valuable macro-scale alternative for regions where
904 computationally intensive three-dimensional numerical models, such as MODFLOW, are unfeasible.
905 The development of high-fidelity regional numerical models relies on dense in-situ borehole data and
906 precise aquifer geometries, presenting substantial challenges in data-scarce regions across Africa and
907 Central Asia. Although our data-driven paradigm does not explicitly model vertical flow fields, it
908 exhibits robust statistical consistency with storage trends validated against a comprehensive global
909 network of 1,518 monitoring wells (Jasechko et al., 2024). Therefore, this 1 km GWSA product can
910 function as a crucial prior hydrological proxy for ungauged basins, providing essential initial
911 conditions and validation fields for the future development of regional groundwater digital twins.

912 **4.2 Mechanistic Differences Between Deep Learning and Tree-Based Models in Downscaling**

913 Identifying the distinct physical representation disparities among underlying algorithms under
914 extreme hydrological forcing conditions is critical for understanding the applicability and limitations of

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

915 the 1 km dataset presented in this study. Although the cross-validation results (Section 3.1) indicate
916 that both DNN and XGBoost achieve high prediction accuracies globally, their mathematical
917 architectures reveal systematic physical representation discrepancies when processing highly nonlinear
918 hydrological forcing and complex spatial heterogeneity.

919 In terms of spatial texture representation, as illustrated in Section 3.2 (Fig. 5), XGBoost exhibits
920 relatively continuous transitions at aquifer boundaries while preserving rich spatial micro-heterogeneity
921 textures at the pixel scale. This spatial representation characteristic stems from its underlying random
922 subspace mechanism and ensemble averaging effect (Breiman, 2001; Tyralis et al., 2019). During
923 training, XGBoost enforces node splitting of each decision tree solely within randomly selected feature
924 subsets. This design effectively reduces the dominance of macro-scale forcing factors (such as
925 precipitation), enabling the model to capture local heterogeneity from secondary environmental
926 variables like digital elevation models (DEM) and soil physical properties. Simultaneously, the
927 independent prediction expectations generated by hundreds of decision trees partially mitigate the hard
928 boundary segmentation effects of individual trees, allowing the model to represent spatial textures
929 shaped by micro-terrain and local capillary hydrodynamics consistent with natural gradient patterns
930 (e.g., Fig. 5-2c).

931 In contrast, the spatial patterns generated by DNNs display sharply defined boundary features with
932 relatively limited local details (e.g., Fig. 5-2b and Fig. 5-3b). This phenomenon is attributable to the
933 priority capture characteristics of gradient-descent-based multi-layer perceptrons for macroscopic
934 dominant signals. During high-dimensional nonlinear mapping, DNN hierarchical networks tend to
935 prioritize fitting climate forcing signals with the highest variance contribution rates. Consequently,
936 localized weak topographic or soil texture variations may be processed as high-frequency noise by the
937 networks, leading to implicit filtering or excessive spatial smoothing (Reichstein et al., 2019; Shen,
938 2018).

939 Additionally, when localized areas experience extreme human-induced pumping activities (e.g.,
940 agricultural zones in New Delhi, Fig. 5-3) causing sudden water storage fluctuations, nonlinear
941 activation functions (e.g., ReLU) in DNN hidden layers may induce directional high-frequency
942 activation. This can result in numerical jumps in predicted values at critical hydrological boundaries,

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

943 forming large-scale deficit patches with sharp visual polarization on maps. Consequently, DNNs tend
944 to emphasize macroscopic trends, whereas XGBoost demonstrates greater sensitivity to heterogeneity
945 at the micro-scale.

946 Beyond spatial texture representation, the temporal responses of groundwater systems to extreme
947 climate events highlight disparities in the extrapolation capabilities between the two models. The
948 mathematical nature of tree-based models inherently confines their prediction outputs within the
949 bounding envelope of the training set label data (Hastie et al., 2009). This implies that during extreme
950 drought events, XGBoost prediction curves may exhibit truncation at the lower end, potentially
951 underestimating the absolute physical magnitude of extreme groundwater deficits. In contrast, DNNs,
952 with continuous mapping capabilities, demonstrate extrapolation potential beyond the training set
953 distribution boundaries, enabling the simulation of deeper trough amplitudes. However, without
954 explicit constraints from hydrological physical laws, such mathematical extrapolation risks
955 non-physical oscillations.

956 A comprehensive understanding of these complementary mechanisms not only explains the
957 rationale behind our dataset selection but also provides critical insights for deploying hybrid integration
958 frameworks with enhanced physical constraints in future Earth system science (Karniadakis et al.,
959 2021). The XGBoost-based high-resolution dataset presented in this study demonstrates an enhanced
960 capability to resolve complex terrain features and localized human pumping boundaries, making it
961 particularly suitable for water resource management and attribution research requiring refined spatial
962 inputs.

963 **4.3 Sources of Uncertainty in Downscaling Models and Future Research Perspectives**

964 Although the 1 km high-resolution dataset constructed in this study exhibits favorable accuracy
965 and spatiotemporal continuity across global aquifer systems, it is imperative to objectively
966 acknowledge the sources of model uncertainty and structural limitations stemming from the complexity
967 of Earth system observations and the inherent constraints of purely data-driven paradigms.

968 Firstly, the cascading propagation of prior errors in the input feature fields represents a crucial
969 external factor influencing local downscaling precision. Meteorological forcing data (e.g., precipitation
970 and temperature) and surface parameters, which serve as baseline covariates for model predictions,

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

inherently display observational biases in regions characterized by extreme topographical complexity or a scarcity of ground monitoring stations (such as alpine mountain ranges and hyper-arid desert hinterlands). Additionally, the native GRACE satellite data, constrained by the truncation of spherical harmonic coefficients and spatial filtering processes, may introduce signal leakage effects that incorporate mass anomalies from adjacent areas into target pixels (Landerer and Swenson, 2012; Long et al., 2015). These multi-source heterogeneous data uncertainties, existing at both physical and observational levels, can propagate through deep nonlinear mapping networks during high-resolution reconstruction processes, potentially amplifying localized noise or creating spatial artifact biases.

Secondly, purely data-driven downscaling models persistently demonstrate structural limitations in dealing with physical mechanism constraints. Current machine learning architectures are highly reliant on the Scale Invariance Assumption, which postulates that the statistical mapping relationships between coarse 0.25° groundwater storage anomalies (GWSA) and surface environmental covariates can be linearly extrapolated to 1 km scales. Nevertheless, actual groundwater flow dynamics display highly nonlinear characteristics at local scales, propelled by micro-topography and the complex three-dimensional geometry of aquifers. Moreover, purely data-driven models lack explicit constraints from mass conservation laws or groundwater dynamic equations. Consequently, aggregated 1 km water storage values within specific closed basins may encounter difficulties in attaining absolute physical mass closure with the original 0.25° observations. Additionally, for sudden localized water resource management interventions or extreme land-use changes that occur outside the temporal range of the training dataset, purely statistical models lacking explicit anthropogenic proxy variables frequently fail to offer accurate real-time dynamic feedback (Kratzert et al., 2019).

Thirdly, the residual attribution framework proposed in Section 3.4 entails its own inherent structural uncertainties. By subtracting GHM-simulated natural backgrounds from the downscaled GWSA, the resulting residuals are assumed to represent human-induced depletion. However, these residuals inevitably incorporate the structural errors of the GHMs themselves—such as biases in precipitation routing, evapotranspiration parameterization, and soil moisture accounting under extreme climatic stress. Therefore, these residuals should be interpreted as observation-constrained approximations of anthropogenic impact, rather than isolated, absolute pumping signals.

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

It is also crucial to clarify the role of the Specific Yield (S_y) parameter used during the independent validation phase. The published 1 km GWSA dataset is provided strictly in units of Equivalent Water Height (EWH). The regional effective average S_y was introduced solely as an intermediate dimensionality conversion parameter to translate satellite-derived EWH into Groundwater Level (GWL) variations for validation against in situ well data. While this approach inevitably introduces uncertainties related to the 3D heterogeneity of local aquifers (Rodell et al., 2009; Scanlon et al., 2012), preliminary sensitivity tests demonstrated that even when introducing ± 20% random perturbations to the empirical S_y averages across subregions, the significant linear correlation between model predictions and observed in situ water level trends ($p < 0.01$) remained robust. This confirms that the macro-scale evolutionary trends revealed in this dataset are fundamentally driven by actual hydro-climatic forcings, rather than artifacts of the dimensionality conversion process.

To address the aforementioned uncertainties and limitations, future high-resolution hydrological reconstruction endeavors can be improved via two dimensions: multi-source data fusion and algorithmic architecture evolution. On one hand, incorporating explicit, high-spatiotemporal-resolution proxy variables of human activities, such as high-frequency land subsidence data derived from Synthetic Aperture Radar Interferometry (InSAR)—into model forcing inputs can substantially enhance the algorithms' capability to directly represent intensive anthropogenic extraction (Castellazzi et al., 2016). On the other hand, integrating hydrological physical prior knowledge into data-driven black-box models has become a critical research frontier for overcoming statistical generalization bottlenecks. Developing Physically Informed Neural Networks (PINNs) that strictly enforce regional water balance closure (Willard et al., 2022), or establishing loose coupling frameworks with high-resolution 3D groundwater flow models (e.g., MODFLOW), represent promising pathways. This integrated research paradigm, bridging massive Earth observation data with classical hydrodynamic laws, will provide the essential scientific foundation for constructing global groundwater digital twin systems that combine high spatial fidelity with robust physical interpretability.

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

5. Conclusion

To overcome the spatial resolution limitations of current satellite gravimetry, this study presents a globally seamless, 1 km resolution monthly groundwater storage anomaly (GWSA) dataset spanning 2002 to 2020. The dataset was generated using an aquifer-stratified Extreme Gradient Boosting (XGBoost) downscaling framework, integrated with Singular Spectrum Analysis (SSA) and 19 high-resolution environmental covariates.

The scientific reliability of this dataset is firmly established through rigorous multi-scale quality control. This includes cross-scale consistency evaluations against native GRACE observations and independent validation utilizing 1,518 global in situ monitoring wells combined with a geographically zoned specific yield matrix.

Unlike traditional coarse-resolution global hydrological models, this 1 km product uniquely resolves the fine-scale spatial heterogeneity of deep aquifer systems. It successfully captures sharp, nonlinear boundaries driven by localized intensive agricultural pumping, revealing critical depletion hotspots that were previously obscured by spatial smoothing. Ultimately, this publicly available dataset provides a robust, observation-constrained physical benchmark for regional water resource management, Earth system modeling, and the development of groundwater digital twins.

6. Data available

The datasets used to develop the High-resolution Global Groundwater Storage Anomalies Dataset are publicly available and are described in detail in the Data Sources section. The monthly downscaled GWSA data and trend products can be previewed through the Google Earth Engine application (https://yifeifan.projects.earthengine.app/view/high-resolution-gwsa). In addition, we provide the original high-precision dataset in multi-band TIFF format, covering the period from April 2002 to December 2020. The dataset is openly available at https://doi.org/10.12072/ncdc.hydrology.db7323.2026 (Fan et al., 2026).

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

7. Author contributions

Chen Zhang: Writing – review & editing, Writing – original draft, Data curation, Visualization, Formal analysis, Investigation. Yifei Fan: Writing – review & editing, Writing – original draft, Data curation, Visualization, Formal analysis, Investigation. Wenxia Han: Writing – review & editing, Writing – original draft, Funding acquisition, Supervision. Tao Che: Writing – review & editing, Writing – original draft, Funding acquisition, Supervision. Jiabei Zhu: Formal analysis, Supervision. Yingchun Ge: Conceptualization, Methodology, Supervision, Investigation, Formal analysis.

8. Declaration of Competing 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.

9. Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant no. 42595534), the Gansu Provincial Science and Technology Program (grant no. 23ZDKA0004), the Qinghai Province Major Science and Technology Project (grant no. 2025-SF-A2), and financial support from Qinghai Province through the Kunlun Talent Program for Wenxia Han. We sincerely acknowledge the National Earth System Science Data Center and the China Centre for Resources Satellite Data and Application for providing the data used in this study. We also acknowledge the Google Earth Engine platform for providing cloud-based geospatial data access and processing capabilities.

Reference

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M.: Optuna: A Next-generation Hyperparameter Optimization Framework, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2623–2631, https://doi.org/10.1145/3292500.3330701, 2019.

Ali, S., Pham, Q. B., Sajjad, A., et al.: A comparative study of machine learning models for

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

groundwater storage anomaly downscaling, Journal of Hydrology, 620, 129424, https://doi.org/10.1016/j.jhydrol.2023.129424, 2023.

Bierkens, M. F. P. and Wada, Y.: Non-renewable groundwater use and groundwater depletion: a review, Environ. Res. Lett., 14, 063002, https://doi.org/10.1088/1748-9326/ab1a5f, 2019.

Breiman, L.: Random forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Campbell, P. C., et al.: Global-scale evaluation of the GLDAS-2.1 Noah Land Surface Model, Journal of Hydrometeorology, 20, 819–839, https://doi.org/10.1175/JHM-D-18-0133.1, 2019.

Castellazzi, P., Martel, R., Rivera, A., Huang, J., Pavlic, G., Calderhead, A. I., Chaussard, E., Garfias, J., and Salas, J.: Groundwater depletion in Central Mexico: Use of GRACE and InSAR to support water resources management, Water Resour. Res., 52, 5985–6003, https://doi.org/10.1002/2015WR018211, 2016.

Chen, J., Wilson, C. R., Famiglietti, J. S., and Rodell, M.: Attenuation effects on seasonal signals from GRACE time-variable gravity, Journal of Geodesy, 81, 237–245, https://doi.org/10.1007/s00190-006-0104-2, 2007.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Chen, Y., Zhang, Z., and Khan, S.: Understanding the impact of irrigation on groundwater storage using XGBoost, Agricultural Water Management, 240, 106297, https://doi.org/10.1016/j.agwat.2020.106297, 2020.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I.: STL: A seasonal-trend decomposition procedure based on loess, Journal of Official Statistics, 6, 3–73, 1990.

Cooley, S. W., Ryan, J. C., and Smith, L. C.: Human alteration of global surface water storage variability, Nature, 591, 78–81, https://doi.org/10.1038/s41586-021-03262-3, 2021.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J. D.: Implementation of Noah land surface model changes in the National Centers for Environmental Prediction operational mesoscale Eta model, J. Geophys. Res., 108, 8851,

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

https://doi.org/10.1029/2002JD003296, 2003.

Famiglietti, J. S.: The global groundwater crisis, Nat. Clim. Change, 4, 945–948, https://doi.org/10.1038/nclimate2425, 2014.

Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K. J., Syed, T. H., Swenson, S. C., de Linage, C. R., and Rodell, M.: Satellites measure recent rates of groundwater depletion in California's Central Valley, Geophys. Res. Lett., 38, L03403, https://doi.org/10.1029/2010GL046442, 2011.

Fan Yifei, Zhang Chen, Han Wenxia, Che Tao, Ge Yingchun. 1km high-resolution monthly-scale GWSA dataset. National Cryosphere Desert Data Center(http://www.ncdc.ac.cn), 2026. https://cstr.cn/CSTR:11738.11.ncdc.hydrology.db7323.2026.

Girotto, M. and Rodell, M.: Terrestrial water storage anomalies and the Earth's energy imbalance, Current Climate Change Reports, 5, 290–301, https://doi.org/10.1007/s40641-019-00140-5, 2019.

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edn., Springer, New York, NY, USA, 745 pp., 2009.

Jasechko, S., Seybold, H., Perrone, D., Fan, Y., Kirchner, J. W., Taylor, R. G., and Tindimugaya, C.: Rapid groundwater decline and some cases of recovery in aquifers globally, Nature, 625, 715–721, https://doi.org/10.1038/s41586-023-06879-8, 2024.

Karger, D. N., Conrad, O., Böhner, J., Kawege, T. W., Kreft, H., RW, M., Zimmermann, N. E., and Kessler, M.: Climatologies at high resolution for the earth’s land surface areas, Scientific Data, 4, 170122, https://doi.org/10.1038/sdata.2017.122, 2017.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L.: Physics-informed machine learning, Nat. Rev. Phys., 3, 422–440, https://doi.org/10.1038/s42254-021-00314-5, 2021.

Kendall, M. G.: Rank Correlation Methods, 4th Edn., Charles Griffin, London, 1975.

Khan, S., et al.: Global groundwater storage anomaly downscaling using tree-based ensembles, Journal of Hydrology, 630, 130635, https://doi.org/10.1016/j.jhydrol.2024.130635, 2024.

Kondrashov, D. and Ghil, M.: Spatio-temporal filling of missing points in geophysical data sets, Nonlinear Processes in Geophysics, 13, 151–159, https://doi.org/10.5194/npg-13-151-2006, 2006.

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

1133 Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning
1134 universal, regional, and local hydrological behaviors via machine learning applied to large-sample
1135 datasets, Hydrol. Earth Syst. Sci., 23, 5089–5110, https://doi.org/10.5194/hess-23-5089-2019,
1136 2019.
1137 Landerer, F. W. and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates,
1138 Water Resour. Res., 48, W04531, https://doi.org/10.1029/2011WR011453, 2012.
1139 LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436–444,
1140 https://doi.org/10.1038/nature14539, 2015.
1141 Li, J., et al.: Global quantification of groundwater seasonality and its contribution to total storage
1142 variability, Earth System Science Data, 17, 2025.
1143 Long, D., Longuevergne, L., and Scanlon, B. R.: Global analysis of spatiotemporal variability in
1144 merged total water storage changes using multiple GRACE products and global hydrological
1145 models, Remote Sens. Environ., 170, 117–133, https://doi.org/10.1016/j.rse.2015.09.016, 2015.
1146 Mankoff, K. D., Solgaard, A., Colgan, W., Ahlstrøm, A. P., Khan, S. A., and Fausto, R. S.: Greenland
1147 Ice Sheet solid ice discharge from 1986 through 2019, Earth Syst. Sci. Data, 12, 1367–1383,
1148 https://doi.org/10.5194/essd-12-1367-2020, 2020.
1149 Mann, H. B.: Nonparametric tests against trend, Econometrica, 13, 245–259,
1150 https://doi.org/10.2307/1907187, 1945.
1151 Miro, M. E. and Famiglietti, J. S.: Downscaling GRACE remote sensing datasets to high-resolution
1152 groundwater storage changes in California's Central Valley, Remote Sens., 10, 143,
1153 https://doi.org/10.3390/rs10010143, 2018.
1154 Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat,
1155 E., Portmann, F. T., Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C.-E., Trautmann, T., and
1156 Döll, P.: The global water resources and use model WaterGAP v2.2d: model description and
1157 evaluation, Geosci. Model Dev., 14, 1037–1079, https://doi.org/10.5194/gmd-14-1037-2021,
1158 2021.
1159 Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K.,
1160 Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, J. Geophys. Res., 116, D12109, https://doi.org/10.1029/2010JD015139, 2011.

Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S.: High-resolution mapping of global surface water and its long-term changes, Nature, 540, 418–422, https://doi.org/10.1038/nature20584, 2016.

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the world with quantified uncertainty, SOIL, 7, 217–244, https://doi.org/10.5194/soil-7-217-2021, 2021.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Rodell, M. and Famiglietti, J. S.: An analysis of terrestrial water storage variations in Illinois with implications for the Gravity Recovery and Climate Experiment (GRACE), Water Resour. Res., 37, 2327–2339, https://doi.org/10.1029/2000WR900306, 2001.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, Bull. Amer. Meteorol. Soc., 85, 381–394, https://doi.org/10.1175/BAMS-85-3-381, 2004.

Rodell, M., Velicogna, I., and Famiglietti, J. S.: Satellite-based estimates of groundwater depletion in India, Nature, 460, 999–1002, https://doi.org/10.1038/nature08238, 2009.

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., and Lo, M. H.: Emerging trends in global freshwater availability, Nature, 557, 651–659, https://doi.org/10.1038/s41586-018-0123-1, 2018.

Save, H., Bettadpur, S., and Tapley, B. D.: High-resolution CSR GRACE RL05 mascons, J. Geophys. Res. Solid Earth, 121, 3544–3564, https://doi.org/10.1002/2016JB013007, 2016.

Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., and Anyamba, A.: Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley, Proc. Natl. Acad. Sci. U.S.A., 109, 9320–9325,

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

1189 https://doi.org/10.1073/pnas.1200311109, 2012.

1190 Scanlon, B. R., Zhang, Z., Save, H., Wiese, D. N., Landerer, F. W., Royetti, D., Longuevergne, L., and
1191 Chen, J.: Global evaluation of new GRACE mascon products for hydrologic applications, Water
1192 Resour. Res., 52, 9412–9429, https://doi.org/10.1002/2016WR019494, 2016.

1193 Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Müller Schmied, H., van Beek, L. P. H., Wiese, D. N.,
1194 Wada, Y., Long, D., Reedy, R. C., Longuevergne, L., Döll, P., and Bierkens, M. F. P.: Global
1195 models underestimate large decadal declining and rising water storage trends relative to GRACE
1196 satellite data, Proc. Natl. Acad. Sci. U.S.A., 115, E1080–E1089,
1197 https://doi.org/10.1073/pnas.1704665115, 2018.

1198 Sen, P. K.: Estimates of the regression coefficient based on Kendall's tau, Journal of the American
1199 Statistical Association, 63, 1379–1389, https://doi.org/10.1080/01621459.1968.10480934, 1968.

1200 Seyoum, W. M., Kwon, D., and Milewski, A. M.: Downscaling GRACE TWSA data into
1201 high-resolution groundwater level anomaly using machine learning-based models in a glacial
1202 aquifer system, Remote Sens., 11, 824, https://doi.org/10.3390/rs11070824, 2019.

1203 Shen, C.: A transdisciplinary review of deep learning research and its relevance for water resources
1204 scientists, Water Resour. Res., 54, 8558–8593, https://doi.org/10.1029/2018WR022643, 2018.

1205 Sun, A. Y., Scanlon, B. R., AghaKouchak, A., and Zhang, Z.: Using Big Data to Advance Our
1206 Understanding of the Earth's Hydrologic Cycle: A Case Study in Space-Based Monitoring of
1207 Water Resources, Front. Earth Sci., 7, 214, https://doi.org/10.3389/feart.2019.00214, 2019.

1208 Theil, H.: A rank-invariant method of linear and polynomial regression analysis, Nederl. Akad.
1209 Wetensch. Proc., 53, 1397–1412, 1950.

1210 Tripathi, V., Groh, A., Horwath, M., and Ramsankaran, R.: Scaling methods of leakage correction in
1211 GRACE mass change estimates revisited for the complex hydro-climatic setting of the Indus
1212 Basin, Hydrol. Earth Syst. Sci., 26, 4515–4535, https://doi.org/10.5194/hess-26-4515-2022, 2022.

1213 Tyralis, H., Papacharalampous, G., and Langousis, A.: A brief review of random forests for water
1214 scientists and practitioners and their recent history in water resources, Water, 11, 910,
1215 https://doi.org/10.3390/w11050910, 2019.

1216 Vautard, R., Yiou, P., and Ghil, M.: Singular-spectrum analysis: A toolkit for short, noisy chaotic

https://doi.org/10.5194/essd-2026-409
Preprint. Discussion started: 24 June 2026

1217 signals, Physica D: Nonlinear Phenomena, 58, 95–126,
1218 https://doi.org/10.1016/0167-2789(92)90103-T, 1992.
1219 Wada, Y., van Beek, L. P. H., van Kempen, C. M., Reckman, J. W. T. M., and Bierkens, M. F. P.: Global
1220 depletion of groundwater resources, Geophys. Res. Lett., 37, L20402,
1221 https://doi.org/10.1029/2010GL044571, 2010.
1222 Wang, J., Yin, W., Hu, L., Zhang, M., and Han, S. C.: Spatiotemporal evaluation of GRACE-derived
1223 groundwater storage anomalies using machine learning, Journal of Hydrology, 612, 128219,
1224 https://doi.org/10.1016/j.jhydrol.2022.128219, 2022.
1225 Weber, T., et al.: Global depletion of groundwater is a critical threat to water security, Environmental
1226 Research Letters, 11, 064003, https://doi.org/10.1088/1748-9326/11/6/064003, 2016.
1227 Willard, J., Jia, X., Xu, S., Steinbach, M., and Kumar, V.: Integrating physics-based modeling with
1228 machine learning: A survey, ACM Comput. Surv., 55, 1–34, https://doi.org/10.1145/3514228,
1229 2022.
1230 Yi, S. and Sneeuw, N.: Filling the data gaps within and between GRACE and GRACE-FO missions
1231 using singular spectrum analysis, Geophysical Journal International, 224, 1393–1411,
1232 https://doi.org/10.1093/gji/ggaa535, 2021.
1233 Yin, W., Hu, L., Zhang, M., Wang, J., and Han, S. C.: Statistical downscaling of GRACE-derived
1234 groundwater storage using evapotranspiration and precipitation, J. Hydrol., 612, 128219,
1235 https://doi.org/10.1016/j.jhydrol.2022.128219, 2022.
1236 Zhang, Y., Kong, D., Gan, R., Chiew, F. H. S., McVicar, T. R., Zhang, Q., and Yang, Y.: Coupled
1237 estimation of 500 m and 8-day resolution global actual evapotranspiration and gross primary
1238 production in 2002–2017, Remote Sensing of Environment, 222, 165–182,
1239 https://doi.org/10.1016/j.rse.2018.12.031, 2019.
1240 Zhang, Z., et al.: Global groundwater storage anomaly downscaling using tree-based ensembles,
1241 Journal of Hydrology, 601, 126620, https://doi.org/10.1016/j.jhydrol.2021.126620, 2021.
1242 Zheng, H., et al.: A global daily surface soil moisture dataset at 1 km resolution, Scientific Data, 10,
1243 143, https://doi.org/10.1038/s41597-023-01991-w, 2023.