Reconstruction of Climate-Driven Global Terrestrial Water Storage Variations (2002–2021) Using a Four-Parameter Linear Recursive Model

Pu Xie, Shuang Yi*

National Key Laboratory of Earth System Numerical Modeling and Application, College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, 101408, China

* Corresponding email: s.yi@ucas.ac.cn

Abstract

Terrestrial water storage anomalies (TWSA), jointly influenced by climatic variability and human activities, ~~serve as a key indicator of global climate change. TWSA~~ exhibits pronounced fluctuations across multiple temporal scales.~~,~~ ~~a~~A substantial portion of ~~which~~ the fluctuations ~~can~~ is ~~be~~ attributed to climatic variability, ~~such as~~like the El Niño–Southern Oscillation (ENSO). Empirical reconstruction of climate-driven water storage based on relationships between GRACE satellite gravity observations and meteorological forcing data has become a common approach; however, existing models often neglect the regulating role of temperature in the transformation of precipitation into water storage. In this study, we propose a linear, four-parameter coupled recursive model that explicitly incorporates temperature effects on both the conversion and dissipation efficiency of water storage. Using GRACE/GRACE-FO satellite observations and meteorological forcing data, we reconstructed climate-driven TWSA over the global land grid (excluding Antarctica) at a ~~daily~~ monthly temporal resolution and 0.5° spatial resolution for the period 2002 to 2021. For 116 major global river basins, we further derived basin-scale TWSA reconstructions and quantitatively evaluated the fraction of precipitation converted into TWSA. ~~Finally, the reconstructed data were compared with existing reconstruction datasets.~~ Compared with existing statistical reconstruction products, the results indicate that: (1) the proposed method achieves substantially faster parameter convergence, improving computational efficiency by several tens of times during the TWSA reconstruction process; (2) the proposed model demonstrates superior performance in approximately 89% of river basins and 62% of global land grid cells. ~~Specifically, the Nash–Sutcliffe efficiency (NSE) exceeds 0.7 in 84 out of 116 basins, and 62% of global land grids exhibit NSE values greater than 0.~~ Additional comparisons with the physically based Catchment Land Surface Model (CLSM) product from NASA's Global Land Data Assimilation System (GLDAS) show that the proposed method better captures the temporal variability of GRACE TWSA in most basins. At the daily scale, the reconstructed TWSA agrees well overall with the ITSG-Grace2018 daily solution and GLDAS-2.2. ~~In addition, our reconstruction demonstrates comparable performance during independent training and validation periods, indicating good robustness and limited overfitting.~~ This study enhances the understanding of the mechanisms governing terrestrial water storage variations at both global and regional scales, provides a quantitative assessment of climate-driven water storage changes, and offers a solid foundation for disentangling the respective impacts of climatic variability and human activities on water resources.

Keywords: GRACE, GRACE-FO, Reconstruction, TWSA, Climate variability

42

1 Introduction

Terrestrial water storage (TWS) is the total amount of water stored in all components, including surface water (e.g., lakes and reservoirs), groundwater, soil moisture, snow and ice, and vegetation water, and it plays a vital role in the global hydrological cycle. Monitoring TWS is essential not only for understanding the climate system and its variability (Tapley et al., 2019), but also for supporting industrial development and ensuring global food security (Abbott et al., 2019; Rodell et al., 2009; Rodell et al., 2018). Temperature and precipitation are the primary climatic drivers of TWS variations. For instance, precipitation anomalies induced by ENSO are a dominant cause of interannual TWS variability (Liu et al., 2020; Ni et al., 2018).

Traditionally, changes in TWS have been estimated using two main approaches. The first is the water balance method, which requires precipitation, evapotranspiration, and runoff data. Precipitation is typically measured by rain gauges, evapotranspiration is monitored using eddy covariance systems (Baldocchi et al., 2001; Pastorello et al., 2020; Ma et al., 2024), and runoff is estimated from discharge records at hydrological gauging stations (Duan et al., 2024). The rate of TWS change ($dTWS/dt$) is then inferred from the water balance ~~Eq~~equation. However, because ground-based observation stations are spatially discrete, this method can only provide localized, point-scale information. In regions constrained by limited financial resources or political barriers to data access, the sparse distribution of monitoring sites further hinders the establishment of continuous spatial coverage (Crochemore et al., 2020; Fekete et al., 2012; Laudon et al., 2017). The second is the model-based approach, in which land surface or hydrological models simulate individual storage components such as soil moisture and groundwater, which are then aggregated to estimate total TWS variations. However, both instrumental and measurement errors can propagate through the modeling process, introducing substantial uncertainties into the final TWS estimates (Long et al., 2015; Humphrey et al., 2023). Moreover, the uncertainty associated with hydrological model outputs remains difficult to quantify (Zheng et al., 2023).

Since its launch in 2002, the Gravity Recovery and Climate Experiment (GRACE) mission has provided precise measurements of temporal variations in Earth's gravity field and delivered monthly gravity field solutions (Tapley et al., 2019; Chen et al., 2022; Humphrey et al., 2023). Compared with traditional approaches that rely on sparse ground-based observations or hydrological models with large uncertainties, GRACE data have substantially improved the spatial continuity and observational reliability of Earth's gravity measurements, providing a crucial tool for monitoring global water storage changes. A major contributor to temporal variations in the terrestrial gravity field is the variations in TWS, which is jointly driven by natural climate variability, long-term

climate change, and human water use (Rodell et al., 2018; An et al., 2021). Distinguishing anthropogenic signals from natural variability in GRACE-observed TWS has therefore become an active research focus (An et al., 2021; Liu et al., 2021; Yi et al., 2016; Zhang et al., 2025).

Currently, reconstruction methods for climate-driven TWSA can be broadly classified into two categories. The first category is data-driven approaches, which establish nonlinear mappings between GRACE-based TWSA and climate–hydrological variables using machine learning or neural network models (Fan et al., 2021; Palazzoli et al., 2025; Li et al., 2021; Sun et al., 2020; Zhang et al., 2016; Long et al., 2014; Sun et al., 2021; Yin et al., 2023; Li et al., 2020). The advantage of such approaches lies in their ability to capture complex patterns without requiring prior assumptions about the underlying physical processes (Long et al., 2014) and without being constrained by the spatial resolution of the input drivers. For instance, Zhang et al. (2016) reconstructed TWSA in the Yangtze River Basin for the period from 1979 to 2012 using an artificial neural network driven by precipitation and ERA-Interim soil moisture, and subsequently estimated drought recovery times. Sun et al. (2021) introduced an automated machine learning framework for GRACE-based TWSA reconstruction, integrating GLDAS model outputs together with meteorological and climatic predictors to combine physical constraints with data-driven learning. Palazzoli et al. (2025) developed unidirectional and bidirectional Long Short-Term Memory networks to construct four climate-driven models, generating a continuous global TWSA record at 0.5° resolution for the period from 1984 to 2021. Li et al. (2021) combined statistical analysis, time series decomposition, and machine learning techniques to reconstruct global (excluding Antarctica) gridded TWSA for the period from 1979 to 2020 using multiple predictors, such as sea surface temperature and climate indices. Yin et al. (2023) also adopted this approach and incorporated a broader range of predictors in addition to conventional climatic variables, including land use data and vegetation indices, to reconstruct a continuous global monthly TWSA dataset at 0.5° spatial resolution for the period 1940 to 2022.

The second category of methods involves data-driven statistical models for reconstructing TWSA (Zhong et al., 2019; Zhong et al., 2025; Humphrey and Gudmundsson, 2019; Xiao et al., 2025). Humphrey and Gudmundsson (2019) proposed a statistical reconstruction approach that uses only precipitation and temperature as predictors to estimate climate-driven TWS at both daily and monthly time scales over the past century. Their method not only successfully filled the ~1-year observational gap between the GRACE and GRACE-FO missions but also demonstrated performance comparable to or exceeding that of several global hydrological models. However, their model incorporates a mathematical fitting procedure to align the reconstructed TWSA with GRACE

observations. This adjustment improves amplitude consistency but does not provide a clear physical interpretation. Building on this framework, Zhong et al. (2025) introduced a key modification by incorporating the concept of the "mean daily fraction of precipitation converted to TWS", thereby adding a physically interpretable representation of the precipitation-to-storage transformation process. Based on this enhancement, they re-estimated that approximately 64% of terrestrial precipitation contributed to TWS across 121 global river basins during the period from 2002 to 2021, with substantial spatial variability linked to climatic and geographic conditions. The specific formulations and reconstruction workflows of both statistical models are detailed in Sections 3.1 and 3.2 of this study.

Expanding upon the frameworks of Humphrey and Gudmundsson (2019) and Zhong et al. (2025), this study further advances parameter design. Whereas previous models relied on two empirical parameters in a residence-time formulation that is only indirectly linked to individual water-balance components~~without clear physical meaning~~, we propose a linear, four-parameter coupled recursive model at the daily time scale. This model explicitly incorporates temperature modulation of both the immediate ~~infiltration~~ conversion efficiency of precipitation and~~, as well as~~ the retention rate of water storage~~ and its temperature sensitivity~~. The model includes four parameters.~~, all of which have clear physical interpretation.~~ The four model parameters are designed to be more closely tied to these hydrological processes. Detailed interpretations of the parameters are provided in Section 3.2.1. In addition, the approach has been optimized for computational efficiency and numerical stability. Using this method, we reconstructed a global (excluding Antarctica) gridded TWSA dataset at 0.5° spatial resolution and monthly time step for the period from 2002 to 2021 (the product name is provided in Section 4.1). The reconstructed TWSA closely matches GRACE observations and achieves accuracy levels comparable to or exceeding those of existing reconstruction products, as demonstrated in comprehensive benchmark evaluations.

2 Data

Table 1 List of the data sets

Dataset	Time	Spatial resolution	Temporal resolution
GSFC RL06v2.0 mascon solution	2002—2021	1°equal area (provided on 0.5°)	monthly

JPL CRI Filtered RL06.1v03 mascon solution	2002—~~2021~~2023	3.0°equal area (provided on 0.5°)	monthly
CSR RL06v2.0 mascon solution	2002—2021	1°equal area (provided on 0.25°)	monthly
Precipitation (ERA5-land)	2000—202~~3~~5	0.1°	hourly
Precipitation (MSWEP v2.8)	2000—2021	0.1°	daily
Temperature (ERA5-land)	2000—202~~3~~5	0.1°	hourly
Temperature (GLDAS~~-v-~~2.2)	2003—2021	0.25°	daily
Evapotranspiration (ERA5-land)	2000—2020	0.1°	hourly
TWS (GLDAS CLSM)	2003—2021	1°	monthly
TWSA (GLDAS-2.2)	2003—2021	0.25°	daily
GRACE-REC (Humphrey's reconstruction)	2002—2019	0.5°	monthly

2.1 GRACE data

This study employs ~~two~~ three GRACE/GRACE-FO mascon solutions (Table 1): the JPL RL06 v03 mascon product (JPLM) provided by the Jet Propulsion Laboratory (Wiese et al., 2019), the CSR RL06 v2.0 mascon product (CSRM) provided by the Center for Space Research (Save et al., 2016), and the GSFC RL06v2.0 mascon product (GSFCM) provided by the NASA Goddard Space Flight Center (Loomis et al., 2019). All datasets cover the period from 2002 to 2021. The native resolution of JPLM is $3° \times 3°$ equal-area caps, and the data are distributed on a 0.5° latitude and longitude grid. The GSFCM product has a native resolution of $1° \times 1°$ equal-area caps and is also provided on a 0.5° grid. Both mascon products have been preprocessed with geocenter correction,

degree-2 replacement, and GIA correction, and thus require no further post-processing (Zhang and Sun, 2022).

2.2 Precipitation and Temperature Data

This study employs two precipitation products and two temperature products (Table 1). For precipitation, the following datasets are used: (1) The ERA5-Land dataset, developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), is the land component of the ERA5 reanalysis. It is driven by atmospheric forcing variables from ERA5 and provides hourly precipitation data at a spatial resolution of $0.1° \times 0.1°$ (~9 km), covering the period from January 1950 to present (Muñoz-Sabater et al., 2021; Muñoz-Sabater, 2019); (2) The Multi-Source Weighted-Ensemble Precipitation (MSWEP) version 2.8 dataset (Beck et al., 2017; Beck et al., 2019) spans from 1979 to present and offers precipitation estimates at $0.1° \times 0.1°$ spatial resolution and multiple temporal resolutions (3-hourly, daily, and monthly). In this study, the daily product is used. MSWEP combines gauge observations, satellite retrievals, and reanalysis data using a weighted ensemble approach to provide high-accuracy global precipitation estimates (Beck et al., 2017).

For temperature, the following datasets are used: (1) The ERA5-Land reanalysis dataset produced by ECMWF (Muñoz-Sabater, 2019); (2) the GLDAS-2.2 daily CLSM product ~~NASA's Global Land Data Assimilation System (GLDAS) developed the GLDAS v2.2 Catchment Land Surface Model (CLSM) product~~ (Li et al., 2019). This model offers a physically based framework suitable for assimilating GRACE-derived TWS observations. The GLDAS ~~v~~ 2.2 ~~dataset~~ daily product includes both a main version and an early product. This study uses the main version, which spans from February 2003 to ~~July~~ December ~~2025~~ 2021 (as of manuscript preparation), with a spatial resolution of $0.25° \times 0.25°$ and a daily temporal resolution. It should be noted that, due to the data usage agreement with the ECMWF, which prohibits redistribution of products from the Integrated Forecasting System (IFS), the GLDAS-2.2 daily product does not include the meteorological forcing fields. Instead, GLDAS-2.2 provides land surface variables simulated by the Catchment land surface model, which is forced by meteorological analysis fields from the operational ECMWF IFS.

Although both ERA5-Land and GLDAS-2.2 are ultimately based on meteorological information produced within the ECMWF framework, their temperature datasets exhibit a discrepancy of approximately 1–2 °C (Fig. S1) due to different data assimilating methods and models.

ERA5-Land precipitation and temperature are used as the primary meteorological inputs in this study. The MSWEP precipitation and GLDAS-2.2 temperature datasets are additionally used to

compare standard deviations with ERA5-Land, in order to assess the uncertainty of ERA5-Land forcing data and support the subsequent parameter uncertainty analysis.

2.3Ancillary Data

This study employs basin boundaries from the publicly available dataset "Major River Basins of the World" provided by the Global Runoff Data Centre (GRDC). This dataset has been widely used in large-scale hydrological studies (e.g., Burek and Smilovic (2023); Zhong et al. (2025); Gao et al. (2026)). We followed the strategy of Zhong et al. (2025) and ranked all basins according to drainage area to ensure spatial representativeness and avoid subjective selection. The 116 largest basins with drainage areas exceeding $10^5\ km^2$ were selected for analysis. ~~This study selected 116 river basins, with basin boundaries obtained from the Global Runoff Data Centre (GRDC).~~

The classification of basin climate types was based on the mean annual aridity index, derived from the Global Aridity Index and Potential Evapotranspiration Climate Database (Zomer et al., 2022; Trabucco and Zomer, 2019), which is available in GeoTIFF format. The aridity index is defined as the ratio of annual precipitation to mean annual potential evapotranspiration, where higher values indicate more humid conditions. Climate types were categorized according to the aridity index as follows: arid (aridity index $<$ 0.2), semi-arid (0.2 $\leq$ aridity index $<$ 0.5), semi-humid (0.5 $\leq$ aridity index $<$ 0.65), and humid (aridity index $\geq$ 0.65).

3 Methods

3.1 Previous Statistical Models for TWSA Reconstruction

3.1.1 Humphrey's Reconstruction Model

Precipitation is the primary input to terrestrial water storage. Statistical models that reconstruct TWSA from daily precipitation and temperature can be used to quantify the influence of climatic variability on terrestrial water storage (Humphrey and Gudmundsson, 2019). For comparison purposes, we first revisit the statistical reconstruction model proposed by Humphrey and Gudmundsson (2019) :

$$TWS\,(t) = TWS\,(t-1) \cdot e^{-\frac{1}{\tau(t)}} + P(t)_{\cdot} \quad (1)$$

Here~~where~~ t is the daily time index, $TWS(t)$ is the water storage on day t, $P(t)$ is the precipitation input on day t, and $\tau(t)$ is the water residence time. A larger τ implies slower losses

through evapotranspiration or runoff, while a smaller τ indicates faster depletion.

$\tau(t)$ is defined as a function of temperature to represent the seasonal variations in water residence time.

$$\tau(t) = a + b \cdot T_z(t), \tag{2}$$

~~where~~ with a and b are model parameters to be calibrated.

To reduce the influence of extreme temperatures on model stability, the original temperature $T(t)$ is first detrended and standardized, then transformed using a sigmoid function:

$$T_z = 1 - \tanh\left(\frac{T_0 - \mathrm{mean}(T_0)}{\mathrm{SD}(T_0)}\right). \tag{3}$$

Since subzero temperatures have negligible effects on residence time (as temperature changes exert little influence on storage losses such as evapotranspiration and runoff under freezing or insufficient thawing conditions), the original temperature series $T(t)$ was preprocessed as follows:

$$T_0 = \begin{cases} 0\,, T < 0 \\ T\,, T \geq 0 \end{cases}. \tag{4}$$

The initial water storage $TWS(0)$ is computed as the ratio of long-term mean precipitation to mean water loss rate:

$$TWS(0) = \frac{\mathrm{mean}(P)}{1 - \mathrm{mean}\left(e^{-\frac{1}{\tau(t)}}\right)}. \tag{5}$$

The daily TWS series is aggregated to monthly means $TWS(t_m)$ and calibrated against deseasonalized and detrended GRACE TWSA at the monthly scale. The model parameters a, b and β are calibrated using a Markov Chain Monte Carlo (MCMC) approach, with the objective of minimizing the sum of squared residuals between the simulated and GRACE-derived TWSA：

$$anom\big(GRACE(t_m)\big) = \beta \cdot anom\big(TWS(t_m)\big) + \epsilon. \tag{6}$$

The coefficient β serves as a scaling factor obtained from MCMC calibration, ϵ denotes the residual term, and $anom(\cdot)$ represents the anomaly series after removal of the seasonal cycle and linear trend.

3.1.2 Zhong's Reconstruction Model

The scaling factor β in Eq. (6) lacks clear physical meaning and serves merely as a calibration parameter, resulting in a strong dependence of the reconstruction on GRACE observations. Building on this framework, Zhong et al. (2025) proposed an improved formulation for reconstructing terrestrial water storage :

$$TWS\,(t) = TWS\,(t-1) \cdot e^{-\frac{1}{\tau(t)}} + \beta \cdot P(t). \tag{7}$$

In this formulation, the parameter β has a clear physical interpretation, representing the mean fraction of daily precipitation that contributes to water storage. The residence time $\tau(t)$ is defined as in Eq. (2). It is important to note that this β is fundamentally different from the scale factor used in Eq. (6).

The daily TWS series is aggregated to monthly means $TWS(t_m)$ and calibrated against deseasonalized and detrended GRACE TWSA at the monthly scale. The model parameters a, b and β are also calibrated using the MCMC approach:

$$anom\big(GRACE(t_m)\big) = anom\big(TWS(t_m)\big) + \epsilon. \quad (8)$$

Owing to the high nonlinearity of the estimation model, parameter optimization is prone to being trapped in local optima. To reduce the bias caused by incomplete chain convergence, the MCMC calibration was repeated 50 times for each basin. All parameter sets falling within the $\pm 1 \cdot \sigma$ range were retained, and their averages were taken as the final estimates of parameters a, b and β. The standard deviation of these filtered parameters was then calculated to represent the inter-basin uncertainty of β.

3.2 Proposed Reconstruction Method

3.2.1 Model formulation

The two preceding approaches describe the residence time (Eq. 2) as a temperature-dependent function. However, the parameters a and b in Eq. (1) and (7)~~its definition~~ are obtained by calibration, and they are less directly linked to specific water-balance components.~~ are merely calibration coefficients without explicit physical meaning when substituted into the exponential decay function (Eq. 1 and 7).~~ In this study, our model is derived directly from the basin-scale water balance equation and provided a systematic clarification of both the derivation and the physical interpretation of the parameters.~~we propose a linearized modification of the water storage reconstruction formula, which simultaneously accounts for the temperature-regulated memory term $TWS(t-1)$, and the temperature-dependent precipitation conversion rate, as follows:~~

We start from the water balance equation:

$$TWS(t) - TWS(t-1) = P(t) - ET(t) - R(t) = P(t) - ETR(t), \quad (9)$$

where $P(t)$, $ET(t)$, and $R(t)$ denote precipitation, evapotranspiration, and runoff at time step t, respectively, and $ETR(t) = ET(t) + R(t)$ represents the total loss term. We assume that ETR consists of two components: (1) the contemporaneous precipitation loss ETR_1, i.e., the direct loss occurring as precipitation is converted into water storage; and (2) the storage-release loss ETR_2, i.e., the loss arising from the depletion of antecedent water storage $TWS(t-1)$ in subsequent

periods, which can be expressed as

$$ETR_1(t) = x \cdot P(t), \tag{10}$$

$$ETR_2(t) = y \cdot TWS(t-1), \tag{11}$$

$$ETR(t) = ETR_1(t) + ETR_2(t) = x \cdot P(t) + y \cdot TWS(t-1). \tag{12}$$

The assumption is validated in Fig. S2, where the ETR is plotted against precipitation and TWS respectively. Both plots show significant correlations, with ETR apparently exhibiting a stronger correlation with the current precipitation than the antecedent TWS.

We then performed a multiple linear regression analysis to further test this assumption. In this step, temperature modulation was temporarily ignored and the coefficients x and y were treated as constants. The results show that all basins pass the F-test at the significance level of $p < 0.01$. Moreover, 74 basins (approximately 64% of all basins) exhibit determination coefficients $R^2 > 0.6$ (Fig. S3(a)), indicating that the total loss term ETR can be statistically explained by the linear combination of precipitation $P(t)$ and antecedent storage $TWS(t-1)$.

We compared the full multiple regression model with a reduced model that includes only precipitation $P(t)$ as the predictor to further examine whether antecedent storage $TWS(t-1)$ provides additional explanatory power beyond precipitation alone. Figure S3(b) presents the spatial distribution of the R^2 obtained from the precipitation-only regression, while Fig. S3(c) shows the spatial difference in R^2 between the full model and the reduced model. Although precipitation alone explains a substantial portion of the variability in basin-scale losses, the inclusion of antecedent storage $TWS(t-1)$ improves model performance. Across all basins, the incorporation of $TWS(t-1)$ increases R^2 by approximately 0.1. Therefore, although ETR is strongly correlated with $P(t)$, the results demonstrate that accounting for $TWS(t-1)$ is both statistically meaningful and physically necessary.

On this basis, we consider the modulation effect of temperature on the hydrological cycle by assuming that the proportional coefficients x and y vary with temperature and apply a Taylor expansion, which yields

$$x = \epsilon \cdot f(T_Z) = \epsilon \cdot \left(1 + \alpha T_z + o(T_z^2)\right). \tag{13}$$

Retaining only the first-order approximation gives

$$x = a' \cdot T_z + b', \tag{14}$$

where T_z denotes standardized temperature.

Similarly, we obtain

$$y = c' \cdot T_z + d'. \tag{15}$$

Substituting Eqs. (12), (14), and (15) into the water balance equation (Eq. 9) leads to

$$TWS(t) = (1 - a' \cdot T_z - b') \cdot P(t) + (1 - c' \cdot T_z - d') \cdot TWS(t-1). \quad (16)$$

After merging constant terms simplifying the signs, the recursive formulation adopted in this study is obtained as

$$TWS\,(t) = (a \cdot T_z + b) \cdot P(t) + (c \cdot T_z + d) \cdot TWS\,(t-1), \quad (17)$$

which is the reconstruction model adopted in this study. In this form, parameters a and c represent temperature modulation effects. The term $1 - \mathrm{b}$ can be interpreted as the fraction of precipitation that directly contributes to the loss component $ETR_1(t)$; therefore, parameter b represents the effective proportion of precipitation that contributes to TWS after basin losses. Similarly, $1 - d$ represents the fraction of antecedent storage contributing to the loss term $ETR_2(t)$, and thus parameter d represents the fraction of previous storage retained in the current storage state. We will continue to discuss these parameters in subsequent sections.

~~All four parameters in the Eq. are endowed with clear physical interpretations. Specifically, parameter a reflects the regulating effect of temperature on the conversion of precipitation into water storage. Parameter b represents the proportion of precipitation that directly contributes to TWS when $T_z = 0$. Parameter c quantifies the temperature-mediated influence of antecedent water storage $TWS(t-1)$ on current storage $TWS(t)$. Parameter d represents the proportion of storage retained from the previous time step that is independent of temperature variations.~~

When solving Eq. (1), (7), and (~~9~~17), it is necessary to specify the initial water storage value, $\mathrm{TWS}(0)$. Humphrey and Gudmundsson (2019) and Zhong et al. (2025) assumed that the initial water storage corresponds to an equilibrium state, as defined in Eq. (5). Although using Eq. (5) provides a unique and reproducible starting condition, it inherently assumes that water storage is close to equilibrium at the beginning of the reconstruction period. However, if a basin experiences an extreme hydroclimatic condition at the initial time step, adopting Eq. (5) may introduce systematic biases and consequently affect the MCMC calibration of other parameters. In this study, $TWS(0)$ is also treated as a variable and is simultaneously estimated together with other model parameters through a joint least-squares and MCMC optimization. This approach eliminates the need to assume equilibrium conditions at the start of the reconstruction period and is therefore more suitable for basins under nonequilibrium states, such as those influenced by prolonged droughts or intensive groundwater extraction (see Section 5.1.2 for details).

The calibration procedure in this section follows exactly the same steps as described in Section 3.1.2, with Eq. (8) retained throughout the process. The reconstruction logic of TWSA is illustrated in Fig. 1.

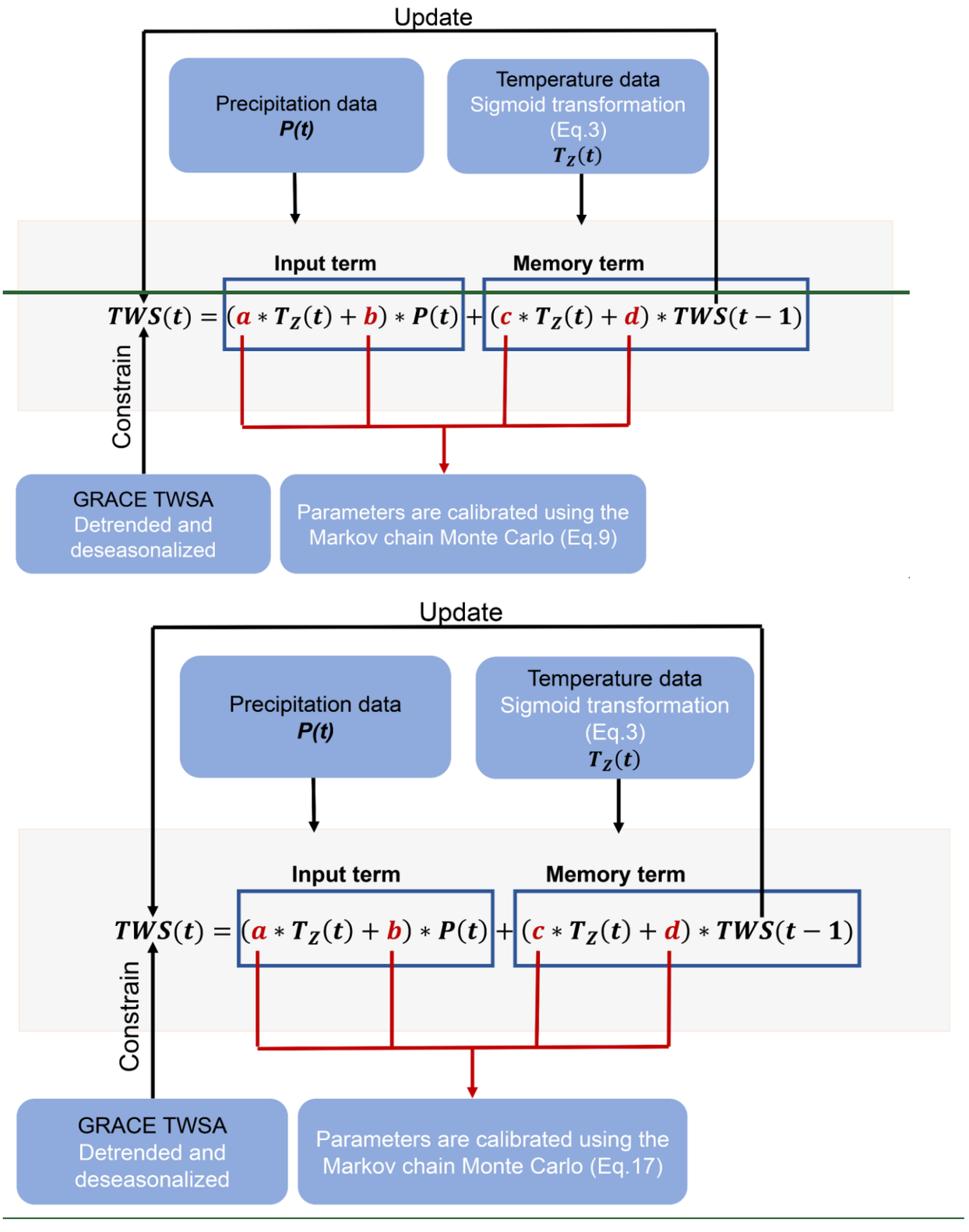


Figure 1: Illustration of the physical meaning of parameters.

Compared with the methods of Humphrey and Gudmundsson (2019) and Zhong et al. (2025), this study linearizes the memory term $TWS(t-1) \cdot e^{-\frac{1}{\tau(t)}}$ as $(c \cdot T_z + d) \cdot TWS(t-1)$. A key

advantage of this linearization is its computational stability and rapid convergence, which substantially improves inversion efficiency. Under the same parameter calibration framework as Zhong et al. (2025), both our model and the Zhong's model were run with 50 independent MCMC chains for the Yellow River to derive the posterior parameter distributions. The parameter dispersion (indicated by the box height) in our model (Fig. 2a) is considerably narrower than that in Zhong's model (Fig. 2b), indicating that our model exhibits smaller parameter uncertainties and faster chain convergence. Consequently, a single MCMC calibration is sufficient for each basin in our method, greatly reducing computation time without a noticeable loss of fitting accuracy. Specifically, on our computing platform (Intel 13th Gen Core i9-13900H processor, 14 cores, base frequency 2.60 GHz, turbo frequency up to ~5.4 GHz, and 16 GB RAM), the Zhong's model required 50 parameter calibrations, taking a total of 353.8 s for a single basin, whereas our method required only one calibration, completing the computation in 6.8 s.

In summary, this study improves upon previous models in three aspects:

(1) Linearization of the memory term. The linearization not only accelerates computational performance but also enhances the physical interpretability of parameters.

(2) Incorporation of the coupling between temperature and precipitation. This mechanism adjusts the fraction of precipitation converted into water storage, leading to a slight improvement in NSE across most basins and substantial enhancement in several cases (see Section 5.2.2).

(3) Consideration of the impact of different initial values on model fitting results (see Section 5.1.2).

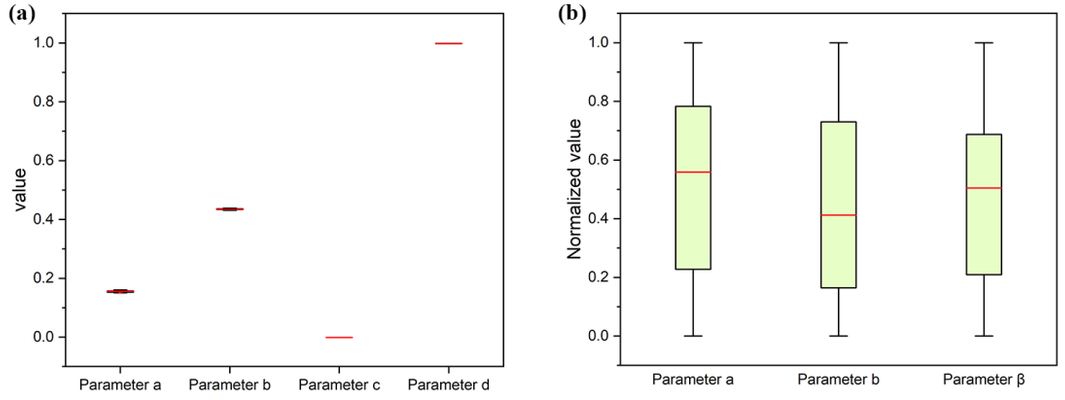


Figure 2: Comparison of parameter stability in the Yellow River Basin between (a) our reconstruction model and (b) the model of Zhong et al. (2025), based on 50 independent calibrations using the MCMC approach. Parameter medians (red lines) are highlighted.

3.2.2 Further interpretation of Our Equations

To demonstrate the compatibility of the proposed model with existing approaches, we consider two extreme cases:

(1) In the model of Zhong et al. (2025), the memory factor is expressed as $e^{-\frac{1}{\tau(t)}}$. When applying a first-order Taylor expansion to the exponential term in Eq. (7), it can be approximated as:

$$TWS\ (t) = \beta \cdot P(t) + \left(1 - \frac{1}{\tau(t)}\right) \cdot TWS\ (t-1). \quad (18\text{~~0~~})$$

Our model (Eq. ~~9~~17) becomes mathematically similar to Eq. (18~~0~~) when the temperature and precipitation modulation is neglected (i.e., $a = 0$). This indicates that our model degenerates to the first-order linear form of the Zhong model when temperature effects are not considered.

(2) Assuming that the memory term is unaffected by temperature ($c = 0$) and exhibits perfect conversion efficiency ($d = 1$), Eq. (~~9~~17) simplifies to:

$$TWSC\ (t) = TWS(t) - TWS(t-1) = (a \cdot T_z + b) \cdot P(t). \quad (19\text{~~1~~})$$

The water balance Eq. is given as:

$$TWSC(t) = P(t) - ET(t) - R(t). \quad (20\text{~~12~~})$$

This implicitly assumes a proportional linear relationship between the loss terms, $ET(t) + R(t)$, and the precipitation input $P(t)$: $R(t) + ET(t) = (1 - a \cdot T_z - b) \cdot P(t)$. In other words, the total output is controlled by the input, so even without explicitly representing the loss terms, the Eq. is still able to account for their integrated contribution.

3.2.3 Uncertainty Estimation

To quantify the uncertainty associated with meteorological forcing data, this study estimates observational errors based on the differences between multiple data sources. Specifically, the standard deviation of daily precipitation differences between ERA5-Land and MSWEP is used as the uncertainty estimate for ERA5-Land precipitation, while the standard deviation of daily temperature differences between ERA5-Land and GLDAS-GLSM is used to estimate the uncertainty in ERA5-Land temperature.

Temperature uncertainty:

$$\sigma_T = std(T_{\mathrm{ERA5-Land}} - T_{\mathrm{GLDAS-GLSM}}). \quad (21\text{~~13~~})$$

Precipitation uncertainty:

$$\sigma_P = std(P_{\mathrm{ERA5-Land}} - P_{\mathrm{MSWEP}}). \quad (23\text{~~14~~})$$

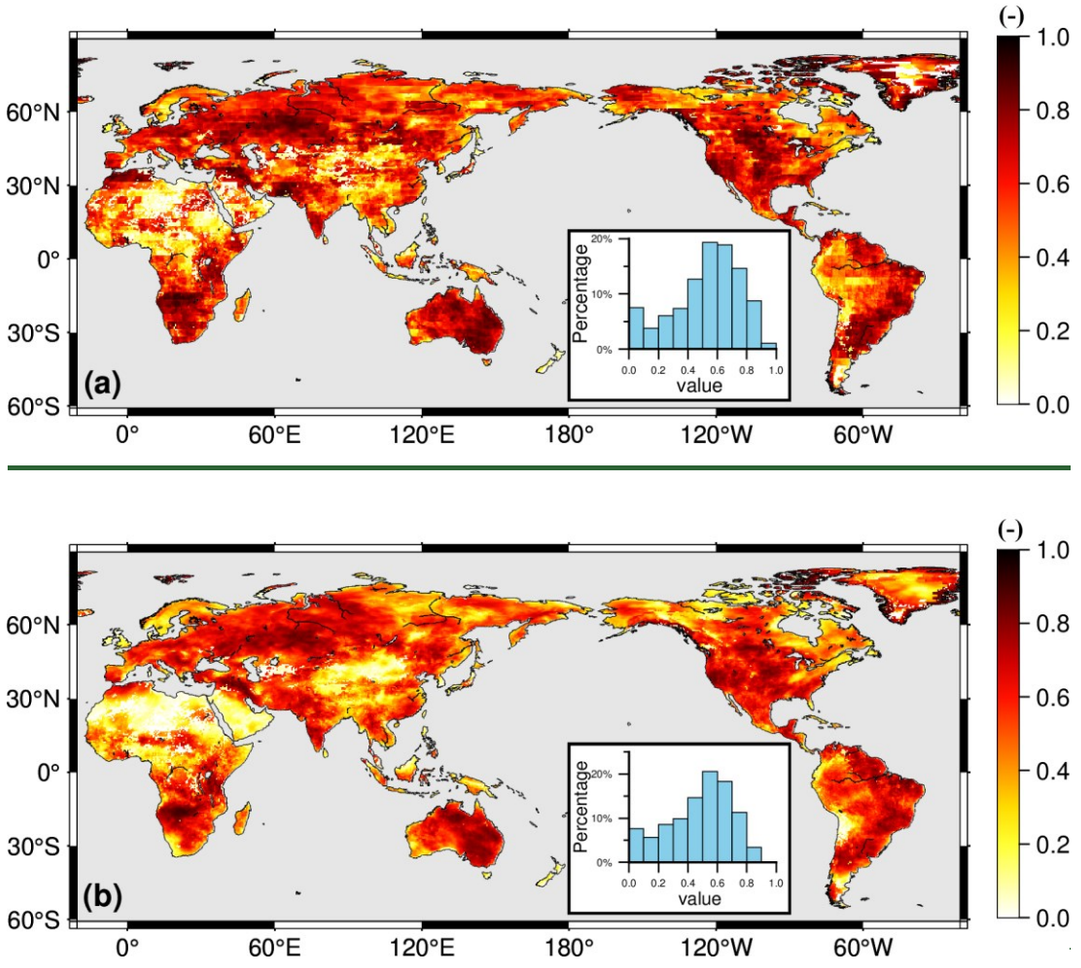
Building on this, we adopt a Monte Carlo simulation framework in which Gaussian noise with zero mean and standard deviations of σ_T and σ_P is added to the temperature and precipitation inputs, respectively. For each perturbed dataset, the model is recalibrated to obtain an optimal set of parameters under the given realization. After completing 1,000 simulations, all sets of estimated parameters are collected to derive the uncertainty distributions of both the model parameters and the reconstructed TWS.

4 Results

4.1 Performance summary

The performance of the GRACE reconstruction model was evaluated using the NSE. Figure 3 presents the spatial distribution of NSE values for 116 major river basins worldwide, and the corresponding numerical results are provided in Table S1. For the reconstruction calibrated against the JPLM product (hereafter referred to as JPL-REC), 53.3% of the 0.5° grid cells exhibited NSE values in the range of 0.5 to 0.8, while 8.6% showed $NSE > 0.8$. High NSE values were primarily concentrated in mid- to high-latitude regions such as North America, Europe, and Australia, where in situ meteorological observations are relatively dense (Fig. 3a). In contrast, regions such as most of Africa and the arid plateau zones of Central Asia showed relatively low NSE values, likely due to the scarcity of precipitation and temperature observations (Xu et al., 2020; Chen et al., 2008). The reconstruction calibrated against the GSFC mascon product (hereafter GSFC-REC) displayed a spatial pattern broadly similar to that of JPL-REC, with 50.2% of grid cells falling within the range from 0.5 to 0.8 and ~~2.81~~2.79% exceeding $NSE > 0.8$. We also compared the reconstruction calibrated against the CSRM product (hereafter referred to as CSR-REC) with the CSRM product to further evaluate robustness of the reconstruction method. The spatial distribution of NSE (Fig. 3c) is broadly consistent with those obtained for JPL-REC and GSFC-REC. The NSE distributions derived from the three mascon-based calibrations are largely consistent, indicating that the reconstruction approach exhibits limited sensitivity to the source of the mascon products, which is consistent with findings reported in previous studies~~The high spatial concordance of NSE between JPL-REC and GSFC-REC indicates that the reconstruction method is largely insensitive to the source of the mascon product, consistent with previous findings~~

(Humphrey and Gudmundsson, 2019).



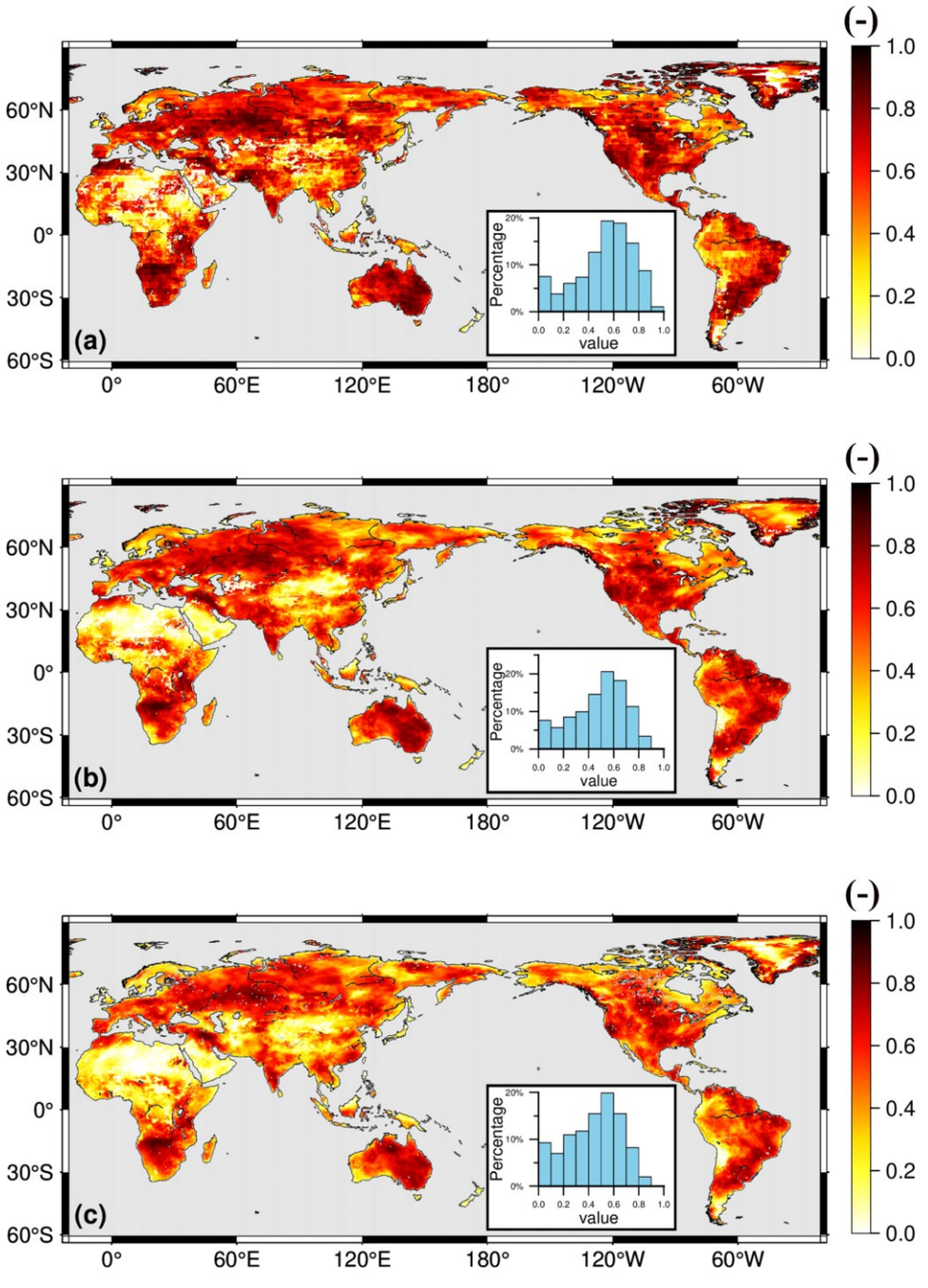


Figure 3: Spatial distribution of NSE (of de-seasonalized, de-trended anomalies) between our reconstruction and (a) JPLM, ~~or~~ (b) GSFCM, and (c) CSRM for the period from 2002 to 2021. Histograms show the distribution of NSE values across all global grid points.

4.2 Comparisons

4.2.1 Comparison with the Humphrey's Model

At the grid scale, we first compared the reconstruction results from this study (JPL-REC and

GSFC-REC) with those of Humphrey and Gudmundsson (2019), which were trained using the JPLM and GSFCM products, respectively (hereafter referred to as Humphrey-JPL-REC and Humphrey-GSFC-REC). Figure 4 presents the spatial distribution of NSE values between the reconstructed and GRACE-derived TWSA after deseasonalization and detrending over the period from 2002 to 2019. Figure 4a and 4d show the NSE performance of JPL-REC and GSFC-REC relative to their corresponding mascon products (JPLM and GSFCM), respectively. Both panels indicate high model performance ($NSE > 0.7$) across large parts of North America, Europe, Australia, and southern Africa, whereas low NSE values ($NSE < 0.4$) are evident in parts of North Africa, Central Asia, and the Arabian Peninsula. Figure 4b and 4e show the NSE performance of Humphrey-JPL-REC and Humphrey-GSFC-REC relative to their corresponding mascon products (JPLM and GSFCM), respectively. While the overall spatial patterns are broadly similar to those in Fig. 4a and 4d, the extent of low-performance areas in Africa and Central Asia is noticeably larger in the Humphrey reconstructions. Figure 4c and 4f present the differences in NSE (ΔNSE), calculated as Fig. 4a minus Fig. 4b and Fig. 4d minus Fig. 4e, respectively. Blue shading ($\Delta NSE > 0$) indicates regions where the present method outperforms the approach by Humphrey and Gudmundsson (2019). Notably, in Fig. 4c, approximately 63% of the grid cells show $\Delta NSE > 0$, with substantial improvements in regions such as the Arabian Peninsula and sub-Saharan Africa, along with scattered enhancements across other continents. Figure 4f demonstrates that under GSFCM constraints, the present method substantially outperforms that of Humphrey and Gudmundsson (2019)，One potential explanation lies in the difference in training datasets: while (Humphrey and Gudmundsson, 2019) utilized the GSFC mascon v2.4 (an older version that is no longer publicly available), the present study employs the updated GSFC mascon RL06v2.0. This difference may account for the discrepancies between our results in Fig. 4e and those reported in the original publication by Humphrey and Gudmundsson (2019).

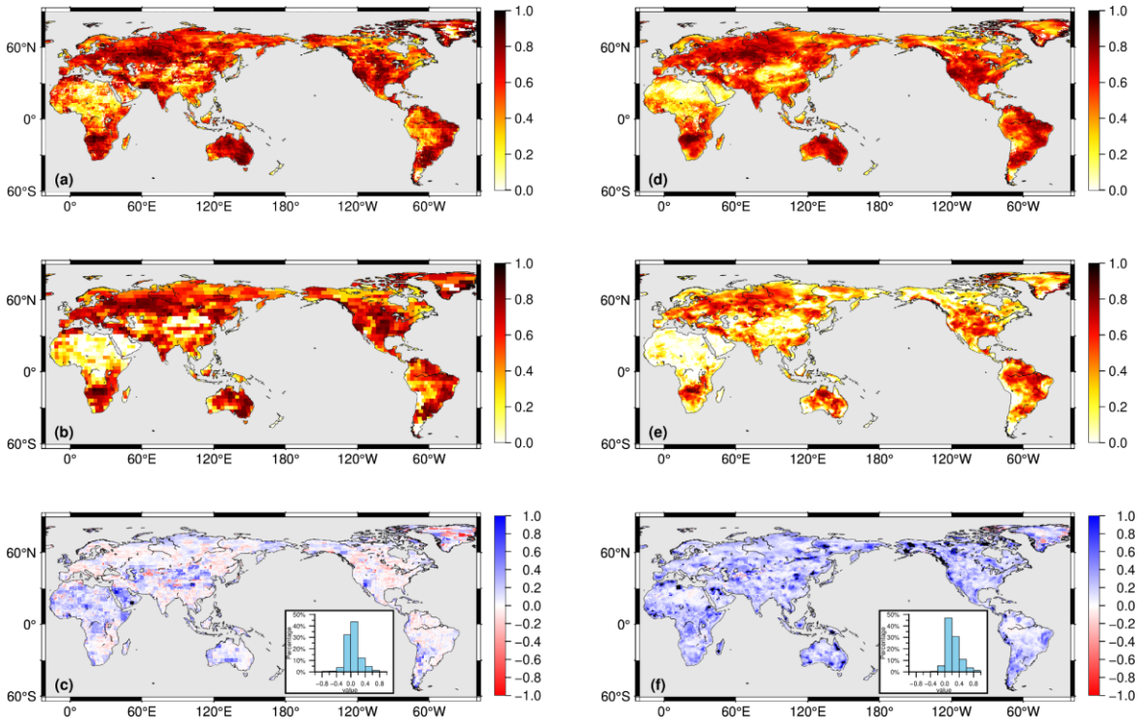


Figure 4: Spatial distribution of NSE (of de-seasonalized, de-trended anomalies) for the period from 2002 to 2019. (a) and (d): NSE of JPL-REC and GSFC-REC against their respective GRACE mascon solutions. (b) and (e): NSE of Humphrey-JPL-REC and Humphrey-GSFC-REC against the corresponding GRACE mascon solutions. (c) and (f): Spatial differences in NSE between our reconstruction and that of Humphrey and Gudmundsson (2019), both relative to the same GRACE mascon solution. Values larger than 0 indicate better performance of our method.

4.2.2 Basin-scale comparison~~Comparison with the Zhong Model~~

We compare our reconstruction (JPL-REC) with two representative types of products to systematically evaluate model performance at the basin scale: (1) a previous GRACE-based reconstruction product (Zhong-REC from (Zhong et al., 2025)), and (2) physically based land surface model simulations forced by meteorological data (GLDAS CLSM).

First, we adopted the JPLM as a common reference to evaluate the relative performance of different statistical reconstruction methods. We compared basin-averaged monthly TWSA from our JPL-REC and Zhong-REC against the JPLM and computed the NSE for each (Fig. 5). By comparing the NSE values for the two approaches, we assessed their relative reconstruction accuracy. ~~We further compared the reconstruction results at the basin scale with those from Zhong et al. (2025).~~ In the comparison across 116 major global river basins, ~~the~~ our proposed model achieved a median NSE of 0.76, outperforming Zhong-REC ~~the reconstruction by Zhong et al. (2025) (hereafter referred to as Zhong-REC)~~, which had a median NSE of 0.70. Among these basins, 84 (73%)

achieved $NSE > 0.7$, compared to only 59 (51%) for Zhong-REC. Model performance was closely linked to the quality of the forcing data and the dominant hydrological processes (Yi et al., 2023). Poor performance ($NSE < 0.5$) was observed in basins such as the Congo, Niger, and Sanaga in Africa, and the Tarim River in Central Asia, where sparse ground-based meteorological observations result in large uncertainties in climate forcing data (Xu et al., 2020; Chen et al., 2008). Cold-region basins such as the Yukon, Brahmaputra, Lena, Indigirka, Kolyma, Olenek, Yana, and Khatanga are strongly influenced by glacial and permafrost-driven seasonal accumulation and ablation processes (Riegger and Tourian, 2014; Yi et al., 2023). The linear reservoir assumption fails to adequately capture the mass variations in these basins (Liu et al., 2022), resulting in NSE values ranging from 0.45 to 0.70. In contrast, river basins where hydrological signals are dominant, such as the Amazon, Mississippi, Nelson, and Amur, showed good reconstruction performance with NSE values greater than 0.80. The histogram of ΔNSE (this model minus Zhong-REC) shown in Fig. 5c reveals a positively skewed distribution, indicating performance improvements in the majority of basins. Specifically, 37 basins (approximately one-quarter of the sample) exhibited $\Delta NSE > 0.1$. The most pronounced improvements were found in arid and semi-arid regions (Fig. ~~S2~~S4), including the Tarim River, Lake Chad, and Yellow River basins. The spatial distribution of regions with the most notable improvements in Fig. 5c closely aligns with the ΔNSE pattern between the proposed model and Humphrey-JPL-REC (Fig.~~S1~~ S5), further supporting the role of the coupling between temperature and precipitation and the joint inversion of initial conditions in enhancing reconstruction accuracy in areas with low NSE.

Second, we compared basin-averaged monthly TWSA from JPL-REC and GLDAS CLSM to highlight the improvement of our reconstruction method relative to the land surface model (Fig. S6). Across 116 major river basins worldwide, NSE values for both products are computed against the JPLM, and the spatial distribution of the resulting ΔNSE is presented (Fig. S6(c)). The results show that NSE values from our reconstruction are higher in most basins, with ΔNSE predominantly positive, indicating that the proposed method more effectively captures temporal TWSA variations at the basin scale.

It should be noted that TWSA estimates from GLDAS CLSM are derived from explicit simulations of individual storage components and are therefore influenced by the completeness and accuracy of the represented physical processes. In contrast, our approach statistically characterizes climate-driven storage responses under direct constraints from GRACE observations, implicitly integrating multiple storage processes. This leads to improved representation of total water storage variations and better agreement across many river basins.

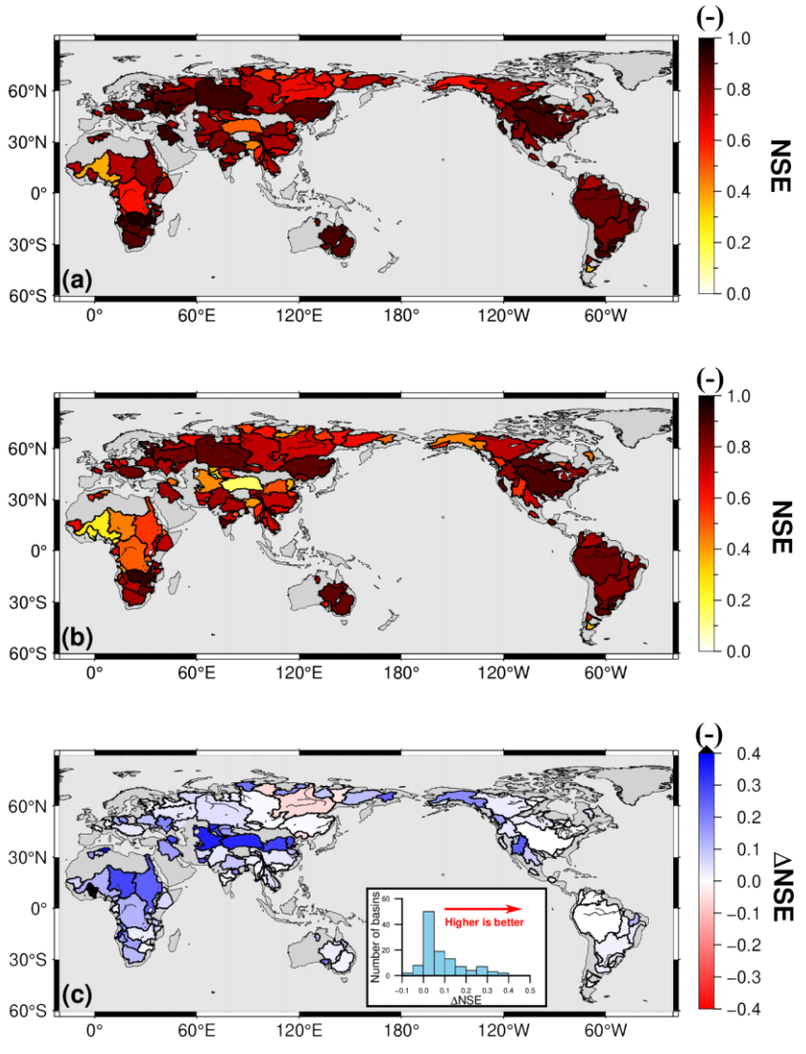


Figure 5: Spatial distribution of NSE (de-seasonalized, de-trended anomalies) between JPL mascon and two reconstruction models across 116 global river basins for the period 2002 to 2021. The NSE between JPLM and JPL-REC (a); the NSE between JPLM and the reconstruction by Zhong et al. (2025) (b); and their difference (ΔNSE =JPL-REC $-$ Zhong-REC) (c). The inset histogram in (c) illustrates the distribution of ΔNSE across all basins, where positive values indicate better agreement with GRACE observations by our reconstruction model.

4.2~~3~~.3 Comparison of Monthly Time Series

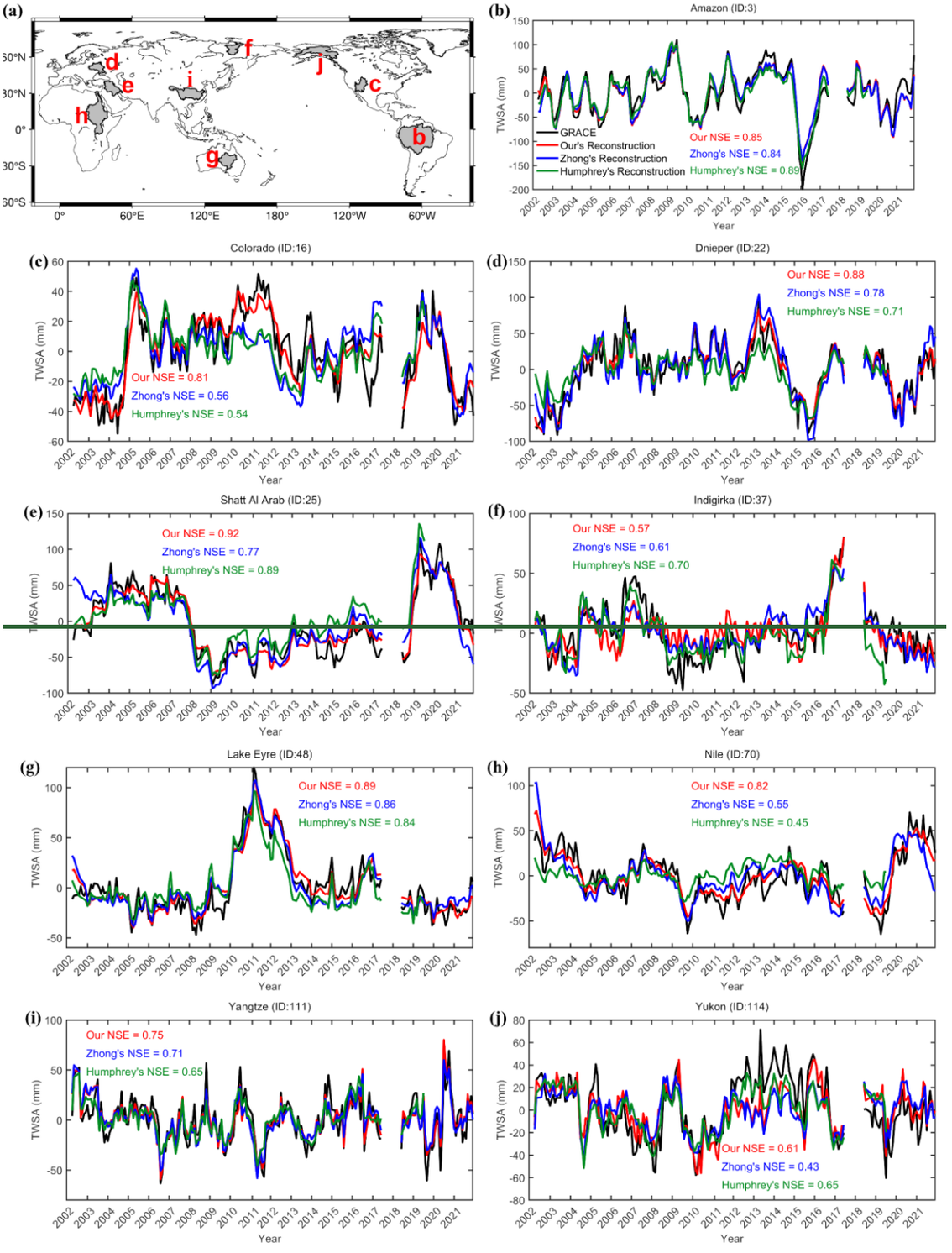
This section presents a comparison of deseasonalized and detrended monthly TWSA time series across ~~nine~~ eight representative river basins. Figure 6 illustrates the differences among GRACE/GRACE-FO observations (black line), the present reconstruction (JPL-REC, red line), Zhong-REC (blue line), and Humphrey-JPL-REC (green line) in terms of basin-averaged TWSA time series.

Interannual variations in terrestrial water storage are primarily driven by atmospheric

circulation anomalies induced by ENSO events, which often trigger hydrological extremes such as floods and droughts, thereby significantly altering regional water balance (Chen et al., 2022). In the Amazon Basin, the La Niña event of 2009 markedly increased regional precipitation (Chen et al., 2010), while the strong El Niño of 2015–2016 led to severe drought conditions (Tian et al., 2021). The proposed reconstruction model successfully captured both the water storage peak in 2009 and the pronounced depletion in 2016. In the Yangtze River Basin, the early 2010 La Niña event was one of the primary drivers of drought, as it weakened the East Asian summer monsoon and caused an eastward retreat of the western Pacific subtropical high, thereby reducing moisture transport and triggering drought in the middle and lower reaches (Li et al., 2023; Zhang et al., 2015). The sharp decline in TWSA observed by GRACE during this period was accurately reproduced by the present model. In the Nile Basin, due to the slower response of TWS compared to precipitation, the reduction in net precipitation caused by the 2008 La Niña event did not immediately lead to a decline in TWSA. Instead, TWSA exhibited a gradual downward trend from 2008 to 2010 (Forootan et al., 2019), and this phase lag was also captured by the model. During the strong La Niña of the period from 2010 to 2011, northern and inland Australia experienced abnormally wet conditions. In the Lake Eyre Basin, intense rainfall and widespread flooding caused a significant rise in TWSA (Chen et al., 2022). The reconstructed TWSA time series in this basin is highly consistent with GRACE observations.

In contrast to the basins discussed above, interannual anomalies in river basins such as the Colorado and Shatt al-Arab are closely coupled with human water extraction. In the Colorado Basin, groundwater levels experienced a brief recovery from 2009 to 2010 due to increased precipitation. However, an anomalous drought event in 2012, combined with record-low snowfall in the Rocky Mountains, led to a rapid decline in water storage (Castle et al., 2014). Similarly, since the onset of drought in 2007, the Shatt al-Arab Basin has experienced a sharp decline in precipitation, compounded by a rapid drop in upstream reservoir levels along the Euphrates River. As surface water supplies became insufficient to meet demand, groundwater abstraction surged and became the primary source for agricultural and domestic use, leading to a rapid depletion of groundwater storage (Voss et al., 2013). In the high-latitude Indigirka Basin, winter precipitation is primarily stored in solid form, with snow water equivalent (SWE) gradually accumulating throughout the frozen season and then melting intensively in the following summer. GRACE effectively captured this “snow accumulation–melt” process, showing a spring peak of approximately 50 mm from 2006 to 2007. By contrast, the four-parameter recursive model developed in this study treats all precipitation $P(t)$ as an immediate input flux, which is attenuated through a memory term $(c \cdot T_z + d) \cdot TWS(t-1)$.

555 This assumption leads to premature dissipation of the snow accumulation signal during winter,
556 resulting in significant underestimation of the seasonal peak. Meanwhile, the relatively large
557 constant memory coefficient d caused part of the residual water to be released with a delay,
558 resulting in overall higher reconstructed TWSA compared to GRACE observations from 2008 to
559 2010. Among all basins examined, this basin exhibited the poorest performance of our method
560 relative to previous models.


(a)
60°N
30°N
0°
30°S
60°S
0°
60°E
120°E
180°
120°W
60°W
d
e
i
f
j
c
h
b
g
(b)
Amazon (ID:3)
TWSA (mm)
Year
GRACE
Our's Reconstruction
Zhong's Reconstruction
Humphrey's Reconstruction
Our NSE = 0.85
Zhong's NSE = 0.84
Humphrey's NSE = 0.89
(c)
Colorado (ID:16)
Our NSE = 0.81
Zhong's NSE = 0.56
Humphrey's NSE = 0.54
(d)
Dnieper (ID:22)
Our NSE = 0.88
Zhong's NSE = 0.78
Humphrey's NSE = 0.71
(e)
Shatt Al Arab (ID:25)
Our NSE = 0.92
Zhong's NSE = 0.77
Humphrey's NSE = 0.89
(f)
Indigirka (ID:37)
Our NSE = 0.57
Zhong's NSE = 0.61
Humphrey's NSE = 0.70
(g)
Lake Eyre (ID:48)
Our NSE = 0.89
Zhong's NSE = 0.86
Humphrey's NSE = 0.84
(h)
Nile (ID:70)
Our NSE = 0.82
Zhong's NSE = 0.55
Humphrey's NSE = 0.45
(i)
Yangtze (ID:111)
Our NSE = 0.75
Zhong's NSE = 0.71
Humphrey's NSE = 0.65
(j)
Yukon (ID:114)
Our NSE = 0.61
Zhong's NSE = 0.43
Humphrey's NSE = 0.65

561

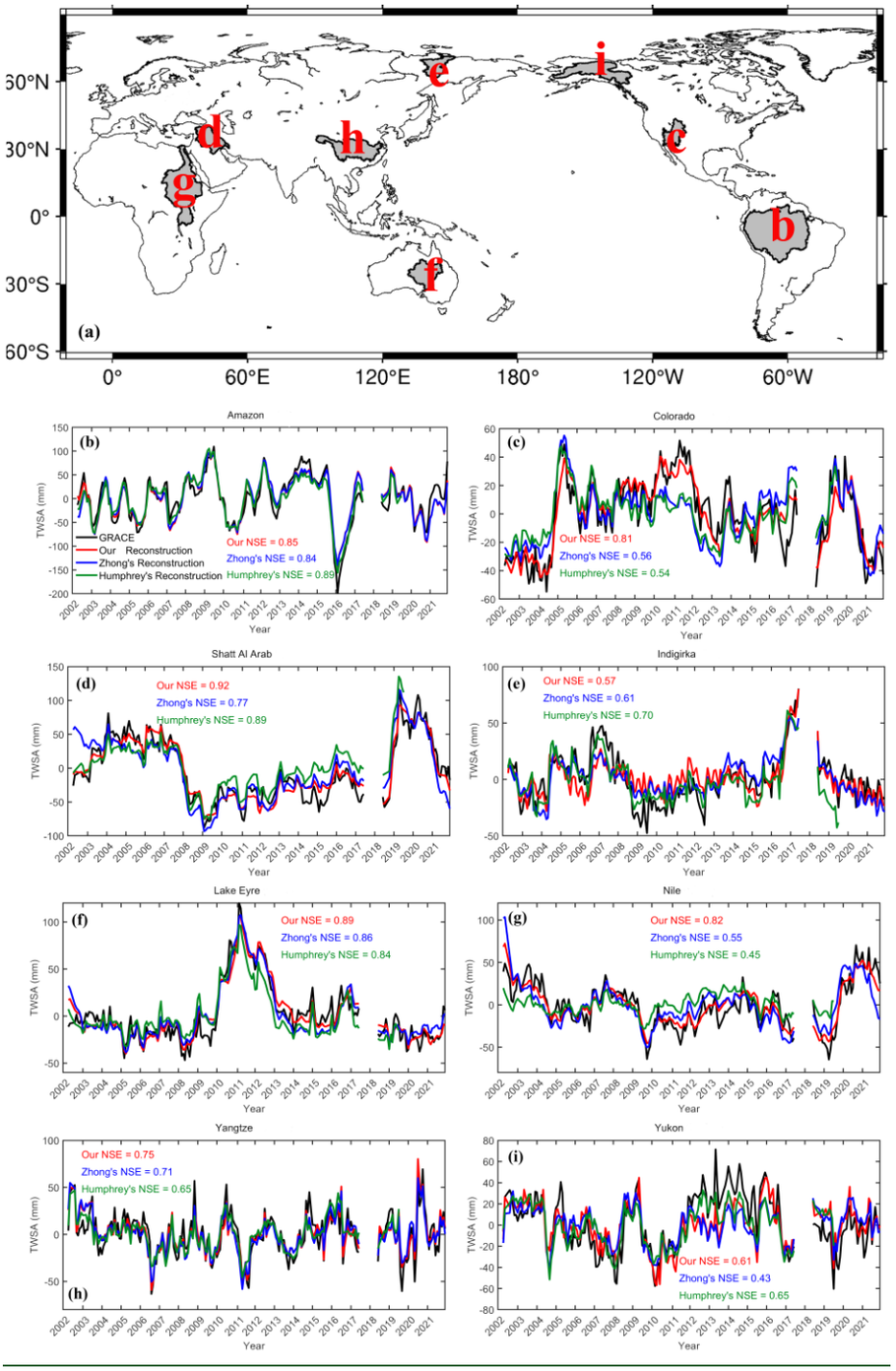


Figure 6: Time series of GRACE/GRACE-FO TWSA and reconstructed TWSA, both de-seasonalized and de-trended, for the ~~nine~~ eight selected river basins (b-j) from 2002 to 2021. The global distribution of the ~~nine~~ eight selected river basins (a).

4.2.4 Comparison of Daily Time Series

Although our model is primarily constrained by monthly GRACE observations, the method can also reconstruct daily characteristics in TWS changes based on the daily climate drivers. Here, we evaluated our reconstructed daily TWSA by comparing it with two different reference datasets,

570 namely the daily ITSG-Grace2018 solution derived using a Kalman smoothing approach and the
571 GLDAS-2.2 Daily product. The comparison with ITSG-Grace2018 evaluates whether the daily
572 reconstruction is consistent with GRACE-based sub-monthly variability, whereas the comparison
573 with GLDAS-2.2 assesses whether the daily reconstruction is consistent with physically based land-
574 surface simulations.

575 We first assessed reconstruction performance at the basin scale by computing basin-averaged
576 daily TWSA over major river basins worldwide. The difference in the time series was quantified
577 using NSE (Fig. 7a&b). The NSE values are positive across most basins, with over half
578 exceeding 0.5, indicating that our reconstruction results align well with the previous models.
579 Although the reconstruction performs well overall relative to both reference datasets, its agreement
580 with ITSG-Grace2018 is stronger in many basins. Specifically, 63 basins show NSE values greater
581 than 0.6 in comparison with ITSG-Grace2018, whereas the count slightly reduces to 51 when
582 compared with GLDAS-2.2. This is likely because our reconstruction is calibrated under constraints
583 from GRACE/GRACE-FO TWSA, while ITSG-Grace2018 is also a daily product derived from
584 GRACE observations.

585 To better illustrate sub-monthly variability, we showed daily TWSA time series in the
586 Mississippi River basin in Fig. 7c. The results show similar sub-monthly variabilities among the
587 reconstructed GRACE-REC, ITSG-Grace2018, and GLDAS-2.2 time series. One exception is found
588 in the Indus basin between our GRACE-REC and GLDAS-2.2. The main reason is that GLDAS-2.2
589 fails to capture a three-year drought spanning from 2016 and 2018 (see Fig. S7 for details). The
590 reconstructed daily TWSA is consistent with the corresponding reduction in the 6-month moving
591 average of precipitation. By contrast, the GLDAS-2.2 daily TWSA remains comparatively flat
592 during this period, suggesting that it may underestimate the sensitivity of basin-scale storage to
593 interannual hydroclimatic forcing in this basin. Figure S7(b) further shows that our monthly
594 reconstruction remains close to the GRACE/GRACE-FO observations, particularly around the
595 resumption of GRACE-FO observations in June 2018, whereas GLDAS-2.2 tends to overestimate
596 TWSA at that time. This discrepancy highlights the indispensable role of observation-based
597 reconstruction approaches.

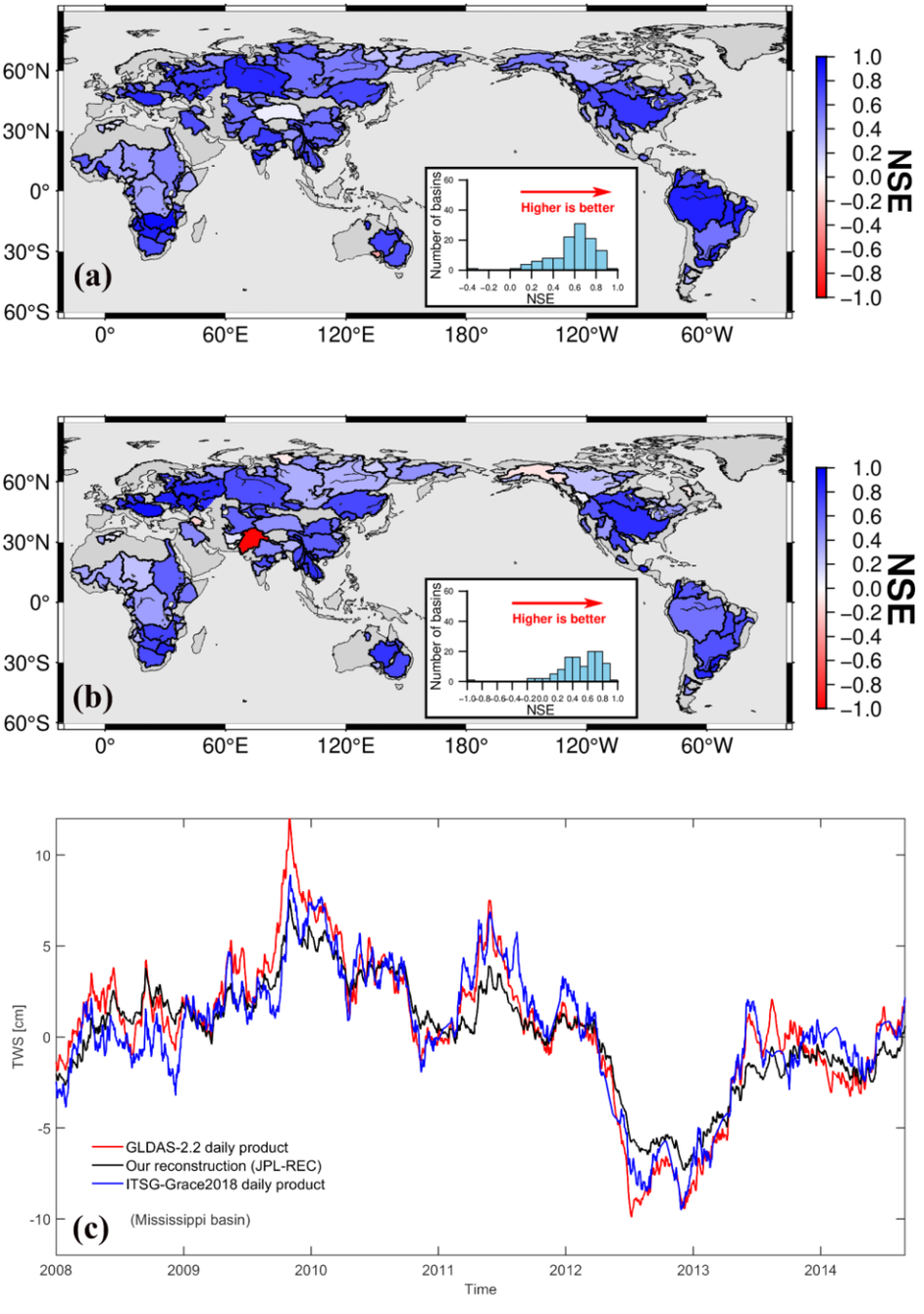


Figure 7: Spatial distribution of NSE (of de-seasonalized and de-trended TWSA) between our daily reconstruction and (a) ITSG-Grace2018 for the period from January 2003 to August 2016 and (b) the GLDAS-2.2 daily product for the period from January 2003 to December 2021 across 116 major river basins worldwide. (c) Time-series comparison of the three daily datasets for the Mississippi River basin, focusing on the period 2008–2014 to improve the readability of high-frequency fluctuations.

4.2.5 Independent validation of reconstruction performance

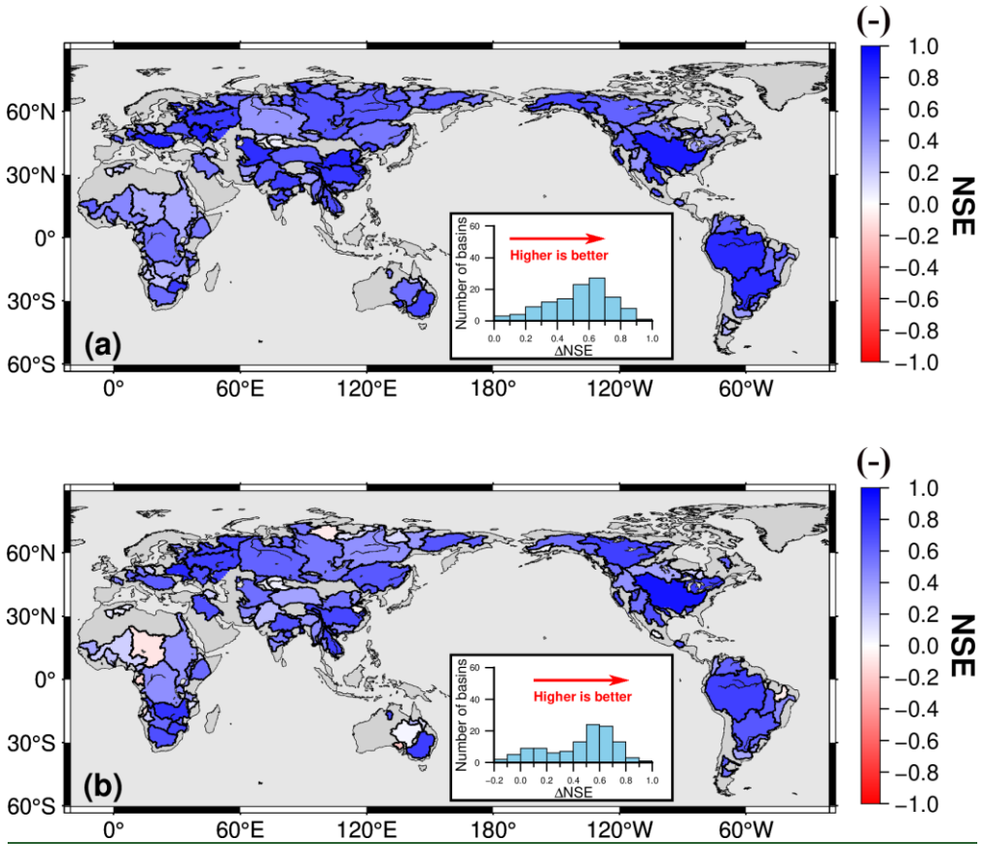
We conducted a self-validation experiment to further evaluate the robustness of the proposed method by dividing the JPLM product into separate training and validation periods. Specifically, the

period from April 2002 to December 2012 was designated as the training period and the model parameters were estimated, while the period from January 2013 to December 2023 was treated as an independent validation period.

We evaluated the model performance using the NSE by comparing JPL-REC with the JPLM in both the training and validation periods (Fig. 8). During the training period (Fig. 8a), most basins exhibit high reconstruction skill, with NSE values exceeding 0.6 over large parts of North and South America as well as Eurasia. A very similar spatial pattern is obtained for the validation period (Fig. 8b), with only a modest reduction in overall skill. Approximately 60 % of the basins retain NSE values above 0.5 in the validation period. This indicates that the model retains stable predictive capability beyond the calibration window and does not exhibit significant overfitting behavior.

The spatial pattern of the results during validation period reveals distinct regional characteristics. High NSE values are mainly found in large, humid basins with relatively dense observations, such as the Amazon, Mississippi, Yangtze River basins. In these regions, variability in TWS is strongly controlled by climate forcing. Consequently, our model can reliably reproduce the TWSA signals observed by GRACE/GRACE-FO.

We further selected eight representative basins spanning high-latitude cold regions, tropical humid regions and typical monsoon-affected areas (Fig. 9). For each basin, the validation period TWSA (blue line) was compared with the reconstructed values during the training period (red line) and GRACE observations (black line). The results show that in certain basins, such as the Amazon, Yangtze, Mississippi and Murray, model performance during the extrapolation period remained comparable to or even exceeded that during the training period. This indicates that under these hydroclimatic conditions, the fixed-parameter model can reliably capture interannual TWSA variability driven by precipitation and temperature. In contrast, basins such as Chad and Khatanga show lower NSE during the validation period and larger discrepancies in certain years. This decline may be attributed to the limitations imposed by parameter nonstationarity and errors in the forcing data. These results suggest that when only precipitation and temperature forecasts are available, the proposed recursive model can be directly applied for estimating TWS changes during missing periods or into the future. Nevertheless, its applicability varies with hydroclimatic conditions, and for regions requiring high prediction accuracy, periodic parameter updates or the inclusion of additional forcing information may still be necessary.


(a)
(b)
60°N
30°N
0°
30°S
60°S
0°
60°E
120°E
180°
120°W
60°W
(-)
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
NSE
Higher is better
Number of basins
ΔNSE

637

Figure 8: Comparison of reconstruction performance against JPLM data during the training and validation periods. (a) Spatial distribution of the NSE between JPL-REC and JPLM data during the training period (April 2002–December 2012); (b) Spatial distribution of NSE during the validation period (January 2013 December 2023)

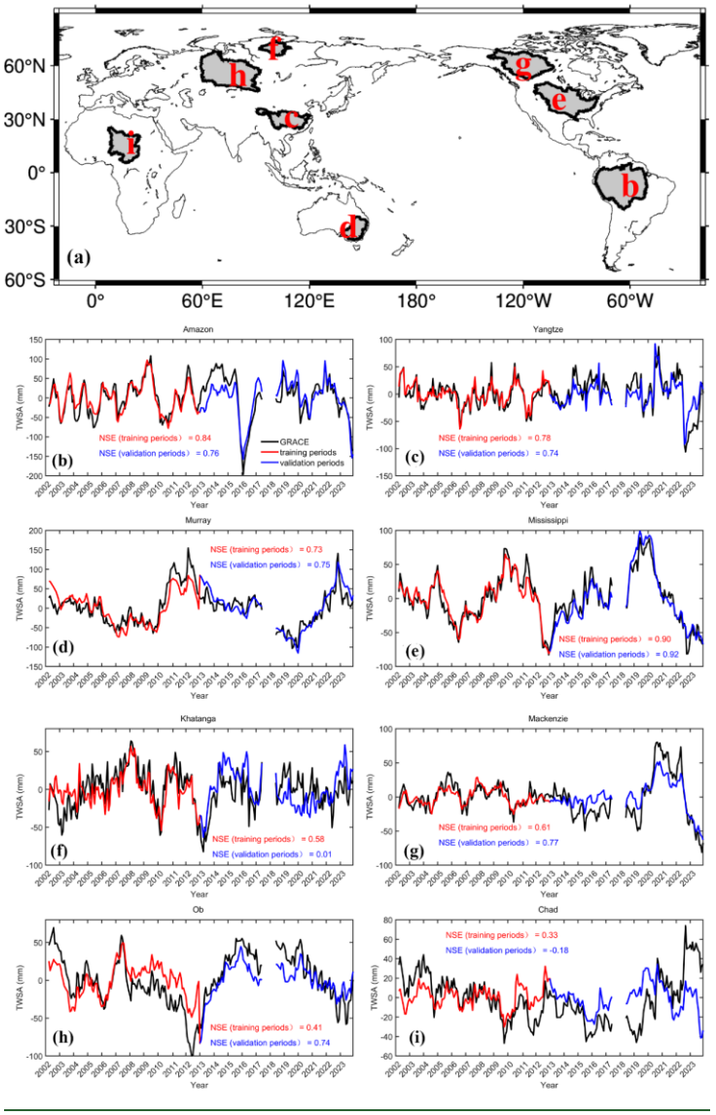


Figure 9: Time series comparison of JPL-REC and JPLM, both de-seasonalized and de-trended, for the eight selected river basins (b-i) during training and validation periods. Black lines represent GRACE TWSA, red lines indicate model reconstructions during the training period, and blue lines during the validation period. The global distribution of the eight selected river basins (a).

5 Discussions

5.1 Sensitivity Analysis of the Model

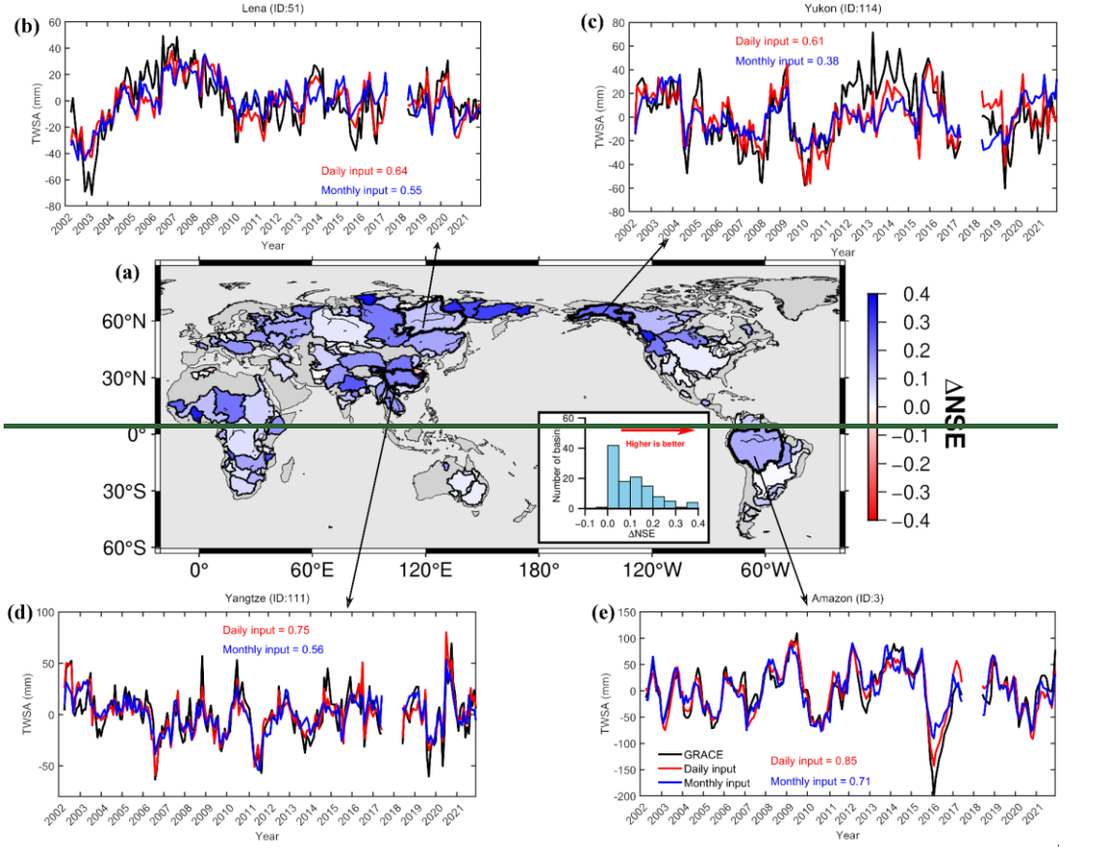
5.1.1 Sensitivity to Temporal Resolution

We first evaluated the impact of the temporal resolution of meteorological forcing data on reconstruction performance, specifically whether daily forcing provides additional benefit compared to monthly forcing. To this end, all 116 basins were recalibrated using both daily and monthly ERA5-Land precipitation and temperature. The monthly forcing datasets were aggregated from daily values: precipitation was summed to monthly totals, and temperature was averaged over each month. The deseasonalized and detrended TWSA reconstructions were then compared on a monthly basis, and NSE was computed for each case. The difference between the two forcing schemes was quantified as $\Delta NSE = NSE_{Daily} - NSE_{Monthly}$, with its spatial distribution shown in Fig. ~~7a~~8a. Positive values indicate better performance under daily forcing.

The results show that 108 out of 116 basins (approximately 95%) exhibit positive ΔNSE values, suggesting that daily forcing generally outperforms monthly forcing. The most significant improvements were observed in high-latitude permafrost–glacier transition zones such as Siberia and Alaska. In contrast, tropical humid basins (e.g., the Paraná Basin) exhibited minimal differences ($\Delta NSE \leq \pm 0.05$). Only seven basins (approximately 5%), such as the Huai River Basin, showed slightly negative differences ($-0.1 < \Delta NSE < 0$), mostly located in regions with strong human regulation or sparse observational data.

From a time series perspective, the reconstructed TWSA driven by monthly climate forcing exhibits significantly lower amplitude compared to that driven by daily forcing. This is mainly due to two factors: daily forcing preserves extreme precipitation events and high-frequency temperature fluctuations on a day-to-day basis, allowing short-term water flux pulses to enter the recursive system as sharp peaks. In contrast, monthly forcing involves pre-integrating or averaging over 30-day precipitation and temperature records before inputting them into the model, which smooths out extreme events and leads to underestimated amplitude in the simulations (Humphrey et al., 2016). In addition, the memory coefficient calibrated at the daily time step typically ranges from $\mathrm{d_{day}} \approx 0.97$ to 0.995, indicating very limited attenuation on a daily basis. When directly scaled to a monthly time step, this corresponds to $\mathrm{d}_{day}^{30} \approx 0.74$; alternatively, recalibration at the monthly scale yields an equivalent memory coefficient of approximately $\mathrm{d}_{day}^{30} \approx 0.7$ to 0.8. Both cases imply a stronger low-pass filtering effect, which further dampens peak values and elevates troughs.

Therefore, we argue that daily forcing is essential for improving model fidelity across nearly all basins and should be considered necessary for accurate reconstruction.



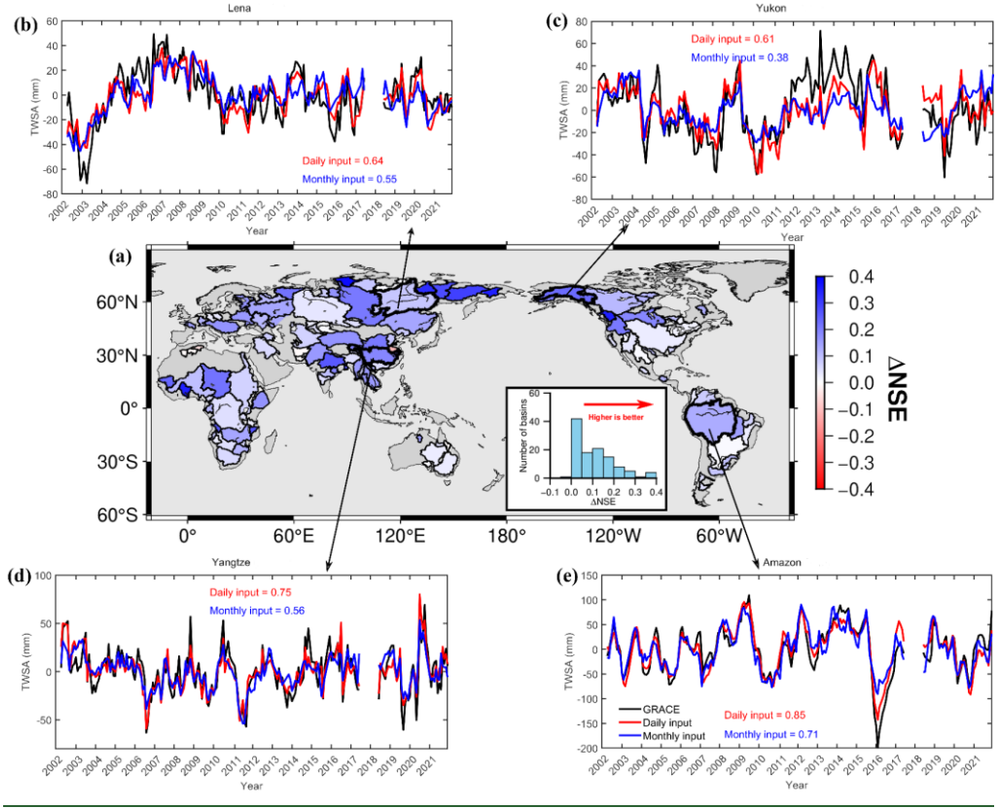


Figure ~~7~~8: Spatial distribution of the difference in NSE between TWSA reconstructions forced by daily and monthly meteorological data, with respect to GRACE JPLM from 2002 to 2021 (a). Time series of GRACE/GRACE-FO TWSA and reconstructed TWSA (de-seasonalized and de-trended) using daily and monthly meteorological data in the (b) Lena, (c) Yukon, (d) Yangtze, and (e) Amazon basins.

5.1.2 Sensitivity to Initial Conditions

This section investigates the influence of the initial terrestrial water storage, $TWS(0)$, on reconstruction performance. Humphrey and Gudmundsson (2019) demonstrated through experiments that if $TWS(0)$ is simply set to zero, the model requires a spin-up period before reaching equilibrium. To avoid discarding multiple years of data, they derived an analytical equilibrium value to serve as a uniform initialization for $TWS(0)$. However, when a basin is not in climatic equilibrium, this equilibrium-based initialization can introduce systematic bias. Since the model's memory decay coefficient $(c \cdot T_z + d)$ is approximately equal to 1, such bias can persist for several years in basins with long water residence times, making the initial condition a critical factor influencing reconstruction accuracy in long-memory basins. To mitigate this bias, $TWS(0)$ in this study is jointly estimated along with the model parameters (a, b, c, d). Three initialization strategies were compared (Fig. ~~8~~9): (1) the optimal solution where $TWS(0)$ is treated as an unknown and solved through least squares estimation; (2) a suboptimal solution, obtained by increasing the optimal value by 10%; and (3) the equilibrium solution, derived from Eq. (5). Using the Chad Basin as an example, the NSE of the equilibrium initialization was 0.48, which increased to 0.57 under the suboptimal case and further improved to 0.73 under the globally optimal initialization. Therefore, in this study, $TWS(0)$ is retrieved through parameter inversion for all basins, substantially reducing systematic bias and enhancing reconstruction accuracy.

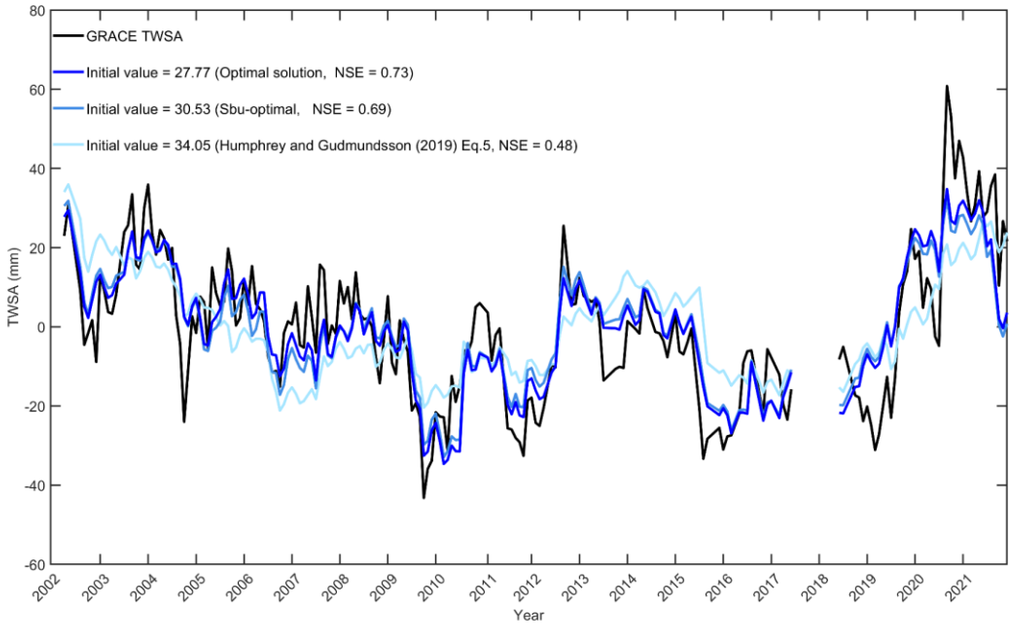


Figure ~~8~~9: Time series of TWSA in the Chad Basin from GRACE observations and model reconstructions under three different initial condition strategies. GRACE data (black) are based on the JPLM; model reconstructions are obtained using (1) an optimized initialization (dark blue), (2) a sub-optimal initialization (medium blue), and (3) an equilibrium initialization following Eq. (5) of Humphrey and Gudmundsson (2019) (light blue).

5.2 Discussion of Parameter Sensitivity

5.2.1 Parameter a

Figure ~~9~~ 10 illustrates the global distribution of parameter a and its associated uncertainty. ~~The spatial pattern of parameter a exhibits~~That parameter a arises from the first-order Taylor expansion of the temperature modulation term in the model formulation (Eq. 13). Therefore, it represents a mathematical sensitivity coefficient describing how temperature perturbs the relationship between precipitation input and TWS response. As shown in Fig. 10a, the sign of parameter a varies among basins, with positive values in some basins and negative values in others. In many high-latitude regions, the calibrated values tend to be negative, whereas some humid basins exhibit positive values. These spatial differences likely reflect variations in regional hydroclimatic conditions and the ways in which temperature influences hydrological processes such as snowmelt, evapotranspiration, and soil moisture dynamics. ~~pronounced contrasts across different climatic zones. In high-latitude cold regions, liquid runoff is nearly absent during winter, and when temperatures rise above 0 °C in summer, rapid snowmelt occurs. However, the presence of permafrost and surface ice layers greatly reduces infiltration capacity, causing meltwater to be discharged primarily as surface runoff rather than being retained in terrestrial water storage. As a~~

~~result, the inversion yields $a < 0$, indicating that warming diminishes the efficiency of precipitation conversion into TWS.~~

~~In low-latitude arid and semi-arid regions, the annual mean ET/P ratio is high (Fig. S5), and evapotranspiration plays a dominant role in the water balance. As temperature increases, potential evapotranspiration intensifies, leading to greater atmospheric loss of precipitation before infiltration, which similarly results in $a < 0$. In contrast, the Yangtze River Basin exhibits a positive response ($a > 0$), as warming coincides with the wet season and ET/P remains relatively low (Fig. S5). Under these conditions, soil and vegetation can effectively retain and infiltrate rainfall, thereby enhancing the conversion of precipitation into terrestrial water storage. Similarly, in the Yellow River Basin, large-scale ecological restoration projects in recent years have substantially improved vegetation cover, soil structure, and infiltration conditions, resulting in significant increases in both surface and groundwater storage (Feng et al., 2025). Consequently, when temperatures rise and evapotranspiration intensifies, the enhanced soil water retention and groundwater recharge are sufficient to offset the additional evaporative losses. As a result, the model likewise yields a positive response with $a > 0$.~~

Figure ~~10a~~ 11a compares the full four-parameter model with a reduced three-parameter version in which the temperature and precipitation coupling term (i.e., parameter a) is excluded. The global distribution of ΔNSE indicates that, on average, the contribution of parameter a is limited. However, under specific hydroclimatic conditions, incorporating parameter a can lead to improvements in model performance with $\Delta NSE > 0.1$. Notable improvements are concentrated in high-latitude snow-dominated basins, such as the Olenek River, and in certain arid regions with high evapotranspiration, such as the Volta Basin. For the remaining 112 basins, ΔNSE values are below 0.1, suggesting no statistically significant difference. From a time series perspective, excluding parameter a in the Volta Basin leads to an overestimation of summer peaks during 2005–2006 and an underestimation during 2009–2012. In regions with high temperatures, removing parameter a impairs the model's ability to capture interannual peak and trough variability.

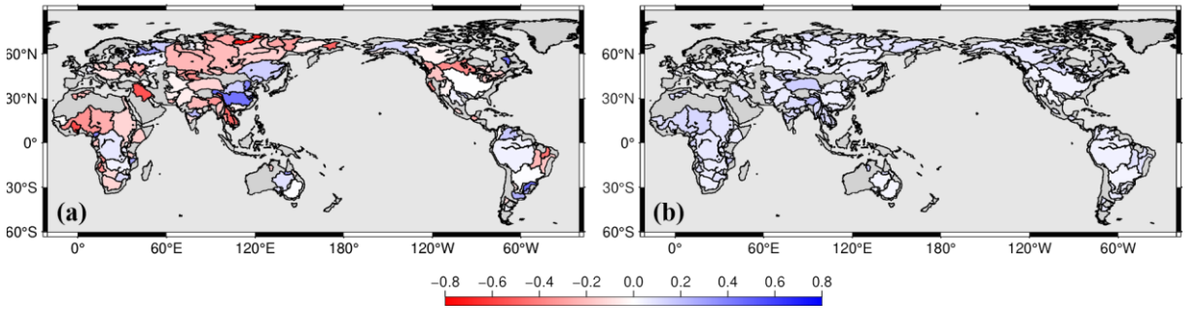
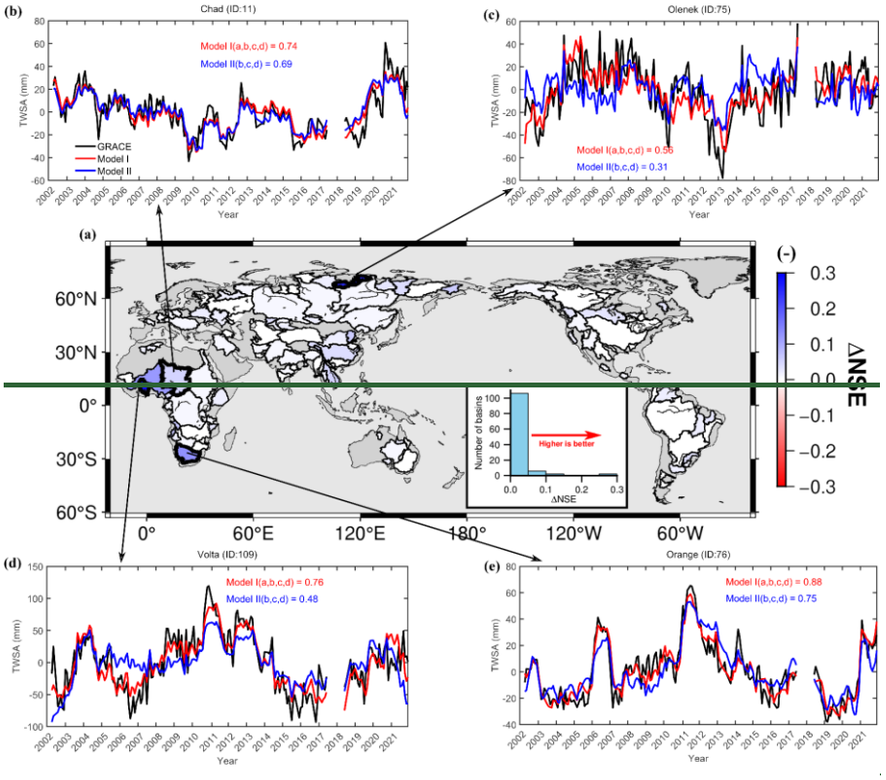


Figure ~~9~~10: Spatial distribution of parameter a (a), and parameter a errors (b) across global 116 river basins. Parameters calibrated by monthly TWSA from JPLM based on ERA5-Land precipitation~~P~~ and temperature from 2002 to 2021. Uncertainty estimation of parameters based on Monte Carlo simulations using precipitation and temperature errors across each basin. Precipitation errors were derived from the standard deviation between ERA5-Land and MSWEP datasets. Temperature errors were derived from the standard deviation between ERA5-Land and GLDAS-2.2 ~~GLSM~~ datasets.



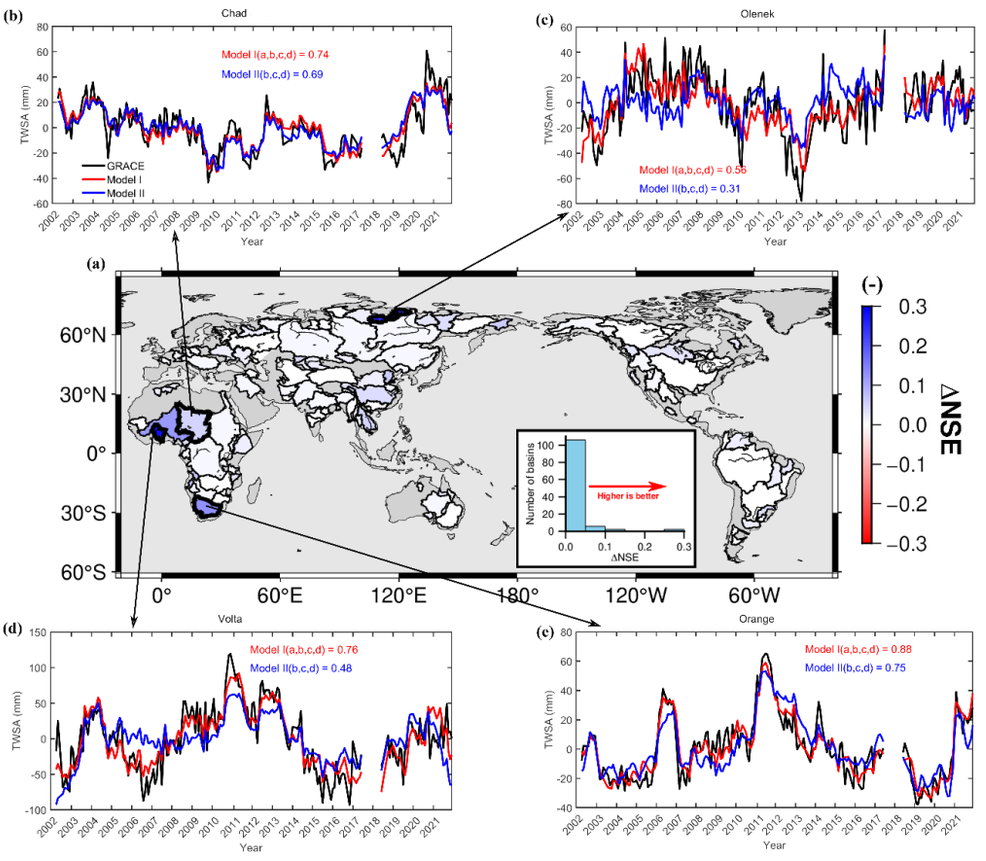


Figure ~~10~~11: Spatial distribution of the NSE difference between TWSA reconstructions from a four-parameter and a three-parameter (excluding parameter a) daily recursive model with respect to JPLM across 116 global river basins during 2002–2021 (a). Time series of GRACE/GRACE-FO TWSA and reconstructed TWSA (de-seasonalized and de-trended) in the (b) Chad, (c) Olenek, (d) Volta, and (e) Orange basins.

5.2.2 Parameter b

From Eq. (17) it follows that parameter b represents the fraction of precipitation that is ultimately converted into TWS: a larger b value indicates that a higher fraction of precipitation contributes to storage, whereas a smaller b suggests that a greater proportion of precipitation is dissipated through evapotranspiration or rapid runoff before becoming part of the basin water storage. In this sense, parameter b characterizes whether precipitation tends to be directly lost from the system or effectively retained as storage.

~~Figure 11~~ Figure 12 shows the global distribution of parameter b and its associated uncertainty. Overall, parameter b exhibits spatial heterogeneity across different regions. As shown in a Fig. 12, relatively low parameter b values are found in the Congo, Niger, and Orange River basins in Africa; the Lake Eyre and Murray basins in Australia; and the Rio Grande and Colorado basins in North America. These regions are characterized by relatively high long-term mean ET/P (Fig. S8); a large

fraction of precipitation is exhausted by evaporation and runoff before contributing to TWS, which yields a low conversion efficiency. Consequently, the efficiency of precipitation-to-storage conversion is low, resulting in smaller calibrated b values.

In contrast, higher b values are observed in high-latitude basins such as the Lena and Yenisey. In these regions, winter precipitation is predominantly stored as snow or ice, delaying immediate losses to the atmosphere or river discharge. During the snowmelt season, the accumulated solid water is released and contributes to TWS. This storage–delay–release mechanism reduces the fraction of direct precipitation loss and enhances the effective conversion of precipitation into TWS, thereby leading to higher b values.

To further substantiate this physical interpretation, we examined the statistical relationship between parameter b and the multi-year mean loss ratio $(ET + R)/P$ at the basin scale (Fig. S9). The results reveal a negative correlation between parameter b and $(ET + R)/P$. This statistical relationship provides independent observational support for interpreting parameter b. Furthermore, the global spatial pattern of b is consistent with the findings of (Zhong et al., 2025).

~~As illustrated, the spatial pattern of b is consistent with the findings of Zhong et al. (2025): high-latitude regions, where winter snow accounts for a substantial portion of total precipitation, require larger values of parameter b to compensate for the temporary storage effect of snow accumulation. In arid and semi-arid basins across the mid- to low latitudes, ET/P ratios are generally high (Fig. S5). As evapotranspiration dominates the depletion of shallow soil moisture, a substantial portion of precipitation is lost to the atmosphere before it can infiltrate or recharge groundwater, resulting in relatively low values of parameter b. In contrast, humid basins at similar latitudes exhibit lower ET/P ratios, where only a small fraction of precipitation is consumed by evapotranspiration. Consequently, a greater share of precipitation contributes to terrestrial water storage, leading to higher b values.~~

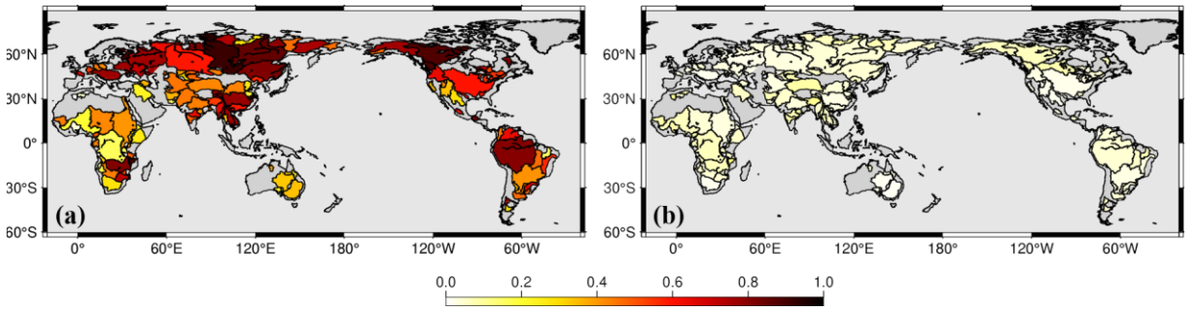


Figure ~~11~~12: Spatial distribution of parameter b (a), and parameter b errors (b) across global 116 river basins. Parameters calibrated by monthly TWSA from JPLM based on ERA5-Land precipitation~~P~~ and temperature from 2002 to 2021. Uncertainty estimation of parameters based on Monte Carlo simulations using precipitation and temperature errors across each basin. Precipitation errors were derived from the standard deviation between ERA5-Land and MSWEP datasets. Temperature errors were derived from the standard deviation between ERA5-Land and GLDAS-2.2 ~~GLSM~~ datasets.

5.2.~~2~~ 3 Parameter c

~~Figure 12~~ Figure 13 shows the global distribution of parameter c and its associated uncertainty. The parameter c arises from the first-order Taylor expansion of the temperature modulation term in the model formulation (Eq. 15). Therefore, it represents a mathematical sensitivity coefficient describing how temperature perturbs the storage memory term. The parameter c exhibit clear spatial variability. In many high-latitude basins, the magnitude of c is relatively larger (mostly negative), whereas in most mid- and low-latitude basins the values are close to zero. This pattern indicates that the influence of temperature on the model's water storage memory term is more pronounced in colder regions, while in many other basins the recursive water storage component is largely temperature-independent within our reconstruction model.~~In high-latitude basins, parameter$c$is predominantly negative, indicating that rising temperatures substantially reduce the contemporaneous water retention rate. This pattern aligns with typical high-latitude hydrological processes: winter precipitation is stored in solid form and rapidly melts during spring warming, resulting in surface runoff and a sharp decline in TWS. In contrast, low-latitude regions lack large-scale snow accumulation and melt processes, resulting in weaker coupling between temperature and the memory term. As a result, parameter c values in these regions are typically close to zero.~~ The spatial distribution of parameter uncertainty (Fig.~~12b~~ 13b) further shows that estimation errors are relatively small in most regions, suggesting that the adopted calibration approach is stable and robust across the majority of basins.

~~Figure 13~~ Figure 14 compares the full four-parameter model with a reduced three-parameter version excluding parameter$c$. Among the 116 basins analyzed, most exhibit limited sensitivity to parameter c; however, incorporating c still yields measurable performance improvements in some cases, with 11 basins showing $\Delta NSE > 0.1$. Notable improvements are concentrated in high-latitude cold regions around the Arctic (e.g., Yana, Yukon, and Lena), while limited improvements are found in parts of North America (e.g., Columbia, Colorado, and Rio Grande) and in certain Chinese basins (e.g., Yangtze and Yellow River).

From a time series perspective, in the Yukon Basin, the average temperatures during the

snowmelt seasons of the period from 2006 to 2007 and in 2009 were markedly below normal (Fig. S10~~3~~). Under these conditions, the decay factor $(c \cdot T_z + d)$ remained close to the baseline value d, resulting in a slower dissipation rate. Meanwhile, SWE during these years was also below the multi-year average, limiting the contribution from snowmelt. GRACE observations captured a slight increase in TWS during spring, followed by a continuous decline. The four-parameter model successfully reproduced this slow depletion process (Fig.~~13c~~ 14c). In contrast, the three-parameter model, due to the absence of a temperature modulation term, maintains a higher retention rate under the same precipitation input and consequently produces an overall overestimation in the reconstructed TWS curve. Conversely, 2019 was an exceptionally warm year, with precipitation anomalies remaining positive from 2017 to 2021. Rising temperatures substantially reduce$(c \cdot T_z + d)$, thereby accelerating the depletion of antecedent water storage. The abundant snowpack was rapidly flushed during the melt season, resulting in a deeper TWS trough recorded by GRACE compared to that of 2006 to 2007. The four-parameter model captured this deep trough due to the dynamic weakening effect of the parameter c, whereas the three-parameter model produced a shallower minimum. A similar pattern is observed in the Yana Basin: exceptionally high SWE in 2007 led to elevated TWS levels, while reduced snowmelt and elevated temperatures in the following year (Fig. ~~S4~~S11) triggered rapid snowmelt and a sharp decline in TWS during 2008. The four-parameter model exhibits better agreement with GRACE observations, whereas the three-parameter model shows excessive smoothing (Fig.~~13b~~ 14b). These comparisons highlight the critical role of the temperature modulation term c in accurately capturing TWS extremes during warm years or periods of rapid snowmelt.

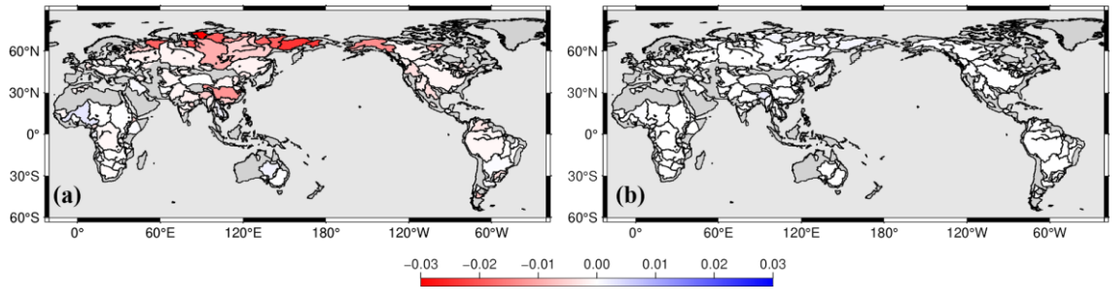
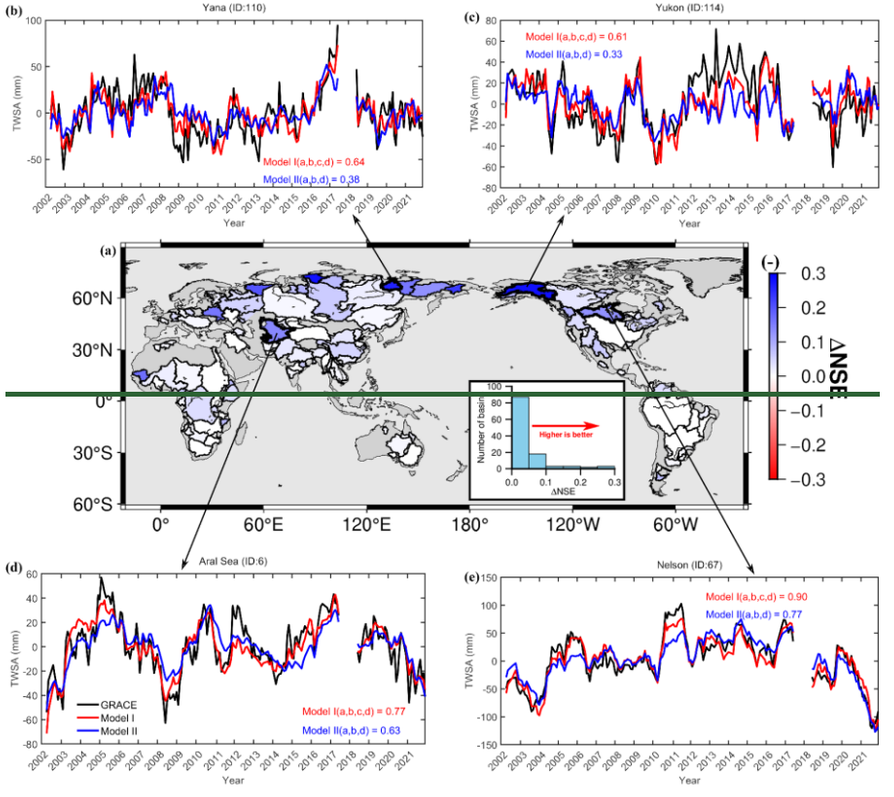
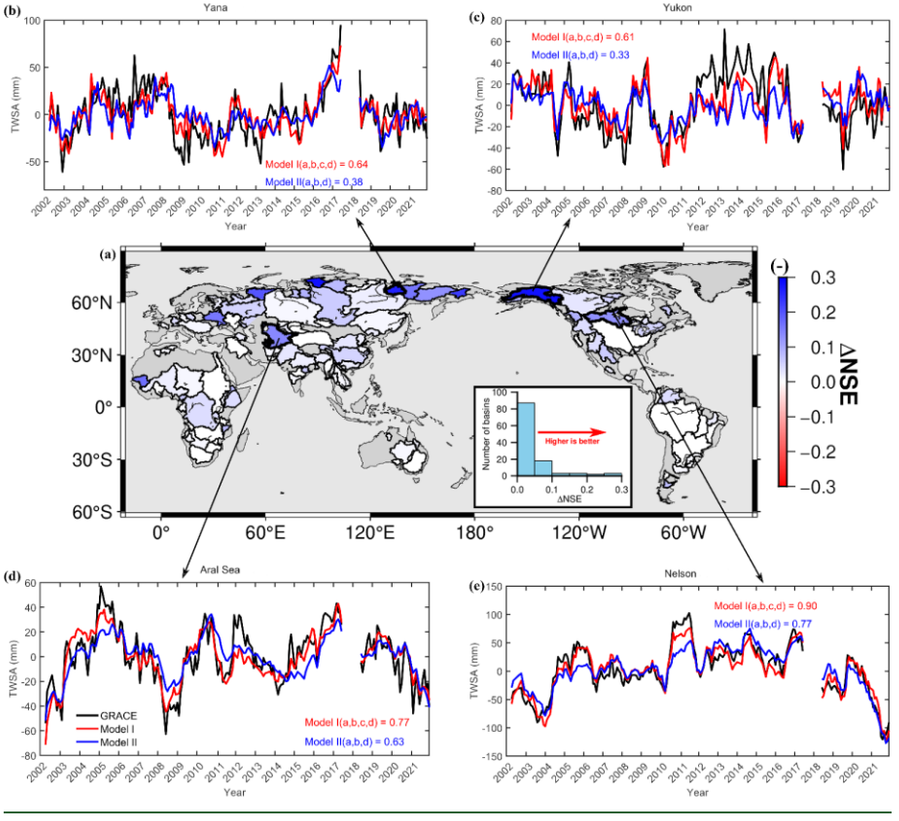


Figure ~~12~~13: Spatial distribution of parameter c (a), and parameter c uncertainties (b) across global 116 river basins.



858



859

Figure ~~13~~14: Spatial distribution of the NSE difference between TWSA reconstructions from a four-parameter and a three-parameter (excluding parameter c) daily recursive model with respect to JPLM across 116 global river basins during 2002-2021 (a). Time series of GRACE/GRACE-FO TWSA and reconstructed TWSA (de-seasonalized and de-trended) in the (b) Yana, (c) Yukon, (d) Aral Sea, and (e) Nelson basins.

5.2.~~3~~ 4 Parameter d

In the linear storage model, when no additional precipitation input is introduced, water depletion follows an exponential decay process. As shown in Fig.~~14~~ 15, when parameter~~the baseline memory coefficient~~ $d = 0.97$, and in the absence of both additional precipitation input and temperature effects, only 40% of the stored water remains after one month. In contrast, when $d = 0.99$, more than 70% of the stored water can still be retained after one month under the same condition. This indicates that the interannual variability of TWS is highly sensitive to the value of parameter d.

To examine spatial variability, Fig. 16 presents ~~Figure 15 presents~~ the global distribution of the 30-day retention factor (d^{30}) and its associated uncertainty (Fig. 16). According to the model formulation, parameter d in Eq. (17) represents the fraction of antecedent storage that remains in the system after a period of decay. From Eq. (10) it follows that, when d approaches 1, the storage at the previous time step is almost entirely preserved, implying that the current total loss $(ET + R)$ is primarily attributable to contemporaneous precipitation losses. Conversely, when d deviates from 1, current losses are also driven by the release of antecedent storage. This mechanism is consistent with the spatial pattern shown in Fig. 16a. In high-latitude regions of the Northern Hemisphere, the widespread presence of permafrost limits the infiltration of meltwater during the spring thaw, causing most of the water to be rapidly discharged as surface runoff. As a result, the calibrated parameter d~~baseline memory coefficient~~ is relatively low.~~, typically ranging from 0.98 to 0.99. Although this value implies a faster daily decay rate, persistent snowfall during winter months largely compensates for water loss. When temperatures rise in spring, the negative value of parameter c quickly reduces the instantaneous retention rate, allowing the model to effectively capture the rapid recession of water storage observed by GRACE.~~

Notably, the Ob and Mackenzie River basins exhibit considerably higher water retention after 30 days of decay compared to other high-latitude basins (Fig. ~~15a~~16a). In the Ob basin, approximately 15% of the upstream and midstream region between the Irtysh River and the upper Ob is occupied by endorheic depressions, where inflow is retained and does not contribute to the main channel (Yi et al., 2023). Additionally, springtime ice jams, channel backwater, and overbank flooding further prolong the average residence time of water, resulting in a higher calibrated 30-day

retention coefficient (d^{30}). In the Mackenzie basin, the presence of multiple large lakes leads to a highly fragmented river network (Yi et al., 2023), impeding the continuous drainage of liquid water and thereby extending the residence time of water within the basin.

~~In contrast, tropical and subtropical basins are characterized by deep groundwater reservoirs and long residence times, with no seasonal clearing effect from freeze and thaw processes. As a result, the calibration favors higher values of parameter d (typically 0.995 to 0.998) to reproduce the slow depletion of water storage over several months.~~

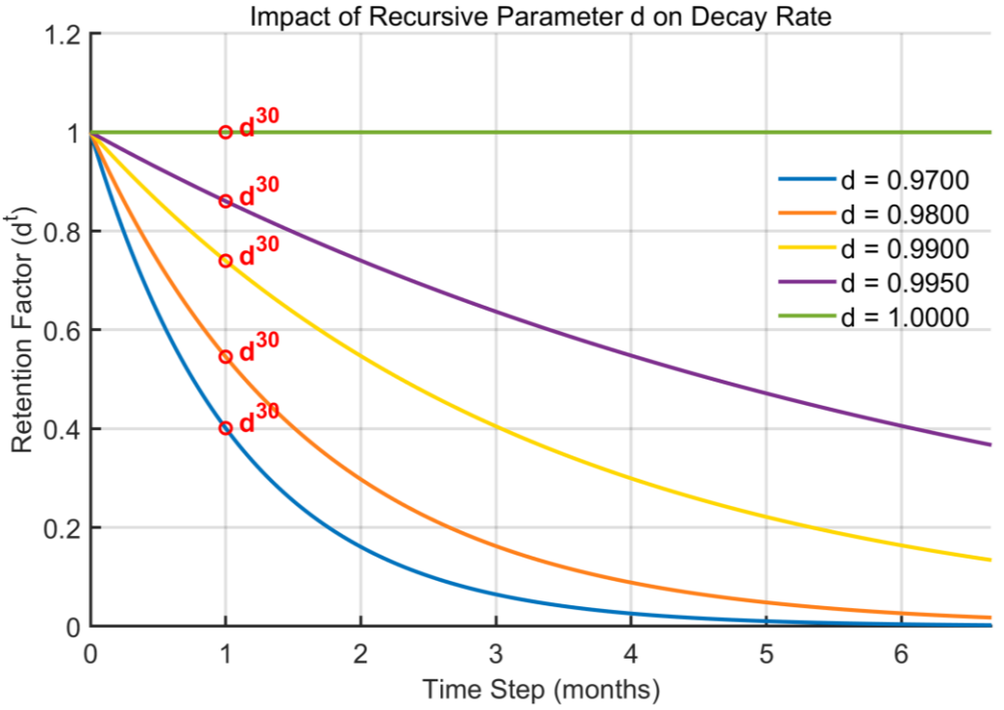


Figure ~~14~~15: Comparison of the temporal evolution of the decay factor under different values of parameter d (t in days).

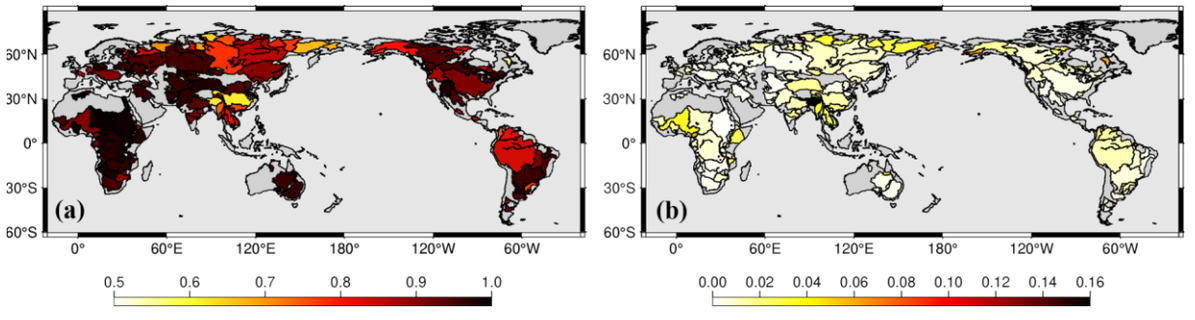
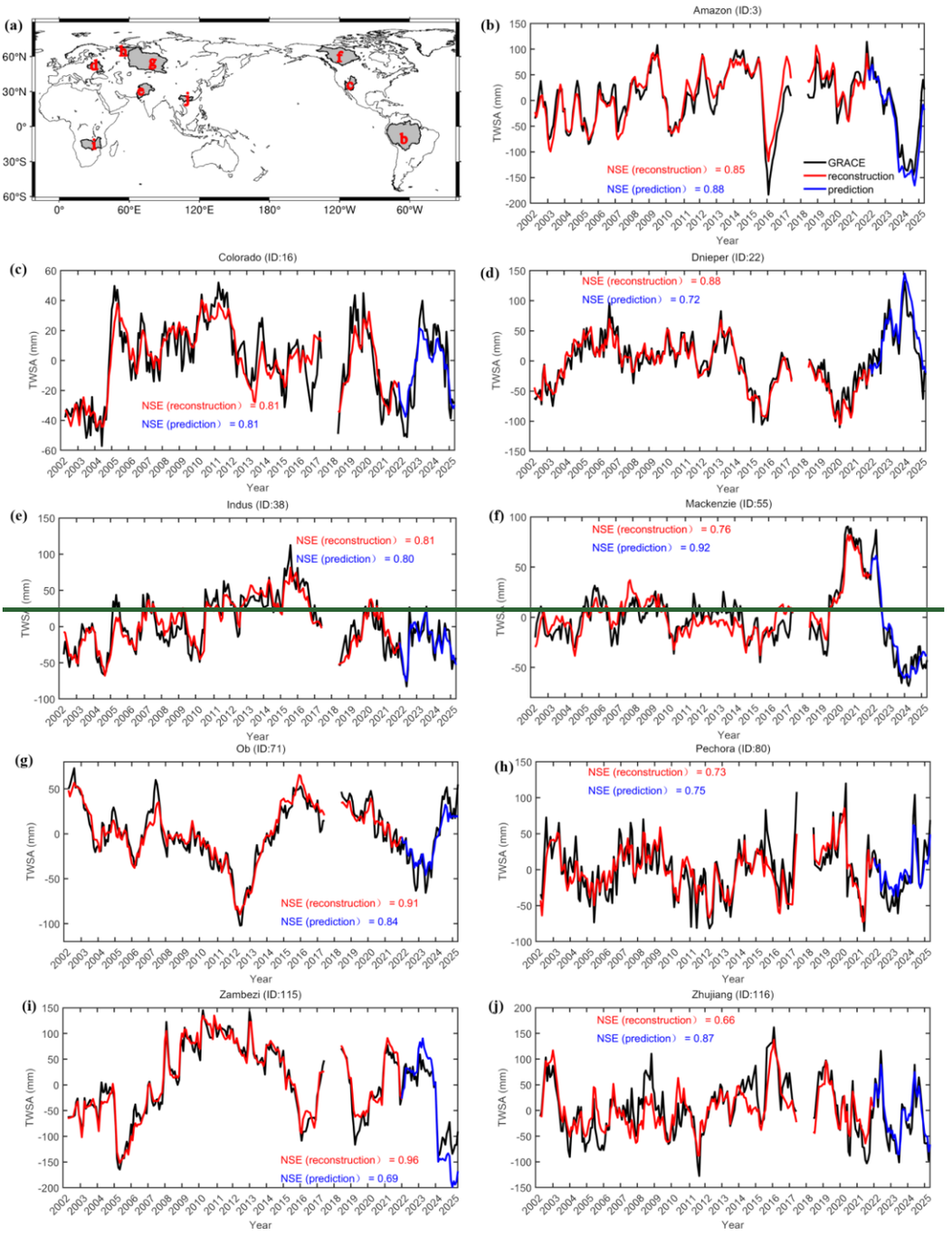


Figure ~~15~~16: Spatial distribution of the 30-day retention factor d^{30} (a) and the uncertainty of parameter d^{30} (b) across 116 global river basins.

~~5.3 Model Extrapolation Evaluation~~

~~In this study, model parameters were calibrated using GRACE data from 2002 to 2021 as the training period. These parameters were then held constant to extrapolate TWSA for the period from January 2022 to June 2025. To assess regional applicability, nine representative basins were selected (Fig. 16), spanning high-latitude cold regions, tropical humid zones, and typical monsoon-affected areas. For each basin, the predicted TWSA (blue line) was compared with the reconstructed values during the training period (red line) and GRACE observations (black line). The results show that in certain basins, such as the Amazon and Colorado, model performance during the extrapolation period remained comparable to or even exceeded that during the training period. This indicates that under these hydroclimatic conditions, the fixed-parameter model can reliably capture TWSA variations driven by precipitation and temperature. For example, during the period from 2022 to 2024, the Amazon Basin experienced a severe extreme drought event (Fig. 16b). This event was primarily driven by two factors: (1) a significant reduction in precipitation due to enhanced atmospheric stability and suppressed convection; and (2) a persistent increase in atmospheric evaporative demand under extreme high-temperature conditions (Fernández-Alvarez et al., 2025). In some other basins, however, the predicted NSE dropped substantially. This decline may be attributed to the limitations imposed by parameter nonstationarity and errors in the forcing data. These results suggest that when only precipitation and temperature forecasts are available, the proposed recursive model can be directly applied for estimating TWS changes during missing periods or into the future. Nevertheless, its applicability varies with hydroclimatic conditions, and for regions requiring high prediction accuracy, periodic parameter updates or the inclusion of additional forcing information may still be necessary.~~



~~Figure 16: Time series of GRACE TWSA (black), reconstructed TWSA (red), and four-parameter climate-driven forecasts (blue) from January 2022 to May 2025 in nine basins (b-j). The global distribution of the nine selected river basins (a).~~

5.4~~-3~~ Limitations of the study

This approach relies on external precipitation and temperature forecasts for predicting future TWSA. In regions where precipitation and temperature forecasts are relatively accurate, the model is capable of producing reliable climate-driven TWSA predictions. Conversely, uncertainties in the

forcing data can significantly affect the reconstruction results. Therefore, both the applicability and the upper limit of prediction accuracy are constrained by the quality of the forcing inputs. In addition, the current model is designed to simulate deseasonalized and detrended TWSA only, and is not capable of independently reproducing the full seasonal component or long-term trend.

In permafrost- and glacier-dominated basins, model performance is further limited. Taking the Yukon Basin as an example (Fig. 6a), during the period from 2011 to 2014, the average temperatures during the snowmelt seasons were significantly lower than normal for three consecutive years, while annual precipitation slightly exceeded the long-term mean (Fig. ~~S3~~S10). GRACE observations show that TWSA remained at a consistently high level during this period, whereas our reconstruction results are notably lower than those from GRACE. The primary source of error is likely associated with the sub-daily freezing and thawing cycles within the snowpack, which is not explicitly represented in the model. Meltwater generated during daytime warming is treated as immediate runoff and removed from the system, while at night, when temperatures drop below 0 ℃, a portion of this water refreezes and is retained in the snowpack. Since the model does not incorporate this refreezing into the subsequent day's initial storage, it systematically underestimates the net accumulation within the snow layer. As a result, the peak in winter and spring is dampened, and the overall reconstructed TWSA is biased low.

6 Conclusions

Building upon the classical linear water storage framework, this study introduces coupled temperature and precipitation factors to develop a four-parameter daily recursive model for reconstructing climate-driven TWSA at a $0.5° \times 0.5°$ spatial resolution over global land areas (excluding Antarctica) for the period 2002–2021. The model is calibrated against the JPL RL06 and GSFC RL06v2 mascon solutions, using hourly ERA5-Land precipitation and temperature data aggregated to daily time steps as forcing inputs.

By treating $TWS(0)$ as an additional parameter to be estimated jointly with the other model parameters, the influence of initial condition errors is substantially reduced, leading to improved reconstruction accuracy. Independent calibrations conducted for 116 major river basins demonstrate that the model parameters are more stable compared to existing approaches, while the reduced number of calibration iterations significantly enhances computational efficiency. At the gridded scale, the reconstruction closely aligns with GRACE/GRACE-FO mascon observations, with 62% of grid cells achieving $NSE > 0.5$ for JPL-REC and 53% for GSFC-REC. A sensitivity analysis

further clarifies the physical interpretation of the four model parameters and their respective roles in governing hydrological processes.

Comparison with existing statistical products demonstrates that the proposed model performs well across multiple spatial scales. At the basin scale, reconstructed TWSA results were compared with those from Zhong-REC. Among the 116 global river basins examined, 84 achieved a NSE greater than 0.7 using the proposed model, whereas only 59 basins (51%) exceeded this threshold with Zhong-REC. In most basins, the model yields improved performance, with the most substantial gains observed in arid and semi-arid regions. At the gridded scale, the reconstructed results were also compared with those from the Humphrey's model. After subtracting the Humphrey-JPL-REC results from the JPL-REC reconstruction, 63% of the grid cells have ΔNSE values greater than zero, indicating an overall improvement in model performance. The most pronounced improvements are observed across the Arabian Peninsula and large parts of Africa.

At the daily scale, the reconstructed TWSA agrees well overall with the ITSG-Grace2018 daily solution and GLDAS-2.2, indicating that the model can capture sub-monthly storage variability. The reconstruction demonstrates comparable performance during independent training and validation periods. The validation results show that about 60% of the basins retain NSE values greater than 0.5, and in representative basins such as the Amazon, Yangtze, Mississippi, and Murray, the validation period reconstructions remain consistent with GRACE/GRACE-FO observations in both amplitude and phase. These results indicate that the proposed framework does not exhibit overfitting and has good temporal extrapolation capability.

In summary, this study develops a climate-driven model of global TWS variability at the interannual scale, based on a simple four-parameter linear response framework that relies solely on precipitation and temperature as inputs. This study introduces optimization strategies to improve both computational efficiency and reconstruction accuracy. The method can be further extended to incorporate additional factors such as wind speed, radiation, and topography, or to adopt nonlinear relationships using machine learning approaches. However, the simplicity and high computational efficiency demonstrated by the proposed framework greatly enhance its applicability and interpretability. Future research will focus on simulating the seasonal and long-term trends of water storage and ultimately aim to separate and quantify the relative impacts of climatic factors and human activities on water resources across multiple temporal scales.

Competing interests

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.

Author contribution

P.X. conducted the data curation, formal analysis, and investigation; developed the model and software; and prepared the original manuscript draft. S.Y. conceptualized the study, acquired the funding, and supervised the research. Both authors contributed to the methodology and participated in the review and editing of the manuscript.

Acknowledgements

This research is funded by the National Natural Science Foundation of China (42374103, E214040201), the University of Chinese Academy of Sciences Research Start-up Grant (E3ER0402A2, 110400M003).

Data availability

The authors are grateful to the NASA Goddard Space Flight Center (https://earth.gsfc.nasa.gov/geo/data/grace-mascons), ~~and the~~ Jet Propulsion Laboratory (https://podaac.jpl.nasa.gov/dataset/TELLUS_GRAC-GRFO_MASCON_CRI_GRID_RL06.1_V3), and the Center for Space Research, University of Texas at Austin (http://www2.csr.utexas.edu/grace) for providing the GRACE mascon solutions (last access: 31 July 2025). The authors thank Zhong et al. (2025) for providing publicly available datasets and source code that supported the reconstruction of terrestrial water storage anomalies, which are accessible from https://doi.org/10.1038/s43247-024-01967-7. The authors also are grateful for the water storage reconstruction dataset provided by Humphrey and Gudmundsson (2019) (https://doi.org/10.6084/m9.figshare.7670849). ERA5-Land precipitation, temperature and evapotranspiration datasets are provided by https://doi.org/10.24381/cds.e2161bac 63(last access: 31 July 2025). MSWEP precipitation can be publicly obtained from http:// www.gloh2o.org/mswep/65(last access: 31 July 2025). GLDAS-GLSM data can be publicly

obtained from https://disc.gsfc.nasa.gov/datasets (last access: 31 July 2025). The boundary of global river basins is available at https://www.bafg.de/SharedDocs/ExterneLinks/GRDC/mrb_shp_zip.html62 (last access: 31 July2025). Aridity index can be publicly obtained from https://doi.org/10.6084/m9.figshare.7504448.v451 (last access: 31 July2025).

References

Abbott, B. W., Bishop, K., Zarnetske, J. P., Minaudo, C., Chapin, F. S., Krause, S., Hannah, D. M., Conner, L., Ellison, D., Godsey, S. E., Plont, S., Marçais, J., Kolbe, T., Huebner, A., Frei, R. J., Hampton, T., Gu, S., Buhman, M., Sara Sayedi, S., Ursache, O., Chapin, M., Henderson, K. D., and Pinay, G.: Human domination of the global water cycle absent from depictions and perceptions, Nat. Geosci., 12, 533-540, 10.1038/s41561-019-0374-y, 2019.

An, L., Wang, J., Huang, J., Pokhrel, Y., Hugonnet, R., Wada, Y., Cáceres, D., Müller Schmied, H., Song, C., Berthier, E., Yu, H., and Zhang, G.: Divergent Causes of Terrestrial Water Storage Decline Between Drylands and Humid Regions Globally, Geophys. Res. Lett., 48, e2021GL095035, https://doi.org/10.1029/2021GL095035, 2021.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., and Evans, R.: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, Bull. Am. Meteorol. Soc., 82, 2415-2434, 2001.

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, Hydrol. Earth Syst. Sci., 21, 589-615, 10.5194/hess-21-589-2017, 2017.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., McVicar, T. R., and Adler, R. F.: MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment, Bull. Am. Meteorol. Soc., 100, 473-500, https://doi.org/10.1175/BAMS-D-17-0138.1, 2019.

Burek, P. and Smilovic, M.: The use of GRDC gauging stations for calibrating large-scale hydrological models, Earth Syst. Sci. Data, 15, 5617-5629, 2023.

Castle, S. L., Thomas, B. F., Reager, J. T., Rodell, M., Swenson, S. C., and Famiglietti, J. S.: Groundwater depletion during drought threatens future water security of the Colorado River Basin, Geophys. Res. Lett., 41, 5904-5911, 2014.

Chen, J., Cazenave, A., Dahle, C., Llovel, W., Panet, I., Pfeffer, J., and Moreira, L.: Applications and challenges of GRACE and GRACE follow-on satellite gravimetry, Surv. Geophys., 43, 305-345, 2022.

Chen, J. L., Wilson, C. R., and Tapley, B. D.: The 2009 exceptional Amazon flood and interannual terrestrial water storage change observed by GRACE, Water Resour. Res., 46, 2010.

Chen, M., Shi, W., Xie, P., Silva, V. B., Kousky, V. E., Wayne Higgins, R., and Janowiak, J. E.: Assessing objective techniques for gauge‐based analyses of global daily precipitation, J. Geophys. Res.-Atmos., 113, 2008.

Crochemore, L., Isberg, K., Pimentel, R., Pineda, L., Hasan, A., and Arheimer, B.: Lessons learnt from checking the quality of openly accessible river flow data worldwide, Hydrol. Sci. J., 65, 699-711, 2020.

Duan, A., Zhong, Y., Xu, G., Yang, K., Tian, B., Wu, Y., Bai, H., and Hu, E.: Quantifying the

2022 extreme drought in the Yangtze River Basin using GRACE-FO, J. Hydrol., 630, 130680, 2024.

Fan, C., Song, C., Liu, K., Ke, L., Xue, B., Chen, T., Fu, C., and Cheng, J.: Century‐scale reconstruction of water storage changes of the largest lake in the inner Mongolia plateau using a machine learning approach, Water Resour. Res., 57, e2020WR028831, 2021.

Fekete, B. M., Looser, U., Pietroniro, A., and Robarts, R. D.: Rationale for monitoring discharge on the ground, J. Hydrometeorol., 13, 1977-1986, 2012.

Feng, J., Li, B., Song, J., Tang, B., Nyein, M. M., and Tani, B. P.: Spatiotemporal Variations of Terrestrial Water Storage and Driving Factors in the Water Towers of Northwest China Based on GRACE and Multi-Source Data Sets, Water Resour. Res., 61, e2024WR039490, https://doi.org/10.1029/2024WR039490, 2025.

Fernández-Alvarez, J. C., Nieto, R., Vicente-Serrano, S. M., Carvalho, D., and Gimeno, L.: Moisture and temperature sources were key drivers of the anomalies for the record-breaking of 2023 Amazon drought, Commun. Earth Environ., 6, 801, 10.1038/s43247-025-02771-7, 2025.

Forootan, E., Khaki, M., Schumacher, M., Wulfmeyer, V., Mehrnegar, N., van Dijk, A. I., Brocca, L., Farzaneh, S., Akinluyi, F., and Ramillien, G.: Understanding the global hydrological droughts of 2003–2016 and their relationships with teleconnections, Sci. Total Environ., 650, 2587-2604, 2019.

Gao, Y., Luo, Z., Liu, H., Wang, L., Chen, X., and Li, H.: Reconstruction of global long-term daily streamflow dataset using machine learning models for revealing streamflow changes, Journal of Hydrology: Regional Studies, 64, 103148, 2026.

Humphrey, V. and Gudmundsson, L.: GRACE-REC: a reconstruction of climate-driven water storage changes over the last century, Earth Syst. Sci. Data, 11, 1153-1170, 10.5194/essd-11-1153-2019, 2019.

Humphrey, V., Gudmundsson, L., and Seneviratne, S. I.: Assessing Global Water Storage Variability from GRACE: Trends, Seasonal Cycle, Subseasonal Anomalies and Extremes, Surv. Geophys., 37, 357-395, 10.1007/s10712-016-9367-1, 2016.

Humphrey, V., Rodell, M., and Eicker, A.: Using Satellite-Based Terrestrial Water Storage Data: A Review, Surv. Geophys., 44, 1489-1517, 10.1007/s10712-022-09754-9, 2023.

Laudon, H., Spence, C., Buttle, J., Carey, S. K., McDonnell, J. J., McNamara, J. P., Soulsby, C., and Tetzlaff, D.: Save northern high-latitude catchments, Nat. Geosci., 10, 324-325, 2017.

Li, B., Rodell, M., Kumar, S., Beaudoing, H. K., Getirana, A., Zaitchik, B. F., de Goncalves, L. G., Cossetin, C., Bhanja, S., Mukherjee, A., Tian, S., Tangdamrongsub, N., Long, D., Nanteza, J., Lee, J., Policelli, F., Goni, I. B., Daira, D., Bila, M., de Lannoy, G., Mocko, D., Steele-Dunne, S. C., Save, H., and Bettadpur, S.: Global GRACE Data Assimilation for Groundwater and Drought Monitoring: Advances and Challenges, Water Resour. Res., 55, 7564-7586, https://doi.org/10.1029/2018WR024618, 2019.

Li, F., Kusche, J., Chao, N., Wang, Z., and Löcher, A.: Long‐term (1979‐present) total water storage anomalies over the global land derived by reconstructing GRACE data, Geophys. Res. Lett., 48, e2021GL093492, 2021.

Li, F., Kusche, J., Rietbroek, R., Wang, Z., Forootan, E., Schulze, K., and Lück, C.: Comparison of data‐driven techniques to reconstruct (1992–2002) and predict (2017–2018) GRACE‐like gridded total water storage changes using climate inputs, Water Resour. Res., 56, e2019WR026551, 2020.

Li, X., Jin, T., Liu, B., Chao, N., Li, F., and Cai, Z.: The Influence of ENSO on the Long‐Term Water Storage Anomalies in the Middle‐Lower Reaches of the Yangtze River Basin: Evaluation and Analysis, Earth Space Sci., 10, e2023EA003007, 2023.

Liu, B., Zou, X., Yi, S., Sneeuw, N., Cai, J., and Li, J.: Identifying and separating climate-and human-driven water storage anomalies using GRACE satellite data, Remote Sens. Environ., 263, 112559, 2021.

Liu, B., Zou, X., Yi, S., Sneeuw, N., Li, J., and Cai, J.: Reconstructing GRACE-like time series of high mountain glacier mass anomalies, Remote Sens. Environ., 280, 113177, 2022.
Liu, X., Feng, X., Ciais, P., and Fu, B.: Widespread decline in terrestrial water storage and its link to teleconnections across Asia and eastern Europe, Hydrol. Earth Syst. Sci., 24, 3663-3676, 10.5194/hess-24-3663-2020, 2020.
Long, D., Longuevergne, L., and Scanlon, B. R.: Global analysis of approaches for deriving total water storage changes from GRACE satellites, Water Resour. Res., 51, 2574-2594, 2015.
Long, D., Shen, Y., Sun, A., Hong, Y., Longuevergne, L., Yang, Y., Li, B., and Chen, L.: Drought and flood monitoring for a large karst plateau in Southwest China using extended GRACE data, Remote Sens. Environ., 155, 145-160, 2014.
Loomis, B., Luthcke, S., and Sabaka, T.: Regularization and error characterization of GRACE mascons, Journal of geodesy, 93, 1381-1398, 2019.
Ma, N., Zhang, Y., and Szilagyi, J.: Water-balance-based evapotranspiration for 56 large river basins: A benchmarking dataset for global terrestrial evapotranspiration modeling, J. Hydrol., 630, 130607, 2024.
Muñoz-Sabater, J.: ERA5-Land hourly data from 1981 to present, Copernicus climate change service (C3S) climate data store (CDS), 10, 2019.
Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., and Hersbach, H.: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349-4383, 2021.
Ni, S., Chen, J., Wilson, C. R., Li, J., Hu, X., and Fu, R.: Global Terrestrial Water Storage Changes and Connections to ENSO Events, Surv. Geophys., 39, 1-22, 10.1007/s10712-017-9421-7, 2018.
Palazzoli, I., Ceola, S., and Gentine, P.: GR Ai CE: reconstructing terrestrial water storage anomalies with recurrent neural networks, Sci. Data, 12, 146, 2025.
Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., and Humphrey, M.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Sci. Data, 7, 225, 2020.
Riegger, J. and Tourian, M. J.: Characterization of runoff‐storage relationships by satellite gravimetry and remote sensing, Water Resour. Res., 50, 3444-3466, 2014.
Rodell, M., Velicogna, I., and Famiglietti, J. S.: Satellite-based estimates of groundwater depletion in India, Nature, 460, 999-1002, 2009.
Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J., Beaudoing, H. K., Landerer, F. W., and Lo, M.-H.: Emerging trends in global freshwater availability, Nature, 557, 651-659, 2018.
Save, H., Bettadpur, S., and Tapley, B. D.: High‐resolution CSR GRACE RL05 mascons, Journal of Geophysical Research: Solid Earth, 121, 7547-7569, 2016.
Sun, A. Y., Scanlon, B. R., Save, H., and Rateb, A.: Reconstruction of GRACE total water storage through automated machine learning, Water Resour. Res., 57, e2020WR028666, 2021.
Sun, Z., Long, D., Yang, W., Li, X., and Pan, Y.: Reconstruction of GRACE data on changes in total water storage over the global land surface and 60 basins, Water Resour. Res., 56, e2019WR026250, 2020.
Tapley, B. D., Watkins, M. M., Flechtner, F., Reigber, C., Bettadpur, S., Rodell, M., Sasgen, I., Famiglietti, J. S., Landerer, F. W., and Chambers, D. P.: Contributions of GRACE to understanding climate change, Nature climate change, 9, 358-369, 2019.
Tian, K., Wang, Z., Li, F., Gao, Y., Xiao, Y., and Liu, C.: Drought Events over the Amazon River Basin (1993–2019) as Detected by the Climate-Driven Total Water Storage Change, Remote Sens., 13, 1124, 2021.
Trabucco, A. and Zomer, R. J.: Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v3, International Center for Tropical Agriculture (CIAT) and World Agroforestry

Centre (ICRAF) [dataset], 10.6084/m9.figshare.7504448.v3, 2019.
Voss, K. A., Famiglietti, J. S., Lo, M., De Linage, C., Rodell, M., and Swenson, S. C.: Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris‐Euphrates‐Western Iran region, Water Resour. Res., 49, 904-914, 2013.
Wiese, D., Yuan, D., Boening, C., Landerer, F. W., and Watkins, M.: JPL GRACE and GRACE-FO Mascon Ocean, Ice, and Hydrology Equivalent Water Height Coastal Resolution Improvement (CRI) Filtered Release 06 Version 02, DAAC: Pasadena, CA, USA, 2019.
Xiao, C., Zhong, Y., Wu, Y., Zhang, Z., Bai, H., and Li, Z.: Flood Evolution in the Past 60 Years Revealed by Reconstructed Daily Terrestrial Water Storage Anomalies in China, Water Resour. Res., 61, e2024WR038712, https://doi.org/10.1029/2024WR038712, 2025.
Xu, L., Chen, N., Moradkhani, H., Zhang, X., and Hu, C.: Improving global monthly and daily precipitation estimation by fusing gauge observations, remote sensing, and reanalysis data sets, Water Resour. Res., 56, e2019WR026444, 2020.
Yi, S., Saemian, P., Sneeuw, N., and Tourian, M. J.: Estimating runoff from pan-Arctic drainage basins for 2002–2019 using an improved runoff-storage relationship, Remote Sens. Environ., 298, 113816, 2023.
Yi, S., Sun, W., Feng, W., and Chen, J.: Anthropogenic and climate-driven water depletion in Asia, Geophys. Res. Lett., 43, 9061-9069, https://doi.org/10.1002/2016GL069985, 2016.
Yin, J., Slater, L. J., Khouakhi, A., Yu, L., Liu, P., Li, F., Pokhrel, Y., and Gentine, P.: GTWS-MLrec: global terrestrial water storage reconstruction by machine learning from 1940 to present, Earth Syst. Sci. Data, 15, 5597-5615, 2023.
Zhang, D., Zhang, Q., Werner, A. D., and Liu, X.: GRACE-based hydrological drought evaluation of the Yangtze River Basin, China, J. Hydrometeorol., 17, 811-828, 2016.
Zhang, J., Liesch, T., and Goldscheider, N.: Impacts of climate change and human activities on global groundwater storage from 2003 to 2022, J. Hydrol., 664, 134298, https://doi.org/10.1016/j.jhydrol.2025.134298, 2025.
Zhang, L. and Sun, W.: Progress and prospect of GRACE Mascon product and its application, Reviews of Geophysics and Planetary Physics, 53, 35-52, 2022.
Zhang, Z., Chao, B., Chen, J., and Wilson, C.: Terrestrial water storage anomalies of Yangtze River Basin droughts observed by GRACE and connections with ENSO, Glob. Planet. Change, 126, 35-45, 2015.
Zheng, S., Zhang, Z., Yan, H., Zhao, Y., and Li, Z.: Characterizing drought events occurred in the Yangtze River Basin from 1979 to 2017 by reconstructing water storage anomalies based on GRACE and meteorological data, Sci. Total Environ., 868, 161755, https://doi.org/10.1016/j.scitotenv.2023.161755, 2023.
Zhong, Y., Feng, W., Humphrey, V., and Zhong, M.: Human-Induced and Climate-Driven Contributions to Water Storage Variations in the Haihe River Basin, China, Remote Sens., 11, 3050, 2019.
Zhong, Y., Tian, B., Kim, H., Yuan, X., Liu, X., Zhu, E., Wu, Y., Wang, L., and Wang, L.: Over 60% precipitation transformed into terrestrial water storage in global river basins from 2002 to 2021, Commun. Earth Environ., 6, 53, 2025.
Zomer, R. J., Xu, J., and Trabucco, A.: Version 3 of the Global Aridity Index and Potential Evapotranspiration Database, Sci. Data, 9, 409, 10.1038/s41597-022-01493-1, 2022.