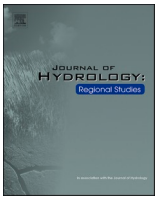
Journal of Hydrology: Regional Studies 50 (2023) 101528

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh



Reconstructing GRACE-derived terrestrial water storage anomalies with in-situ groundwater level measurements and meteorological forcing data

Peijun Li [a], Yuanyuan Zha [a,b,*], Chak-Hau Michael Tso [c]

[a] State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan 430072, China
[b] Guangxi Key Laboratory of Water Engineering Materials and Structures, Guangxi Institute of Water Resources Research, Nanning 530015, China
[c] UK Centre for Ecology and Hydrology, Lancaster, UK

ARTICLE INFO

Keywords:
GRACE
Terrestrial water storage
Groundwater level
Land surface model
North China Plain

ABSTRACT

Study region: North China Plain (NCP), China, a semi-arid region with intense groundwater withdrawals.
Study focus: This paper developed a framework using meteorological data, model-simulated terrestrial water storage anomalies (TWSA), and additional in-situ (groundwater level, GL) data to improve the unsatisfactory GRACE-TWSA reconstruction in arid and semi-arid regions due to the intense anthropogenic influence on groundwater. The inconsistency between point-scale data (GL) and grid-scale data (GRACE-TWSA and predictors other than GL) is handled by feature extraction techniques. Moreover, to deal with temporal non-stationarity, the time series are separated into trend and detrended components, the patterns of which are further learned by linear and nonlinear machine learning models, respectively.
New hydrological insights for the region: Multi-site GL observations in NCP can not only serve as validation data but also as predictors providing invaluable information on human effects for the reconstructed TWSA improvement (from 6.51 to 3.86 cm for Root Mean Square Error and from 0.56 to 0.82 for Nash-Sutcliffe Efficiency). Our results show that multi-site GL data in NCP are highly inter-correlated and can be represented by several principal components, demonstrating the strong hydraulic connectivity in NCP. We also find a significant one-month lag and linear relationship between the trends of GRACE-TWSA and GL changes in NCP. These deeper understandings of hydrologic processes have implications for enhancing the GRACE-TWSA estimations in other similar regions.

1. Introduction

Launched in March 2002, the Gravity Recovery and Climate Experiment (GRACE) satellites have provided unprecedentedly accurate terrestrial water storage anomaly (TWSA) observations (Tapley et al., 2004). Due to a battery issue (Landerer et al., 2020), GRACE ended its science operations in October 2017, which generated data from April 2002 to June 2017. In May 2018, the successor mission GRACE Follow-On was launched to continue the measurements and provided its first observation in June 2018. Hence, the GRACE-TWSA data is missing in the 11-month data gap and the time before the GRACE launches. This affects studies requiring long

* Corresponding author at: State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan 430072, China.
E-mail address: zhayuan87@whu.edu.cn (Y. Zha).

https://doi.org/10.1016/j.ejrh.2023.101528
Received 3 February 2023; Received in revised form 7 September 2023; Accepted 9 September 2023
Available online 16 September 2023

and continuous time series, e.g., agricultural water and groundwater resource management (Frappart and Ramillien, 2018), long-term climate change (Tapley et al., 2019), hydrological drought (Mo, Zhong, Forootan, Shi et al., 2022), and links between TWSA and teleconnections (Guo et al., 2021; P. Li et al., 2022).

Nonetheless, there are studies on the GRACE-TWSA reconstruction both on a global (e.g., F. Li et al., 2020; Mo, Zhong, Forootan, Mehrnegar et al., 2022; Z. Sun et al., 2020; Wang et al., 2021; Yi and Sneeuw, 2021; X. Zhang et al., 2022) and regional (e.g., Becker et al., 2011; A. Y. Sun et al., 2019; X. Yang et al., 2021) scale. However, these studies generally show good performances in humid regions but unsatisfying results in arid and semi-arid areas where intense groundwater withdrawals often exist (e.g., Z. Sun et al., 2020; X. Zhang et al., 2022). The reason is that meteorological data, model-based products (e.g., Scanlon et al., 2018), or coarse satellite observations (e.g., Forootan et al., 2020) used to predict GRACE-TWSA are not competent to reflect the impact of human influences on groundwater storage (B. Liu et al., 2022). The ineptitude of model-based products (e.g., TWSA simulated by the NOAH model (Rodell et al., 2005)) is because the existing land surface models usually do not contain modules to describe groundwater storage changes influenced by human activities. For instance, A. Y. Sun et al. (2019) reported that the mismatch values between the GRACE observations and reconstructions over India are mainly induced by the missing groundwater storage component in NOAH-TWSA. Although the long-term TWSA variations caused by human activities can be partially extracted by trend analysis of the GRACE data (F. Li et al., 2021; Mo, Zhong, Forootan, Mehrnegar et al., 2022; X. Yang et al., 2021), short-term variation induced by human activity without GRACE observation can only be detected with the help of in-situ groundwater level (GL) data (Ahmed et al., 2021; B. Liu et al., 2022).

It seems that the inclusion of GL data in regions with intense groundwater withdrawals is the only way to provide information on the anthropogenic influence on groundwater during GRACE-TWSA reconstruction (C. Zhang et al., 2021). However, it is not straightforward to include this type of data since the GL data from groundwater observation wells have very different spatial properties compared to GRACE-TWSA and other predictors for reconstruction. For instance, GRACE-TWSA has a fixed horizontal support scale, while that of the GL data is highly variable, depending on local hydrogeological properties and regional groundwater flow field, and it is data-demanding to obtain this parameter based on variogram analysis (Varouchakis et al., 2019). In the meantime, the GL data are often observed at a specific depth in a multi-layer aquifer-aquitard system, and their information must be synthesized to obtain the

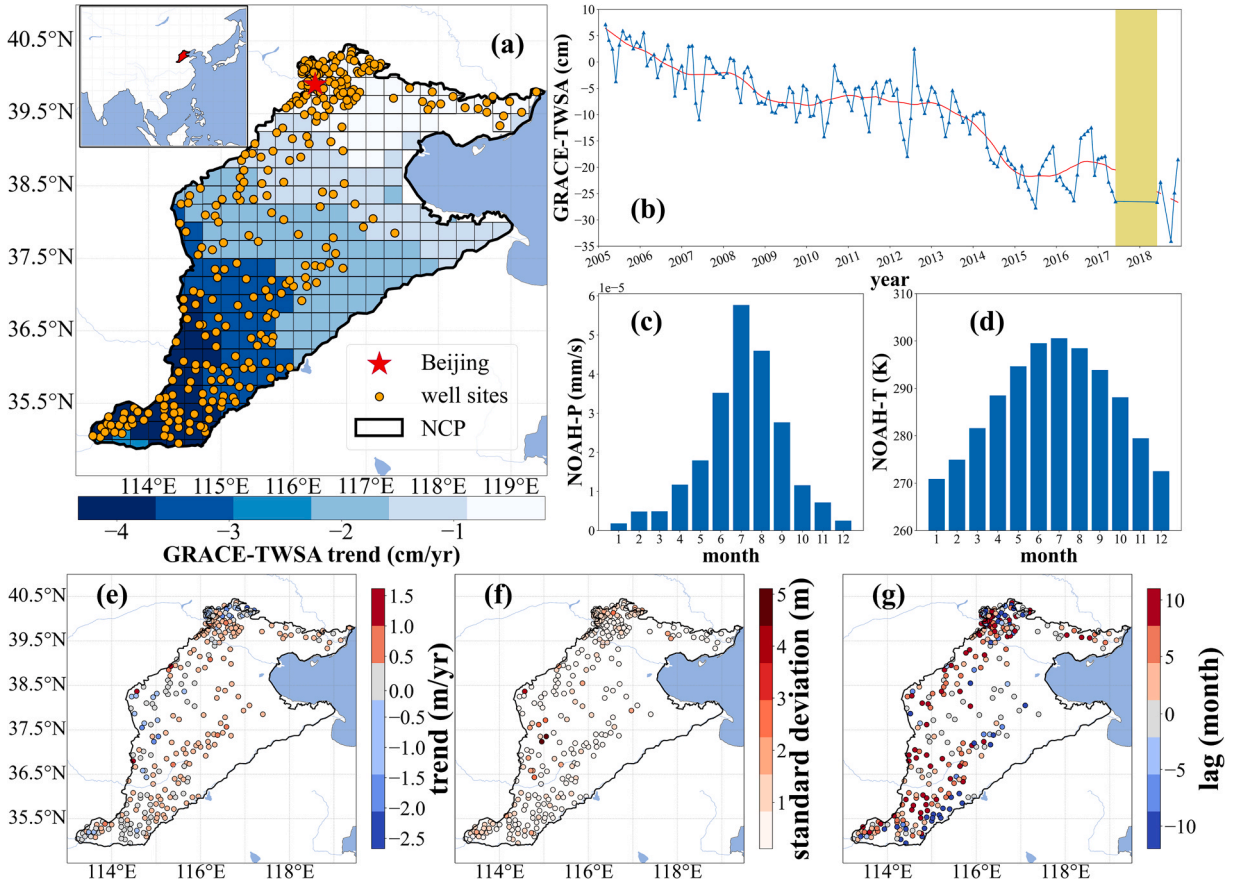


Fig. 1. (a) The location of the North China Plain (NCP), with grid color indicating the trend of GRACE-TWSA during 2005 – 2018; (b) the spatially averaged (blue line) GRACE-TWSA and its trend (red line) with data gap between GRACE and GRACE Follow-On highlighted by the olive box; (c)~(d) the monthly variations of precipitation and temperature derived from NOAH products during 2002–2016; (e)~(g) the trends, standard deviations, and lags (month) of multi-site groundwater levels.

depth-integrated TWSA information. Although spatial interpolation methods (e.g., kriging) can unify the spatial scale of GL and GRACE-TWSA, it has several technical difficulties, including: 1. the depth/aquifer layer information for the GL observation may not be available to ensure accuracy during vertical integration of GL data for deriving TWSA; 2. it is difficult to obtain spatial variogram information based on time series from the sparse well network since the spatiotemporal kriging method is quite computational demanding and requires human expertise (van Zoest et al., 2020).

Besides the difficulty induced by spatial scale mismatch, temporal non-stationarity of GL and GRACE-TWSA time series brings additional difficulties in TWSA reconstruction, since a single machine learning method usually cannot perform well simultaneously for trend inter/extrapolation and characterization of short-term variation (P. Li et al., 2020; Shen et al., 2023). For instance, nonlinear machine learning approaches (e.g., random forest - RF) have a powerful ability to characterize complex, short-term variation, while linear approaches, such as linear regression, are more stable for learning long-term trends. Combining the advantages of different machine learning models is necessary to learn the two-scale temporal features of GRACE-TWSA and its predictors.

Aiming at these two issues, we propose a novel framework for the GRACE-TWSA reconstruction. Feature extraction techniques, with the powerful ability to extract the main features from large amounts of low-quality data (Hsiao et al., 2017; Zhao et al., 2019), tend to be useful tools for solving the spatial scale mismatch issue of GL data. Among them, principal component analysis (PCA, Pearson, 1901) and independent component analysis (ICA, Hyvärinen et al., 2001) are successfully applied in deriving the main variations or natural source signals of groundwater level data (e.g., Hsiao et al., 2017; Tsai and Hsiao, 2020), which can be important predictors for the GRACE TWSA reconstruction. Therefore, PCA and ICA are selected to be respectively used to derive the main features of multi-well GL data. To learn the two-scale temporal feature (long-term trend, and short-term variation) simultaneously, we propose to decompose the time series into trend and detrended parts (F. Li et al., 2020), which are trained by linear (multiple linear regression (MLR) (Montgomery et al., 2021)) and nonlinear machine learning (RF) (Ho, 1995; Jing et al., 2020) approaches, respectively, to learn the functional relationship between GRACE-TWSA and predictors. Compared to previous studies (F. Li et al., 2020; Mo, Zhong, Forootan, Mehrnegar et al., 2022; X. Yang et al., 2021) without the inclusion of in-situ GL data, the temporal decomposition is a key step in this study since it can learn the critical trend information impacted by groundwater withdrawal more effectively. In summary, the proposed framework only requires several hyperparameters (e.g., the number of leading principal components) to be determined manually, and we can avoid data-demanding spatial variogram analysis and instability induced by data extrapolation via nonlinear machine learning.

2. Material and methods

2.1. Study area

The North China Plain (NCP) is a well-known plain (~140,000 km^2, Fig. 1(a)) providing ~20% of the total crop production in China each year (Mukherjee et al., 2021): It is located in a semi-arid region based on the Global Aridity Index (Trabucco and Zomer, 2019) (Readers please refer to Fig. S1 in the supplementary material). Intensive groundwater extraction (R. Liu et al., 2022) for irrigation has caused significant groundwater depletion in NCP (Mukherjee et al., 2021; Zhao et al., 2019). For capturing the groundwater dynamic changes, a national groundwater monitoring well network has been established since the last century, providing abundant GL data (C. Zhang et al., 2021; Z. Zhang et al., 2009). Complex hydrogeology (Sakura et al., 2003; Z. Zhang et al., 2009), groundwater withdrawal activity, and impact from the South-to-North Water Diversion (SNWD) (R. Liu et al., 2022; Long et al., 2020) have led to a multi-scale spatiotemporal variable pattern of the groundwater storage. The reconstruction of GRACE-TWSA at NCP generally yielded unsatisfactory performance (e.g., Humphrey and Gudmundsson, 2019; Z. Sun et al., 2020), probably due to the ignorance of in-situ GL data that contains valuable groundwater withdrawal information. Thus, NCP is an ideal study area to test our hypothesis and the robustness of the proposed TWSA reconstruction framework.

2.2. Data

Reported to have several advantages over standard spherical harmonic solutions without some dedicated processing (Xiang et al., 2022; Zhao et al., 2019), the RL06 GRACE mascon-solution monthly TWSA in equivalent water thickness based on 0.25° grids (real resolution 3°) by the University of Texas at Austin's Center for Space Research are used in this study, which performs similarly in most scenarios to mascon data of the Jet Propulsion Laboratory, USA, based on 0.5° grids (Save et al., 2016; Scanlon et al., 2016) and have the same grid size as the Global Land Data Assimilation System (GLDAS) Version 2.1 NOAH 0.25° products as described in the following contents. As shown in Fig. 1(a), the study area includes 270 0.25° grid cells, and the study period is between January 2005 and December 2018 (155 epochs) (Fig. 1(b)). The scale factors are not used (Scanlon et al., 2016) here since GRACE-TWSA is reconstructed grid-by-grid.

The GLDAS NOAH monthly simulations (Rodell et al., 2005) relate well with GRACE-TWSA (Long et al., 2014; T. Yang et al., 2013) and are widely used for GRACE reconstruction (Gyawali et al., 2022). Therefore, the NOAH products (i.e., precipitation (NOAH-P), temperature (NOAH-T), and TWSA (NOAH-TWSA)) based on GLDAS Version 2.1 with 0.25° grid size (Rui and Beaudoing, 2020) are used in this study. The multi-year monthly variations of NOAH-P and NOAH-T, which exhibit climatology features in NCP, are shown in Fig. 1(c) ~ (d). It should be noted that the NOAH model only simulates water storage up to the top 2 m, and does not include forcing induced by human activities.

To complement such missing information, we collected the in-situ GL monthly observations in NCP from Jan 2005 - Dec 2018 from the Data Sharing Repository of the National Earth System Science Data Center. The dataset contains GL observations from 559 observation wells. Only 300 sites with 168 epochs (Fig. 1(a)) remain after excluding some discontinuous and low-quality data. It should be noted that the observation depth and auxiliary hydrogeology are not publicly available from the GL dataset. The GL data are further significantly heterogeneous (Fig. 1(e) ~ (g)), and the observation wells are not evenly distributed with two clusters in the southwest and far north while wells are missing in the east and southeast, making it difficult to establish the relationship between point-scale GL data and depth-integrated, grid-scale TWSA data using traditional kriging techniques (Adams et al., 2022).

To show the superiority and significance of the reconstructed GRACE-TWSA in this study, two available datasets of reconstructed GRACE-TWSA published in two previous studies, i.e., Mo et al. (2022) and Rateb et al. (2022) are collected. The spatiotemporal comparisons between their and our datasets are conducted based on similarity to the raw GRACE-TWSA data. Additionally, to quantify the uncertainties of our reconstructed GRACE-TWSA, the precipitation and temperature data from two different data sources are also collected, i.e., Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) (Gelaro et al., 2017) and ERA5-Land (Muñoz Sabater, J, 2019). Both of them are resampled into the same spatiotemporal resolution as the NOAH products. Therefore, three sets of reconstructed GRACE-TWSA data in total can be generated based on our developed method to quantify their uncertainties.

To evaluate the link between the reconstructed GRACE-TWSA and the teleconnection, we collect the Multivariate El Niño–Southern Oscillation (ENSO) Index version 2 (T. Zhang et al., 2019), which has been reported to have significant effects on TWSA globally (Ni et al., 2018; Phillips et al., 2012; Zizhan Zhang et al., 2015)).

2.3. Reconstruction framework

We intend to reconstruct the grid-scale GRACE-TWSA data using the grid-scale model outputs (i.e., NOAH-P, NOAH-T, and NOAH-TWSA) and point-scale in-situ GL data. Precipitation and temperature are major meteorological factors influencing TWSA. In addition, NOAH-TWSA can potentially provide information about TWSA from a modeling perspective, but neither groundwater storage deeper than 2 m nor anthropogenic activities (e.g., groundwater irrigation) are considered in the model. Thus, GL data is a necessary sup-

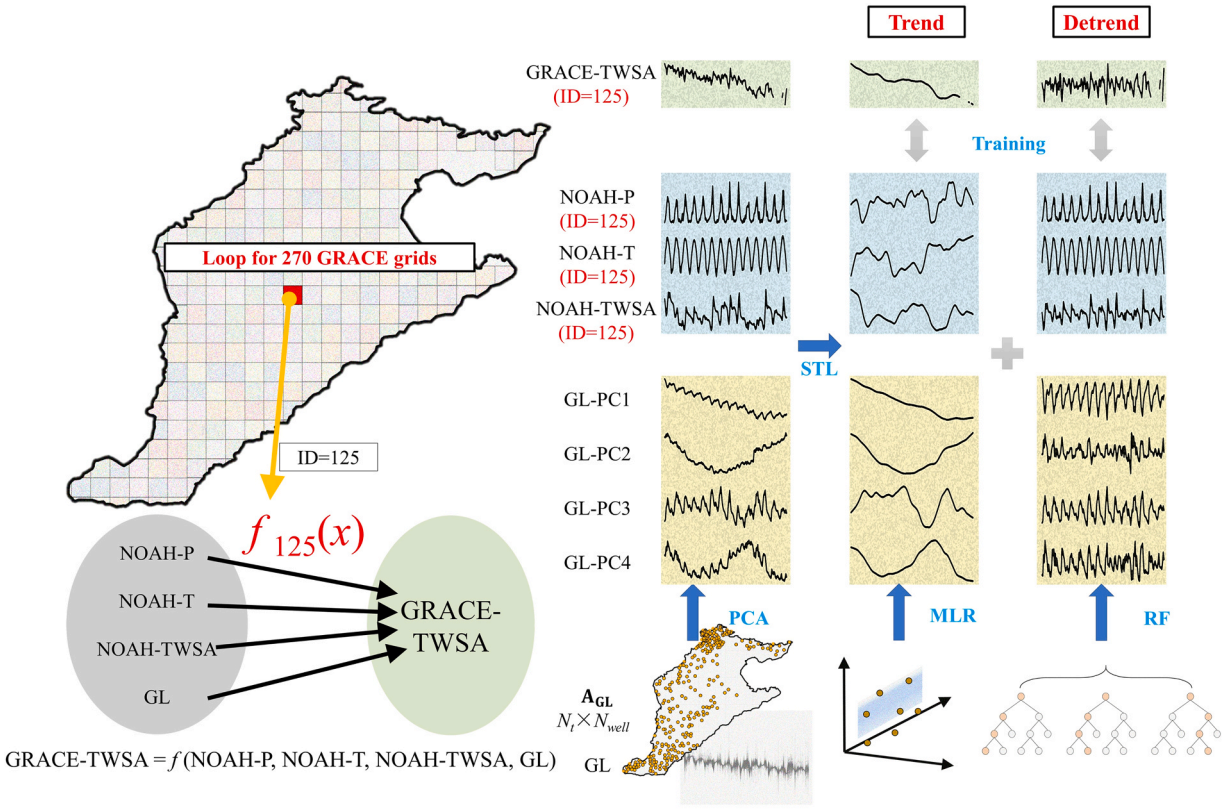


Fig. 2. Scheme of grid-wise data processing in this study. NOAH-P, NOAH-T, and NOAH-TWSA denote the precipitation, temperature, and terrestrial water storage anomaly derived from the NOAH models in one grid, respectively. GL-PCs represent principal components derived from all groundwater level (GL) data via the principal component analysis (PCA). Multiple linear regression (MLR) and random forest (RF) are applied to simulate the trend and detrended components decomposed by the seasonal decomposition of time series by loess (STL) approach, respectively.

plementary. The relationship between GRACE-TWSA and these predictors is nonlinear (Fig. 2) and can be described as

$$\text{GRACE-TWSA} = f(\text{NOAH-P, NOAH-T, NOAH-TWSA, GL}) \tag{1}$$

We can find this relationship f using groundwater flow modeling, spatiotemporal co-kriging, or machine learning. The machine learning methods are preferable since groundwater flow modeling requires much computational time during calibration, while spatiotemporal co-kriging requires a delicate spatiotemporal variogram analysis, but the GL observation depth information is not available from the GL dataset.

The relationship f is built on a grid basis. In other words, the reconstruction via the machine learning method will be conducted separately for the 270 grids. We do not consider the spatial correlation of the grid-scale predictors (i.e., NOAH-P, NOAH-T, and NOAH-TWSA) in the prediction since the correlation is not obvious across such coarse grids, and this simplification will save training time significantly. In contrast, the spatial correlation of GL data is implicitly considered since we extract the overall feature (e.g., principal components) of those 300 observation wells.

2.3.1. *Spatial preprocessing*

The predictors from the NOAH model (i.e., NOAH-P, NOAH-T, and NOAH-TWSA) have the same spatial grid as GRACE-TWSA, while point-scale GL data require special treatment for spatial matching before grid-wise reconstruction. We employ powerful feature extraction methods, i.e., PCA and ICA, to derive the major temporal patterns of groundwater storage at the 300 observation wells.

Herein we use PCA (see Fig. 2) to demonstrate the spatial preprocessing of the GL data matrix $\mathbf{A}_{\mathbf{GL}}$:

$$\mathbf{A}_{\mathbf{GL}} = \mathbf{S}_{\mathbf{PCA}}\mathbf{V}^T, \tag{2}$$

where $\mathbf{A}_{\mathbf{GL}}$ is the GL data matrix with size $N_t \times N_{well}$ (N_t is the number of months for observations and N_{well} is the number of observation sites), $\mathbf{S}_{\mathbf{PCA}}$ ($N_t \times N_{well}$) is the matrix storing principal components, and the column importance is usually arranged in descending order, $\mathbf{V}$ ($N_{well} \times N_{well}$) is the weight matrix with normalized columns (i.e., $\mathbf{V}\mathbf{V}^T = \mathbf{I}$). From a standard variance analysis, we find that keeping the first few principal components can approximate $\mathbf{A}_{\mathbf{GL}}$ well:

$$\mathbf{A}_{\mathbf{GL}} \approx \widetilde{\mathbf{S}}_{\mathbf{PCA}}\widetilde{\mathbf{V}}^T, \tag{3}$$

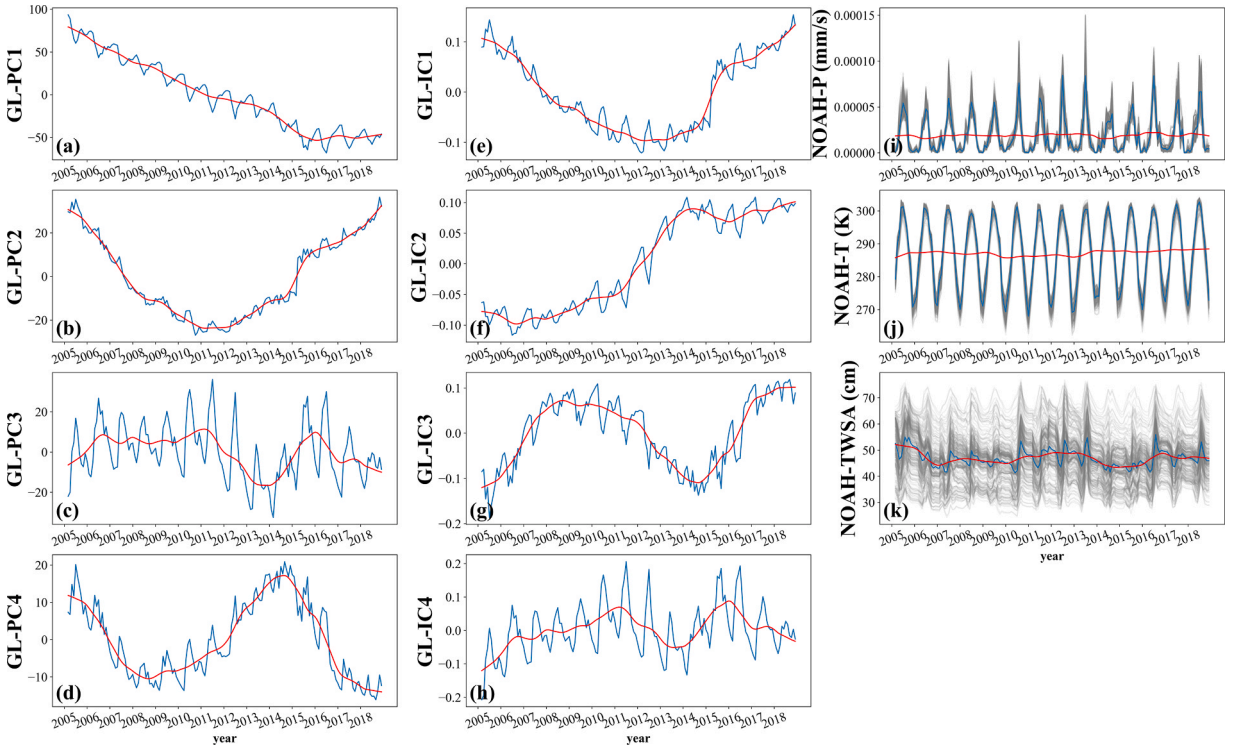


Fig. 3. (a)~(d) The principal components (GL-PCs) and (e)~(h) the independent components (GL-ICs) of the GL data. (i)~(k) The spatial average changes of NOAH-derived precipitation (NOAH-P), temperature (NOAH-T), and terrestrial water storage (NOAH-TWSA). The light grey lines in (i)~(k) denote the time series changes in different grids. The red line denotes the averaged trends of temporal changes (blue line) derived via the seasonal decomposition of time series by loess.

where $\widetilde{\mathbf{S}}_{\mathbf{PCA}}$ ($N_t \times n$) and $\widetilde{\mathbf{V}}$ ($N_{well} \times n$) are reduced-order matrices of $\mathbf{S}_{\mathbf{PCA}}$ and $\mathbf{V}$, respectively. We rewrite $\widetilde{\mathbf{S}}_{\mathbf{PCA}} = [\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}, \ldots, \mathbf{p_n}]$, where $\mathbf{p}_i$ ($N_t \times 1$, $i = 1, 2, 3, \ldots, n$) is one of the first n principal components, and $\widetilde{\mathbf{V}} = [\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \ldots, \mathbf{v_n}]$, where $\mathbf{v}_i$ ($N_{well} \times 1$, $i = 1, 2, 3, \ldots, n$) are the corresponding n column weights. Afterward, the observation GL time series at an arbitrary well k (i.e., $\mathbf{a}_k$, column k of $\mathbf{A}_{\mathbf{GL}}$), even at unobserved places, can be approximated as:

$$\mathbf{a}_k \approx \sum_{i=1}^{n} v_{ik}\mathbf{p}_i, \tag{4}$$

where weight v_{ik} is the k^{th} element of the vector $\mathbf{v}_i$. As indicated in this equation, the first n principal components $\mathbf{p}_i$ ($i = 1, 2, 3, \ldots, n$) represent the major temporal dynamic of GL in NCP, and the GL data at any location, including unobserved places in different spatial scales, can be characterized by the weighted (with weights v_{ik}, $i = 1, 2, 3, \ldots, n$) summation of the n principal components $\mathbf{p}_i$. Analogically, we then use $\mathbf{p}_i$ ($i = 1, 2, 3, \ldots, n$) instead of GL data $\mathbf{A}_{\mathbf{GL}}$ in machine learning models for predicting GRACE-TWSA. ICA has the same process as PCA except that the GL data matrix $\mathbf{A}_{\mathbf{GL}}$ is decomposed as the independent components $\mathbf{S}_{\mathbf{ICA}}$ and its weight matrix instead of Eq. (2). The difference is that $\mathbf{S}_{\mathbf{PCA}}$ tends to explain the global variance with a few elements while $\mathbf{S}_{\mathbf{ICA}}$ focuses on the separation of statistically independent signals that induce the temporal variability of TWSA.

2.3.2. Temporal decomposition

Due to weather variability and anthropogenic activity, GRACE-TWSA and its predictors (i.e., NOAH-P, NOAH-T, NOAH-TWSA, GL derivations from PCA or ICA) exhibit significant two-scale features (Fig. 3) with non-stationary long-term trends and short-term seasonal variations. Hence, we first split the time series of GRACE-TWSA and its predictors into trend and detrended parts using the widely-used seasonal decomposition of time series by loess (STL) (Cleveland et al., 1990):

$$Y_{total} = Y_{trend} + Y_{detrend}, \tag{5}$$

where Y_{total} indicates the original GRACE-TWSA or its predictors, Y_{trend} and $Y_{detrend}$ are decomposed long-term trend and detrended parts, the latter of which includes seasonal signal and residual that has random nature. Note that the enhanced version of the STL method (Hafen, 2016) that can handle data with missing values is used here; thus the trends of the GRACE data before and after the data gap can be estimated simultaneously. Then, the classical linear machine learning method, multiple linear regression (MLR), and the powerful nonlinear one, random forest (RF), are used to train and predict the trend and detrended parts, respectively.

To implement the RF and MLR method, the dataset is randomly shuffled and divided into training (70%) and testing (30%). In addition, the training process is based on fourfold cross-validation of the training data (using 25% of the training data for validation). In the MLR method, we choose the least-squares method to estimate the MLR coefficients. In the RF method, we optimized two parameters, i.e., the number of regression trees (N_{tree}, from 100 to 1000 with intervals of length 100) and the number of input variables per node (N_{input}, based on the two algorithms, i.e., the square root and log base 2). We optimize them based on the mean square error using the training dataset.

2.4. Testing cases

To test our methodological framework and compare the data worth of different inputs, four test cases are designed (Table 1). In case 1, only NOAH derivations (i.e., NOAH-P, NOAH-T, and NOAH-TWSA) are used for the reconstruction as widely done in previous studies (e.g., A. Y. Sun et al., 2019 and Z. Sun et al., 2020), and thus it is regarded as the benchmark case. Case 2 only includes principal components of GL data (i.e., columns of $\widetilde{\mathbf{S}}_{\mathbf{PCA}}$, denoted as GL-PCs) derived from PCA. Cases 3 and 4 integrate both NOAH derivations and GL features from ICA (i.e., columns of $\widetilde{\mathbf{S}}_{\mathbf{ICA}}$, denoted as GL-ICs) and PCA, respectively.

This study considers the hysteresis of hydrological processes and includes lagged inputs up to two-time steps as done by A. Y. Sun et al. (2019) and F. Li et al. (2020). That means the predictors at month t, month t-1, and month t-2 are used to reconstruct GRACE-TWSA at month t. For more implementation details, please refer to Section 3 of the Supplementary material.

Table 1
Summary of cases with different inputs.

Case Number	Inputs
Case 1 (benchmark)	NOAH-P, NOAH-T, NOAH-TWSA
Case 2	GL-PCs
Case 3	NOAH-P, NOAH-T, NOAH-TWSA, GL-ICs
Case 4	NOAH-P, NOAH-T, NOAH-TWSA, GL-PCs

* NOAH-P, NOAH-T, and NOAH-TWSA denote the precipitation, temperature, and terrestrial water storage anomalies derived from the NOAH model, respectively. GL-PCs and GL-ICs represent principal components and independent components from groundwater levels, respectively.

2.5. Evaluation metrics

We apply the root mean square error (*RMSE*) and the Nash-Sutcliffe efficiency (*NSE*) to evaluate the quality of the reconstructed data. The metrics (Moriasi et al., 2015) are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i^t - S_i^a)^2}, \tag{6}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(S_i^t - S_i^a)^2}{\sum_{i=1}^{n}(S_i^t - \overline{S^t})^2}, \tag{7}$$

where S^t is the GRACE-TWSA observation, and S^a is the corresponding reconstructed value; n is the total number of S; $\overline{S}$ denotes the mean value of S. Under ideal conditions, the best fit between S^t and S^a would yield $RMSE = 0$ and $NSE = 1$.

To investigate the relative contributions of different inputs on the GRACE-TWSA reconstruction, we derive the absolute slope coefficients (β) of MLR based on the normalized inputs for the trend simulation (Montgomery et al., 2021) and the variable's importance (*VI*) of the RF method based on variable importance ranking (Breiman, 2001). A larger β or *VI* indicates a more important predictor in machine learning.

In addition, since ENSO and TWSA have a strong connection globally (P. Li et al., 2022; Ni et al., 2018), their cross-correlation (*CORR*) can be used as a proxy to assess the reconstructed TWSA:

$$CORR = \frac{\sum_{i=1}^{n}(T_i^a - \overline{T^a})(T_i^t - \overline{T^t})}{\sqrt{\sum_{i=1}^{n}(T_i^a - \overline{T^a})^2}\sqrt{\sum_{i=1}^{n}(T_i^t - \overline{T^t})^2}}, \tag{8}$$

where T^t is the ENSO (T. Zhang et al., 2019), and T^a is the corresponding TWSA variation (P. Li et al., 2022); a closer *CORR* to 1 indicates more similarity between them. Due to hysteresis, the links to 0–24 month historical ENSO are also considered.

3. Results

3.1. Features of inputs for reconstruction

The first four GL-PCs (168 ×4), comprising around 85% of the variance in the GL data, which are well extracted from the rest residuals, are selected as the main features of GL. Same as PCA, four GL-ICs (168 ×4) are derived from the GL data via ICA (the eigenvalue spectrum of the GL data via principal component analysis PCA can be seen in Tables S1 in the supplementary material, and the uncertainties of derived GL-ICs can be seen in Fig. S2 in the supplementary material). The derived temporal patterns of the inputs for the GRACE-TWSA reconstruction are presented in Fig. 3(a) ~ (k). All GL-PCs (Fig. 3(a) ~ (d)) are different from each other, and so are GL-ICs (Fig. 3(e) ~ (h)), indicating little redundant information remained. Besides, GL-PCs and GL-ICs have some similar temporal and spatial (Fig. 4(a) ~ (h)) patterns (i.e., GL-PC1 & GL-IC2, GL-PC2 & GL-IC1, GL-PC3 & GL-IC4, and GL-PC4 & GL-IC3). GL-IC2 and GL-IC3 can be corrected via a negative unit coefficient to have positive *CORR* values with GL-PC1 and GL-PC4 (Hyvärinen et al., 2001). Furthermore, we can see that the spatial distributions of weight (v_i) of GL-PC1 (Fig. 4(a)) and GL-IC2 (Fig. 4(f)) are similar to those patterns of GL trends and standard errors (Fig. 1(e) and (f)), indicating that both GL-PC1 and GL-IC2 characterize the main temporal features of groundwater level. However, GL-ICs have more obvious seasonal features than GL-PCs (e.g., GL-IC1 & GL-PC2) due to the

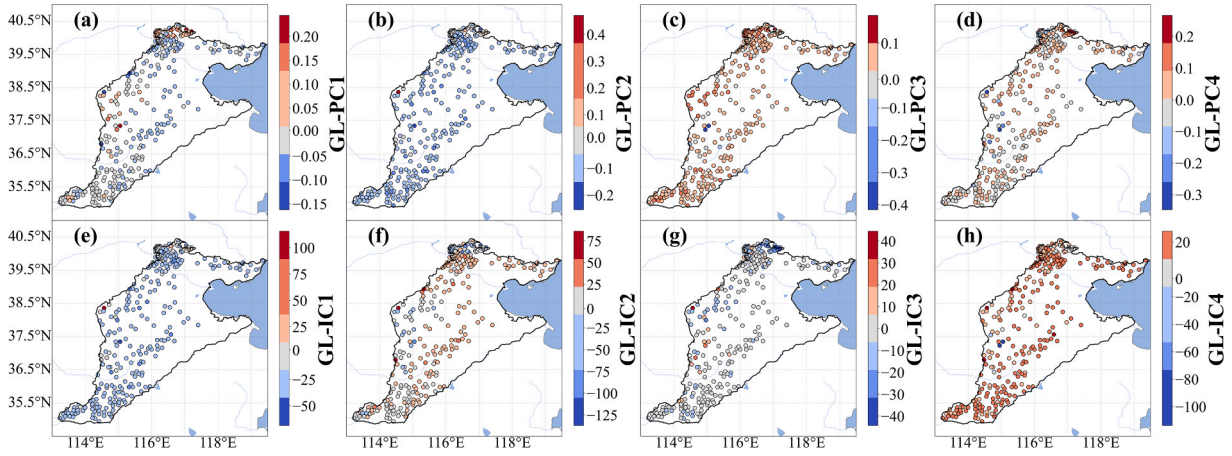


Fig. 4. Spatial distributions of point-scale weights of principal components (GL-PCs) ((a)~(d)) and independent components (GL-ICs) ((e)~(h)) derived from multi-site groundwater levels. The color bars are used to characterize the dimensionless spatial weight values (v_{ik}) in Eq. (4).

nature differences between PCA and ICA.

As for the NOAH derivations (Fig. 3(i) ~ (k), NOAH-P, NOAH-T, and NOAH-TWSA), they do not exhibit a significantly decreasing trend as shown in GRACE-TWSA (Fig. 1(b)), indicating that the NOAH derivations, missing some information (i.e., groundwater storage), are incapable of characterizing GRACE-TWSA fully.

3.2. Reconstruction performances

The statistical results of different cases are shown in Fig. 5. It is found that the trend component performances via the MLR method during training and test periods have smaller differences than those of detrended components via the RF method. The results exhibit great stability of the MLR method. Based on the comparisons between case 1 (using only NOAH derivations) and case 2 (using only GL derivations), the GL derivations are more informative for the trend simulations (please see performances of case 2 and case 1 during the test period in Fig. 5(a) with RMSE values from 1.87 to 5.52 cm and Fig. 5(b) with NSE values from 0.941 to 0.602), while the effects of NOAH derivations are more visible on the detrended simulations (please see performances of case 1 and case 2 during the test period in Fig. 5(c) with RMSE values from 3.77 to 4.02 cm and Fig. 5(d) with NSE values from 0.244 to 0.191). Nevertheless, case 1 (using only NOAH derivations) shows an advantage in characterizing the TWSA trend features compared to the case that does not simulate the trend and detrended parts separately, as shown in Section 4.2 (please see performances of case 1 and case 1′ in Fig. 11(b) with RMSE values from 0.558 to 1.221 cm and NSE values from 0.919 to 0.764). In general, case 3 and case 4 using both kinds of data outperformed case 1 using only NOAH derivations and case 2 using only GL derivations, demonstrating the significant data worth of both NOAH derivations and GL derivations. Case 4 (i.e., using GL-PCs) outperforms case 3 (i.e., using GL-ICs) based on different GL data preprocessing methods, indicating that PCA extracting the main features of data is more appropriate for the GRACE-TWSA reconstruction than ICA, which is also implicitly observed by F. Li et al. (2021).

The spatial distributions of two different statistical (i.e., *RMSE* and *NSE*) metrics of results are shown in Fig. 6 and Fig. 7, respectively. In general, from case 1 (using only NOAH derivations), case 2 (using only GL derivations) to case 3 and case 4 (using both NOAH and GL derivations), the spatial variability of metrics for the trend simulation gradually decreases (also refer to the distribution widths in Fig. 5(a) ~ (b)), underscoring benefits of the multi-source data utilization. Note that the trends of GRACE-TWSA in the north

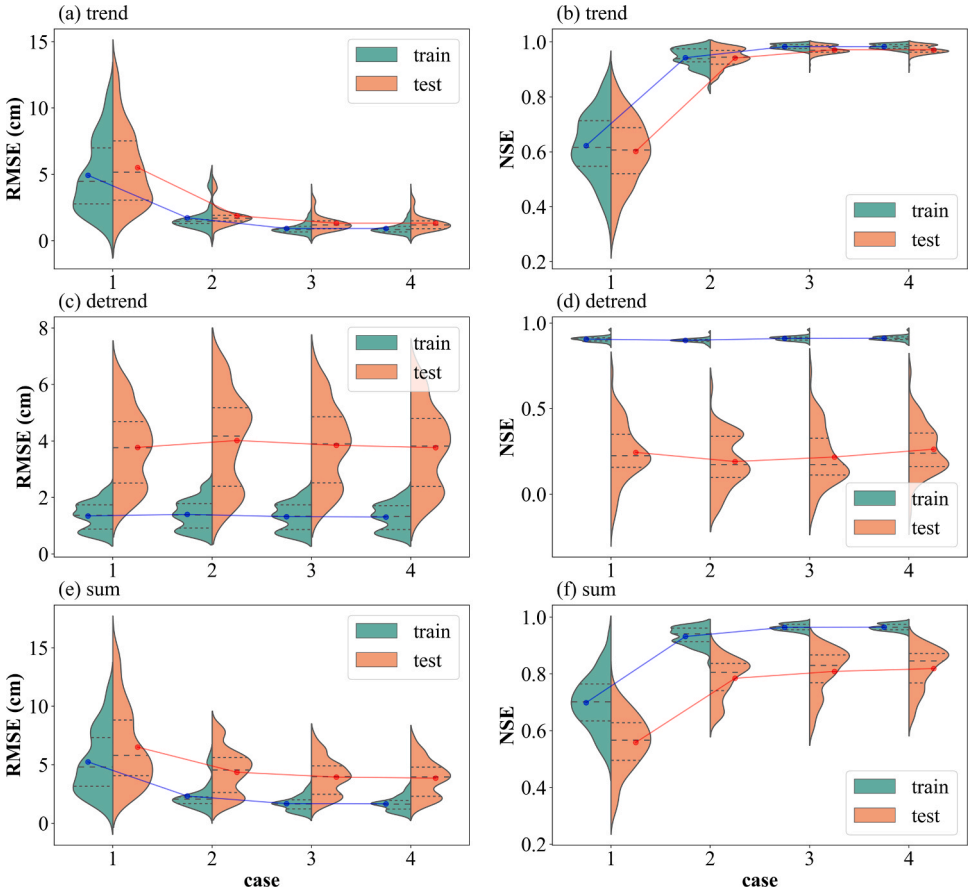


Fig. 5. The violin plots for the performances of the reconstructed trend ((a)~(b)), detrended ((c)~(d)) components of GRACE-TWSA and their sum ((e)~(f)) in cases 1–4. The blue and red parts represent the training and testing phases, respectively.

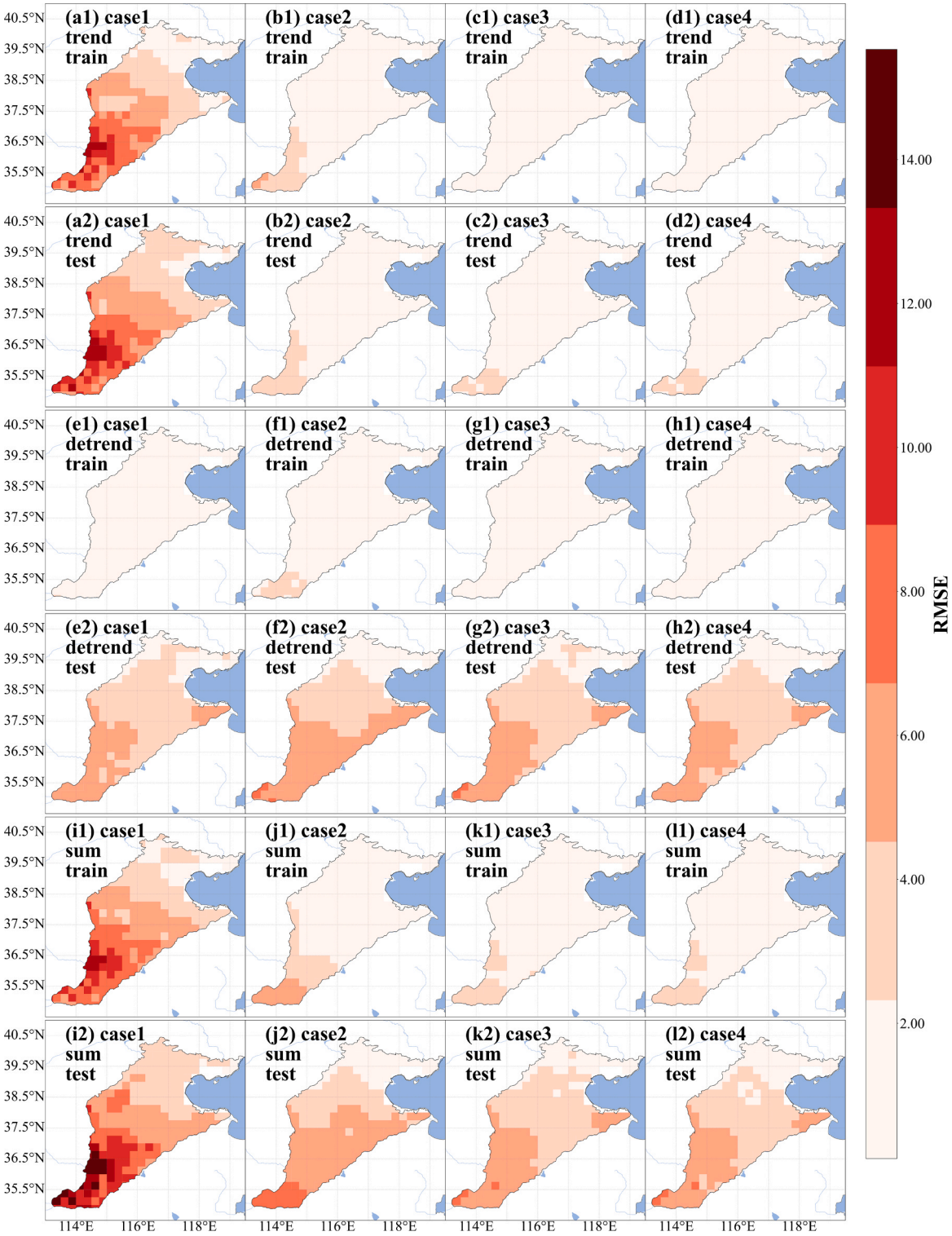


Fig. 6. The spatial distributions of *RMSE* values of reconstructed GRACE-TWSA data in cases 1–4.

are relatively less than that in the south (Fig. 1(a)). Since *RMSE* can be affected by the magnitude, and *NSE* gives more weight to the peak value (McCuen et al., 2006; Waseem et al., 2017), to weaken the effects from magnitudes and evaluate the long-term model simulations (Moriasi et al., 2015), we focus on results based on *NSE* (Fig. 7). In cases using both NOAH and GL derivations (i.e., cases 3 and 4) for the final trend and detrended components summed results (Fig. 7(k) ~ (l)), the performances in the northeast and coastal

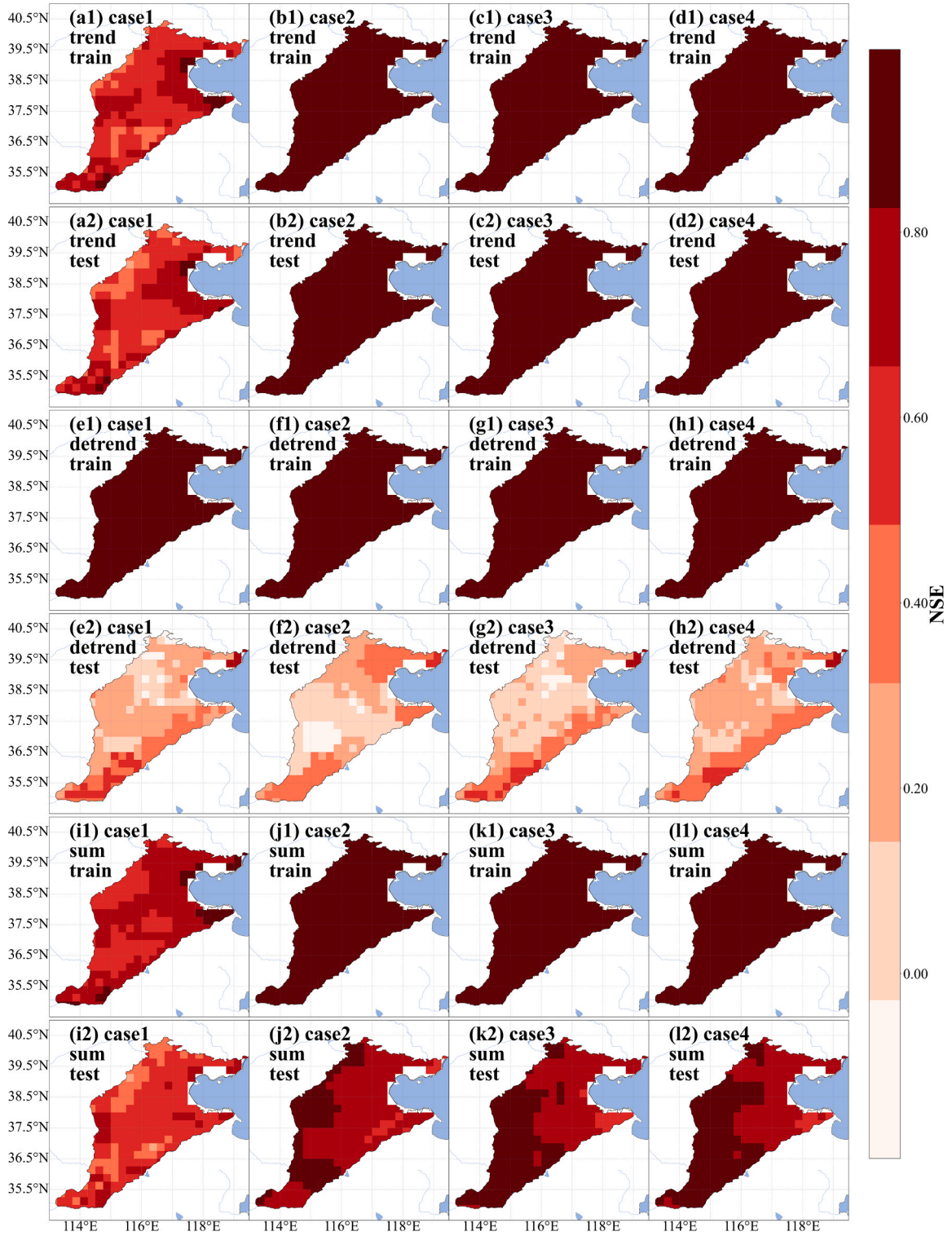


Fig. 7. The spatial distributions of *NSE* values of reconstructed GRACE-TWSA data in cases 1–4.

regions are generally worse than those in the southwest. One possible reason may be the missing anthropogenic influence on surface water storage (e.g., water diversions, such as SNWD, which are particularly significant in Beijing and Tianjin). Another reason for the relatively poor performance in the coastal region is potentially due to the sparseness of the groundwater observation wells along the coastal lines (Fig. 1(a)), and the groundwater dynamics along the coastal lines are different from inland ones due to tidal influence and

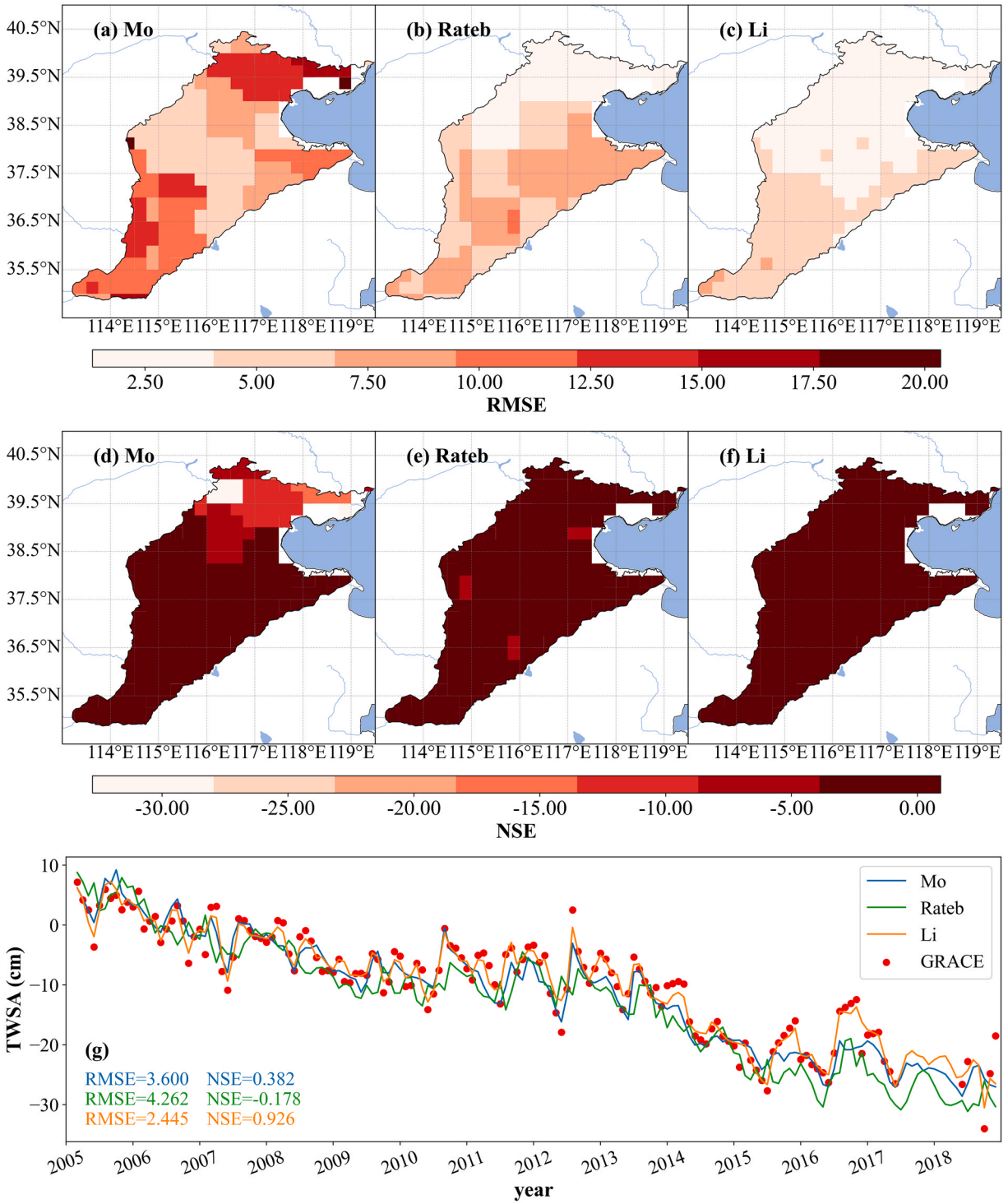


Fig. 8. Comparisons between spatial patterns of reconstructed GRACE-TWSA obtained by (a) and (d) Mo et al. (2022), (b) and (e) Rateb et al. (2022), and (c) and (f) this study as well as (g) their temporal comparisons. The *RMSE* and *NSE* values are calculated based on the reconstructed values and the CSR GRACE-TWSA during their respective test periods.

human activity differences (e.g., less irrigation).

The snapshots of the reconstructed GRACE-TWSA data via case 4 and the corresponding GRACE-TWSA observations during the test are also exhibited in Figs. S3 and S4 of the Supplementary Material, which provides an intuitive way to demonstrate the great consistency of the raw and reconstructed GRACE-TWSA data.

3.3. Comparisons to previous studies and uncertainty analyses

We have compared our results with the previous reconstructions by Mo et al. (2022) and Rateb et al. (2022) as shown in Fig. 8. These datasets with different spatial grids size are mapped into the consistent grid size as our reconstructed GRACE-TWSAs. The *RMSE* and *NSE* values are calculated based on the reconstructed values and GRACE-TWSA during their respective test periods. The results show that our reconstructed data have some advantages over the results from the two previous studies on both spatial and temporal patterns. As for the two competing products, the results provided by Rateb et al. (2022) outperform the results via Mo et al. (2022) on the spatial pattern, while the results via Mo et al. (2022) have the better temporal performances than the results via Rateb et al. (2022).

To conduct the uncertainty analyses, we have collected additional two sets of precipitation and temperature from other data sources, i.e., Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2) and ERA5-Land. Therefore, three sets of precipitation and temperature from different data sources are used to reconstruct the GRACE-TWSA, respectively. The standard deviations of three sets of reconstructed GRACE-TWSA are shown in Fig. 9. The results show that uncertainties in input variables caused by different data sources do affect the reconstructed GRACE-TWSA. However, these uncertainties are much lower than the deviations between the reconstructed and original GRACE-TWSA values (~0.3 cm versus 1.597 cm as shown in Fig. 8).

4. Discussion

4.1. Importance of different predictors

The relative importance of each predictor in different processes is analyzed and shown in Fig. 10. It exhibits that GL-ICs (Fig. 10(a)) and GL-PCs (Fig. 10(c)), especially for GL-IC2 and GL-PC1 which indicate the GL main trend changes, are the most important predictors for the trend component simulations. In addition, GL derivations with one lag are the most important predictors among the GL

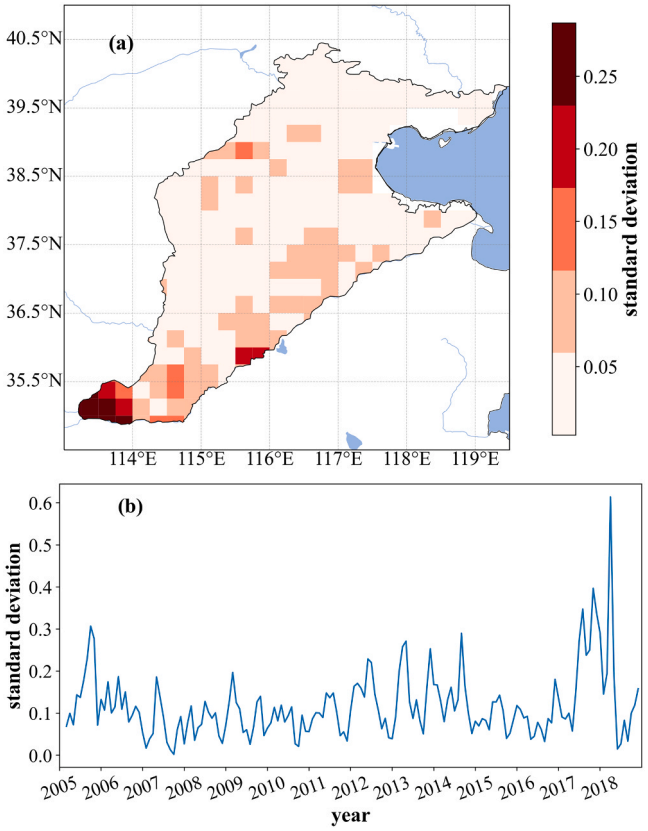


Fig. 9. The (a) spatial and (b) temporal patterns of standard deviations of three sets of reconstructed GRACE-TWSA based on precipitation and temperature from different data sources, i.e., GLDAS NOAH, MERRA-2, and ERA5-land.

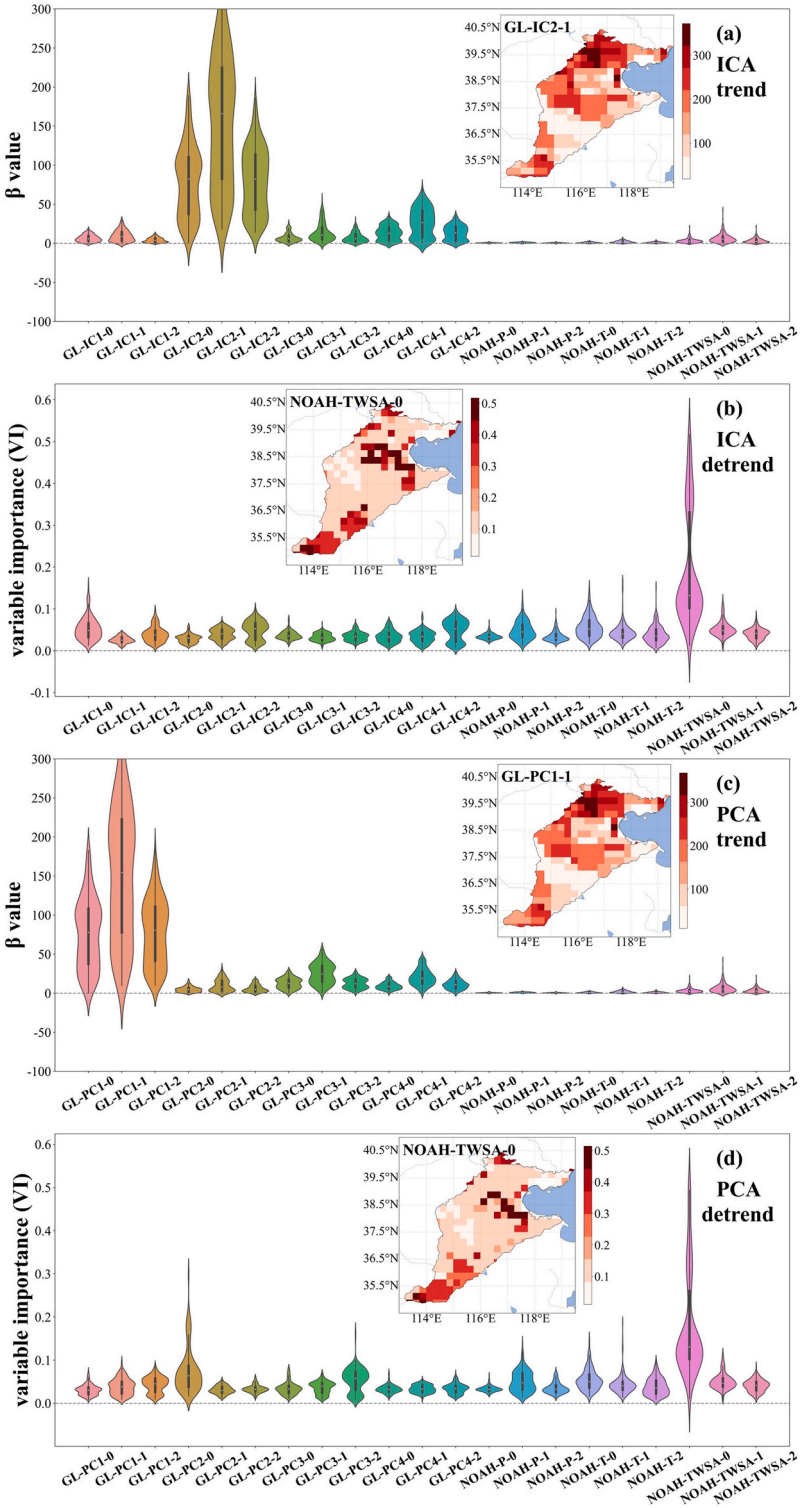


Fig. 10. The β value distributions of the trend components of GRACE-TWSA in case 3 (a) and case 4 (c). The variable importance (*VI*) value distributions of the detrended components of GRACE-TWSA in case 3 (b) and case 4 (d). The larger β and *VI* values of one predictor indicate that it is more important to the model prediction. The spatial distributions of the corresponding first typical predictor (with the highest β and *VI* values) in each subplot are also shown. Numbers at the right of each predictor denote the lags used for the variable, where number 0 denotes zero lag.

derivations with three different lags. It denotes that the trend changes of TWSA lag behind the trend changes of groundwater storage by a month, which is reasonable since the groundwater withdrawal is first made to the surface when TWSA does not change, and then this part of the water is consumed to make TWSA decrease in NCP (P. Li et al., 2021). Spatial distributions of β values of GL-IC2–1 and GL-PC1–1 shown in Fig. 10(a) and (c) show the high values in the northwest and southwest of NCP, which is consistent with that the urban water in the northwest and agricultural water in the southwest rely heavily on groundwater.

As for the detrended component simulation, the average values of the *VI* values of NOAH-TWSA decrease from Fig. 10(b) to Fig. 10 (d), while the average values of the *VI* values of GL derivations increase. The spatial *VI* distributions of NOAH-TWSA are similar from ICA to PCA, but the number of high values decreases in some areas where significant fluctuations of groundwater storage happen (e.g., the northwest and the southwest of NCP). These indicate the superiority of PCA for mining information from the GL data to reconstruct GRACE-TWSA.

4.2. Feature of reconstructed TWSA

To evaluate the effectiveness of additional trend simulation and anthropogenic fluctuations revealed in reconstructed TWSA (Fig. 2), an additional case 1′, which is similar to case 1 in Table 1 but directly simulates without separating trend and detrended parts, is conducted to compare with case 1 (using only NOAH derivations) and case 4 (using both GL and NOAH derivations) in Table 1.

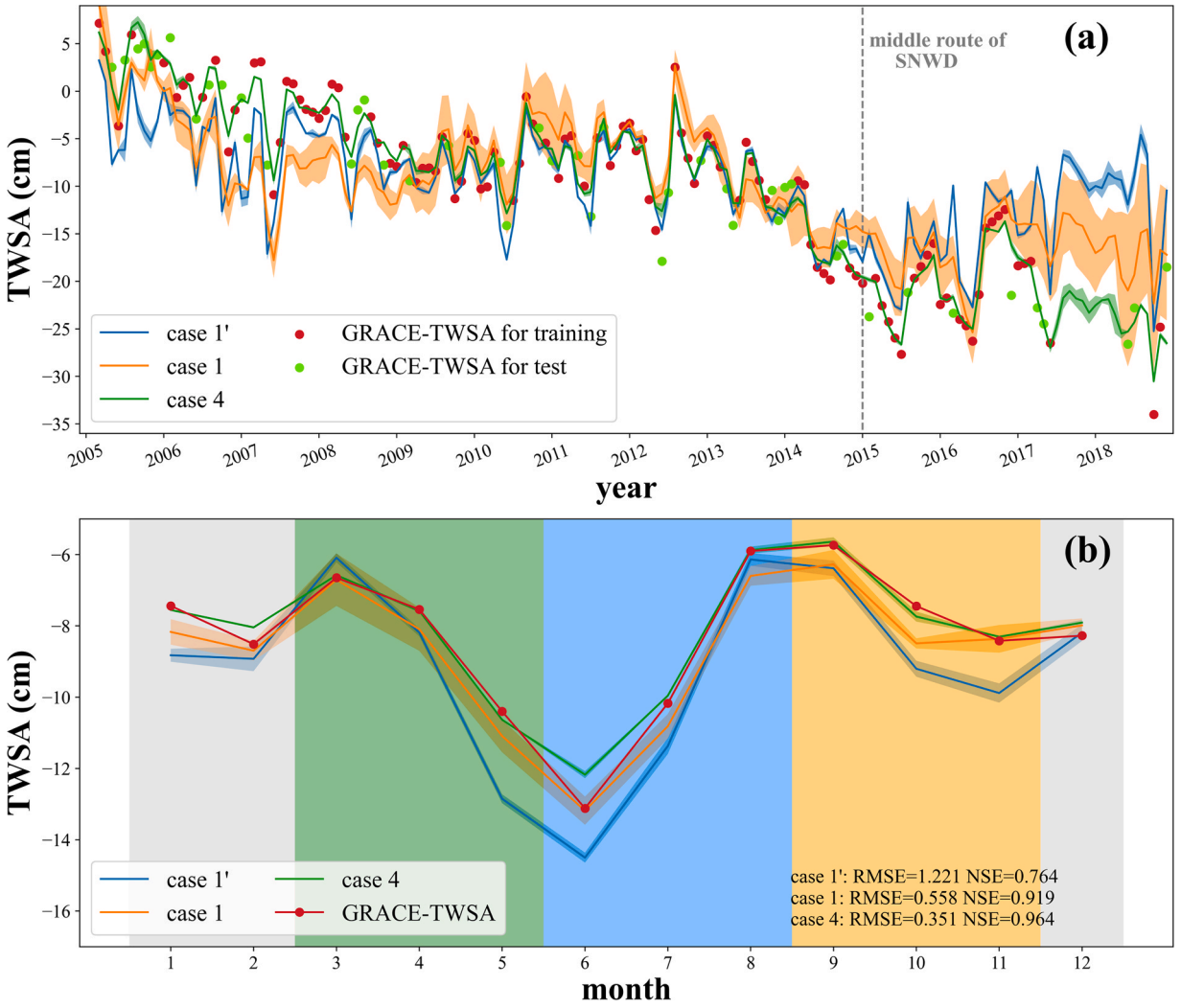


Fig. 11. (a) Comparison between spatial averages of GRACE-TWSA and the reconstructed GRACE-TWSAs via cases 1′, 1, and 4. Case 1′ is similar to case 1 in Table 1 but does not simulate trend and detrended components separately. The vertical dash line in (a) denotes the commencement of operation of the middle route of the South-to-North Water Diversion (SNWD). (b) Comparison between multi-year monthly variations of GRACE-TWSA and the reconstructed GRACE-TWSAs via different scenarios during 2005-2016. Spring, summer, autumn, and winter are highlighted with green, blue, orange, and gray background colors, respectively. The shadow denotes three standard deviations of different GRACE-TWSA reconstructions via different sets of precipitation and temperature datasets.

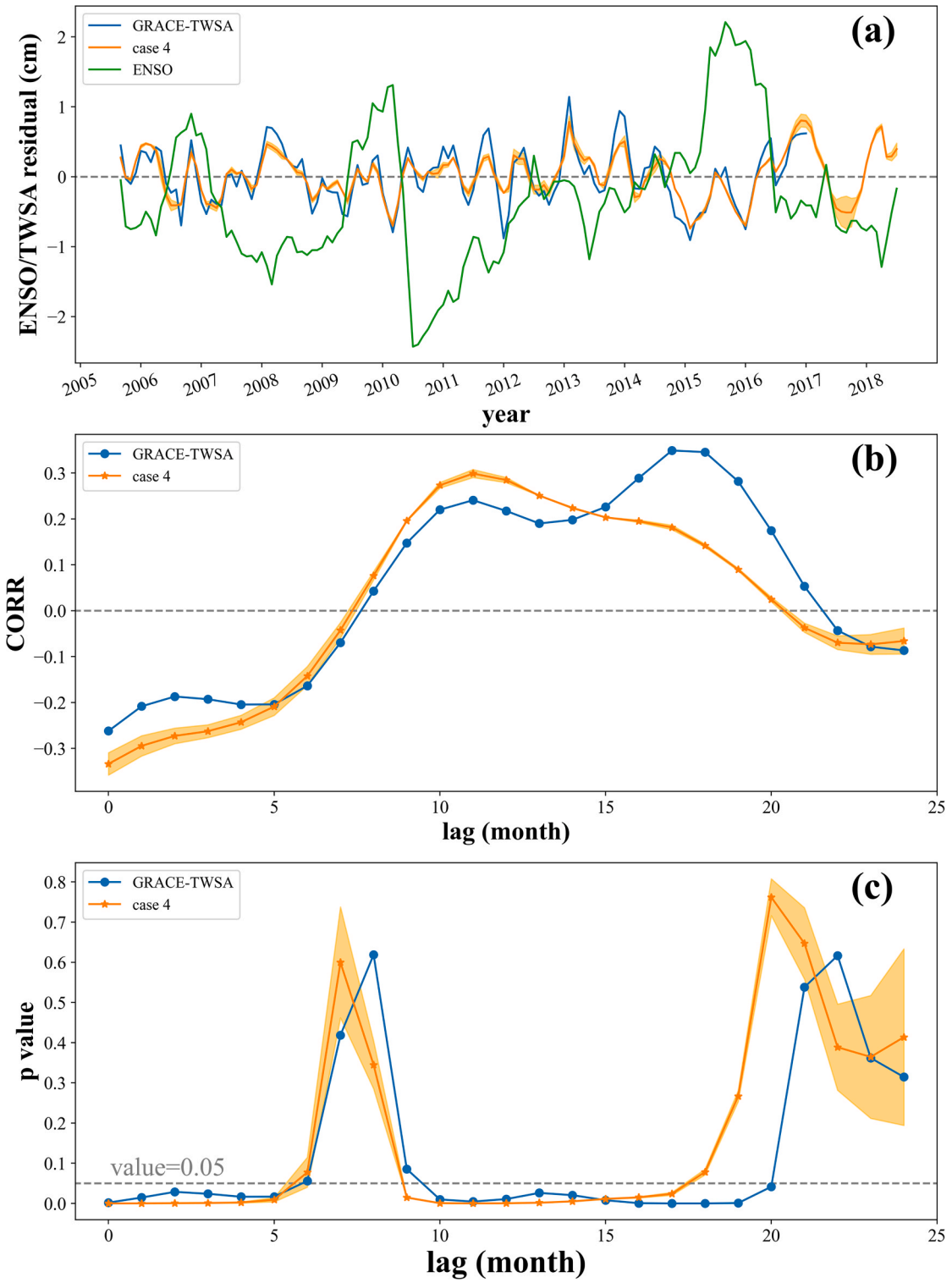


Fig. 12. The relationships between the El Niño-Southern Oscillation (ENSO) and corresponding variations of spatial averages of GRACE-TWSA and the reconstructed GRACE-TWSA in case 4 (case 4): (a) their temporal changes; (b) the correlation coefficient (*CORR*) v.s. different lags; (c) the corresponding *p* values. The shadow denotes three standard deviations based on different GRACE-TWSA reconstructions with different sets of precipitation and temperature datasets.

Fig. 11(a) shows that case 1 characterizes the long-term trend better than case 1′ due to the additional trend simulation, but it has some biases in some local trend simulations even during training since the information on groundwater storage is missing. Such biases are expected to be magnified during the pre-GRACE era. Compared with cases 1′ and 1, case 4 simulates the trend components significantly better and also improves the local fluctuations with the GL data assimilated and additional trend simulation. Also, the trends of reconstructed TWSA data of cases 1 and 4 have been relatively stable since 2015 (C. Zhang et al., 2020), which is widely considered to be an anthropogenic result of the commencement of operation of the middle route of the SNWD in December 2014 (Long et al., 2020).

The multi-year monthly averages of all reconstructed TWSA data in Fig. 11(b) show a significant double-peak seasonal pattern consistent with previous reports (Pan et al., 2017). However, the reconstructed TWSA data in case 1′ has poorer fits against GRACE-TWSA before June when intense groundwater irrigation from March to May (i.e., spring) happens (Pan et al., 2017). Note that the reconstructed TWSA data in case 1′ generally perform better during summer (i.e., from June to August) than in other seasons when intense precipitation and evapotranspiration generally occur (see Fig. 1(c) and (d)). This is reasonable since the reconstructed TWSA data in case 1′ are purely driven by the meteorological data, tending to perform better with insignificant anthropogenic influence. Although the issue without predictors to reflect anthropogenic influence has been mitigated to some extent with additional trend simulation in case 1, case 4 achieves the best performance with the GL data incorporated.

The climatic response in the reconstructed GRACE-TWSA via case 4 is evaluated by investigating its link to ENSO. Fig. 12(a) shows that the derived signals and ENSO without phase differences have a moderate negative correlation, obviously shown in the lag cross-correlation analysis (Fig. 12(b)). Based on it, we can see that the reconstructed TWSA with small lags (< 5 months) in case 4 has higher *CORR* values than the raw GRACE-TWSA, even though the raw GRACE-TWSA data have a shorter data length (in case they tend to have higher *CORR* absolute values). Besides, as for the numbers of lags that correspond to a significant relationship between ENSO and TWSA ($p < 0.05$, Fig. 12(c)), the reconstructed data generally have smaller p values than GRACE-TWSA. The analyses in Fig. 12 successfully exhibit that the reconstructed GRACE-TWSA contains fluctuations affected by climatic factors. Also, the uncertainty analyses (the shadows shown in Fig. 11 and Fig. 12) based on different sets of precipitation and temperature datasets indicate that our results are stable.

4.3. Uncertainties and limitations

Although the reconstructed GRACE-TWSA data in NCP (a semi-arid region with intense groundwater withdrawals) with the GL data included via our proposed framework perform well, the relatively short temporal length and uneven spatial distribution of the collected GL data limit its further assessment and may even bring some adverse effects on the GRACE-TWSA reconstruction. The rapid development of a global public in-situ or reanalysis database of GL data (IGRAC, 2022) is required, which helps extend this study to improve the GRACE-TWSA reconstruction in regions with intense groundwater withdrawals, especially with the semi-arid and arid climate classes, on a global scale for a longer period. It is promising that more data like water diversions are used via our proposed framework to provide more comprehensive information about anthropogenic activities for the reconstruction. It is especially important for some developing countries or remote areas where few or even no historical GL data are recorded. In this case, other independent data that can reflect similar information about human activities may be a potential alternative. For example, electricity consumption (Xu et al., 2018) and global navigation satellite system (R. Liu et al., 2022) measurements can be used to evaluate groundwater storage changes in areas with intense groundwater withdrawals. The effects of different machine learning models and GRACE data from different sources on the reconstruction are not investigated in this study since it is not the focus of our research. Nevertheless, we believe these uncertainties and limitations do not change the comparative performances under different scenarios. The results may be further improved when the GL data is high-quality, a detailed hydrogeological description is available, and some advanced interpolation techniques (e.g., spatiotemporal Kriging (Snepvangers et al., 2003)) are used. However, such data are difficult to obtain, and more labor and time costs will be taken to calibrate such an interpolation model.

5. Conclusions

In this paper, we developed a novel framework to assimilate in-situ GL measurements to improve the GRACE-TWSA reconstruction in regions with intense groundwater withdrawals. The spatial scale inconsistency between the GL data and GRACE-TWSA is handled by feature extraction methods. The trend and detrended components of the time series are separately simulated by linear and nonlinear machine learning approaches, respectively.

Our findings further reveal the strong hydraulic connectivity of multi-site GL observations, and the significant linear correlation between the trend changes of GRACE-TWSA and GL in NCP, with a one-month lag. The case study demonstrates the high value of GL data as predictors for the GRACE-TWSA estimation in regions with semi-arid and arid climates and intense groundwater extraction. Compared to ICA, PCA is more appropriate for extracting the spatiotemporal features of groundwater in NCP to reconstruct the GRACE-TWSA. Also, the results suggest that the trend and detrended components of TWSA can be simulated via linear and nonlinear machine learning methods, respectively. Compared with those without using the in-situ groundwater observations as predictors and previous results, our reconstructed GRACE-TWSA improves in terms of accuracy and properly reflects anthropogenic influences on groundwater and the moderate connection with ENSO, which can be applied to many other studies in NCP.

Our study underscores the benefits of integrating multi-source datasets of regional meteorological forcing data and in situ groundwater levels for the GRACE-TWSA reconstruction, given the scale mismatch issue has been properly treated. In the future, we will attempt to reconstruct global-wide GRACE-TWSA by integrating in-situ GL records all over the world (IGRAC, 2022) before the GRACE launches and during the gap.

CRediT authorship contribution statement

Peijun Li: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization, Investigation. **Yuanyuan Zha:** Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization, Investigation, Funding acquisition, Supervision. **Chak-Hau Michael Tso:** Writing – review &editing.

Declaration of Competing Interest

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

Data Availability

We have shared the link to our data/code in the section of acknowledgments.

Acknowledgments

The GRACE RL06 mascon solutions can be downloaded from the Center for Space Research (http://www2.csr.utexas.edu/grace/); the GLDAS Version 2.1 NOAH 0.25° data are available from the National Aeronautics and Space Administration (https://disc.sci.gsfc.nasa.gov/); the groundwater level data are publicly available on the Data Sharing Repository of National Earth System Science Data Center (https://www.doi.org/10.12041/geodata.10392011028646.ver1.db); the scripts and reconstructed data are available at https://github.com/PercyLeeMaxWell/ReconstructedTWS. This study is supported by the National Key Research & Development Program of China (2019YFC1805701), the Natural Science Foundation of China (52279042), and the Open Research Fund of Guangxi Key Laboratory of Water Engineering Materials and Structures, the Guangxi Institute of Water Resources Research (GXHRI-WEMS-2022-01, GXHRI-WEMS-2022-07). The authors acknowledge Yun Pan and Wenjie Yin for their invaluable assistance in the revision and supervision of this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2023.101528.

References

Adams, K.H., Reager, J.T., Rosen, P., Wiese, D.N., Farr, T.G., Rao, S., et al., 2022. Remote sensing of groundwater: current capabilities and future directions. Water Resour. Res. 58 (10), 1–27. https://doi.org/10.1029/2022WR032219.

Ahmed, M., Aqnouy, M., Stitou El Messari, J., 2021. Sustainability of Morocco's groundwater resources in response to natural and anthropogenic forces. J. Hydrol. 603 (PA), 126866 https://doi.org/10.1016/j.jhydrol.2021.126866.

Becker, M., Meyssignac, B., Xavier, L., Cazenave, A., Alkama, R., Decharme, B., 2011. Past terrestrial water storage (1980-2008) in the Amazon Basin reconstructed from GRACE and in situ river gauging data. Hydrol. Earth Syst. Sci. 15 (2), 533–546. https://doi.org/10.5194/hess-15-533-2011.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1201/9780429469275-8.

Cleveland, R.B., Cleveland, W.S., Mcrae, J.E., 1990. STL: A seasonal-trend decomposition procedure based on loess. J. Stat. 6, 3–73.

Forootan, E., Schumacher, M., Mehrnegar, N., Bezdĕk, A., Talpe, M.J., Farzaneh, S., et al., 2020. An iterative ICA-based reconstruction method to produce consistent time-variable total water storage fields using GRACE and Swarm satellite data. Remote Sens. 12, 1639. https://doi.org/10.3390/rs12101639.

Frappart, F., Ramillien, G., 2018. Monitoring groundwater storage changes using the Gravity Recovery and Climate Experiment (GRACE) satellite mission: A review. Remote Sens. 10, 829. https://doi.org/10.3390/rs10060829.

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., et al., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J. Clim. 30 (14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1.

Guo, L., Li, T., Chen, D., Liu, J., He, B., Zhang, Y., 2021. Links between global terrestrial water storage and large-scale modes of climatic variability. J. Hydrol. 598, 126419 https://doi.org/10.1016/j.jhydrol.2021.126419.

Gyawali, B., Ahmed, M., Murgulet, D., Wiese, D.N., 2022. Filling temporal gaps within and between GRACE and GRACE-FO terrestrial water storage records: an innovative approach. Remote Sens. 14 (7) https://doi.org/10.3390/rs14071565.

Hafen, R. , 2016. stlplus: Enhanced Seasonal Decomposition of Time Series by Loess.

Ho, T.K., 1995. Random decision forests. Proc. 3rd Int. Conf. Doc. Anal. Recognit. 278–282.

Hsiao, C.T., Chang, L.C., Tsai, J.P., Chen, Y.C., 2017. Features of spatiotemporal groundwater head variation using independent component analysis. J. Hydrol. 547, 623–637. https://doi.org/10.1016/j.jhydrol.2017.02.021.

Humphrey, V., Gudmundsson, L., 2019. GRACE-REC: a reconstruction of climate-driven water storage changes over the last century. Earth Syst. Sci. Data Discuss. 11 (3), 1153–1170. https://doi.org/10.5194/essd-2019-25.

Hyvärinen, A., Karhunen, J., Oja, E. , 2001. Independent Component Analysis. Wiley Online Library. ⟨https://doi.org/10.1016/b978–0-444–64165-6.02006–1⟩.

IGRAC, 2022. Global Groundwater Information System. *Published: Ggis.Un-Igrac.Org.*

Jing, W., Zhang, P., Zhao, X., Yang, Y., Jiang, H., Xu, J., et al., 2020. Extending GRACE terrestrial water storage anomalies by combining the random forest regression and a spatially moving window structure. J. Hydrol. 590, 125239 https://doi.org/10.1016/j.jhydrol.2020.125239.

Landerer, F.W., Flechtner, F.M., Save, H., Webb, F.H., Bandikova, T., Bertiger, W.I., et al., 2020. Extending the global mass change data record: GRACE follow-on instrument and science data performance. Geophys. Res. Lett. 47 (12), 1–10. https://doi.org/10.1029/2020GL088306.

Li, F., Kusche, J., Chao, N., Wang, Z., Löcher, A., 2021. Long-term (1979-Present) total water storage anomalies over the global land derived by reconstructing GRACE data. Geophys. Res. Lett. 48 (8), 1–10. https://doi.org/10.1029/2021GL093492.

Li, F., Kusche, J., Rietbroek, R., Wang, Z., Forootan, E., Schulze, K., Lück, C., 2020. 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 https://doi.org/10.1029/2019WR026551.
Li, P., Zha, Y., Shi, L., Zhong, H., 2021. Identification of the terrestrial water storage change features in the North China Plain via independent component analysis. J. Hydrol.: Reg. Stud. 38, 100955 https://doi.org/10.1016/j.ejrh.2021.100955.
Li, P., Zha, Y., Shi, L., Zhong, H., 2022. Assessing the global relationships between teleconnection factors and terrestrial water storage components. Water Resour. Manag. 36, 119–133. https://doi.org/10.1007/s11269-021-03015-x.
Li, P., Zha, Y., Shi, L., Tso, C.H.M., Zhang, Y., Zeng, W., 2020. Comparison of the use of a physical-based model with data assimilation and machine learning methods for simulating soil water dynamics. J. Hydrol. 584, 124692 https://doi.org/10.1016/j.jhydrol.2020.124692.
Liu, B., Zou, X., Yi, S., Sneeuw, N., Li, J., Cai, J., 2022. Reconstructing GRACE-like time series of high mountain glacier mass anomalies. Remote Sens. Environ. 280 (April), 113177 https://doi.org/10.1016/j.rse.2022.113177.
Liu, R., Zhong, B., Li, X., Zheng, K., Liang, H., Cao, J., et al., 2022. Analysis of groundwater changes (2003–2020) in the North China plain using geodetic measurements. J. Hydrol.: Reg. Stud. 41 (December 2021), 101085 https://doi.org/10.1016/j.ejrh.2022.101085.
Long, D., Shen, Y., Sun, A., Hong, Y., Longuevergne, L., Yang, Y., et al., 2014. Drought and flood monitoring for a large karst plateau in Southwest China using extended GRACE data. Remote Sens. Environ. 155, 145–160. https://doi.org/10.1016/j.rse.2014.08.006.
Long, D., Yang, W., Scanlon, B.R., You, L., Wada, Y., Zhao, J., et al., 2020. South-to-north water diversion stabilizing Beijing's groundwater levels. Nat. Commun. 11, 3665 https://doi.org/10.1038/s41467-020-17428-6.
McCuen, R.H., Knight, Z., Cutter, A.G., 2006. Evaluation of the Nash–Sutcliffe efficiency index. J. Hydrol. Eng. 11, 597–602. https://doi.org/10.1061/(ASCE)1084-0699(2006)11.
Mo, S., Zhong, Y., Forootan, E., Mehrnegar, N., Yin, X., Wu, J., et al., 2022. Bayesian convolutional neural networks for predicting the terrestrial water storage anomalies during GRACE and GRACE-FO gap. J. Hydrol. 604, 127244 https://doi.org/10.1016/j.jhydrol.2021.127244.
Mo, S., Zhong, Y., Forootan, E., Shi, X., Feng, W., Yin, X., Wu, J., 2022. Hydrological droughts of 2017-2018 explained by the Bayesian reconstruction of GRACE(-FO) fields. Water Resour. Res., e2022WR031997 https://doi.org/10.1029/2022WR031997.
Montgomery, D.C., Peck, E.A., Vining, G.G., 2021. Introduction to Linear Regression Analysis. John Wiley & Sons.
Moriasi, D.N., Gitau, M.W., Pai, N., Daggupati, P., 2015. Hydrologic and water quality models: performance measures and evaluation criteria. Trans. ASABE 58 (6), 1763–1785. https://doi.org/10.13031/trans.58.10715.
Mukherjee, A., Scanlon, B.R., Aureli, A., Langan, S., Guo, H., McKenzie, A.A., 2021. Global Groundwater: Source, Scarcity, Sustainability, Security, And Solutions. Elsevier. https://doi.org/10.1016/c2018-0-03156-4.
Muñoz Sabater, J, 2019. ERA5-Land hourly data from 1981 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)* , (Accessed on < 30–08-2022 >), ⟨10.24381/cds.e2161ba⟩.
Ni, S., Chen, J., Wilson, C.R., Li, J., Hu, X., Fu, R., 2018. Global terrestrial water storage changes and connections to ENSO events. Surv. Geophys. 39 (1), 1–22. https://doi.org/10.1007/s10712-017-9421-7.
Pan, Y., Zhang, C., Gong, H., Yeh, P.J.F., Shen, Y., Guo, Y., et al., 2017. Detection of human-induced evapotranspiration using GRACE satellite observations in the Haihe River basin of China. Geophys. Res. Lett. 44 (1), 190–199. https://doi.org/10.1002/2016GL071287.
Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Lond., Edinb., Dublin Philos. Mag. J. Sci. 2 (11), 559–572. https://doi.org/10.1080/14786440109462720.
Phillips, T., Nerem, R.S., Fox-Kemper, B., Famiglietti, J.S., Rajagopalan, B., 2012. The influence of ENSO on global terrestrial water storage using GRACE. Geophys. Res. Lett. 39 (16), 1–7. https://doi.org/10.1029/2012GL052495.
Rateb, A., Sun, A., Scanlon, B.R., Save, H., Hasan, E., 2022. Reconstruction of GRACE mass change time series using a bayesian framework. Earth Space Sci. 9 (7) https://doi.org/10.1029/2021EA002162.
Rodell, M., Houser, P.R., Berg, A.A., Famiglietti, J.S., 2005. Evaluation of 10 methods for initializing a land surface model. J. Hydrometeorol. 6 (2), 146–155. https://doi.org/10.1175/JHM414.1.
Rui, H.L., & Beaudoing, H.2020. README Document for NASA GLDAS Version 2 Data Products. *Goddard Earth Sciences Data and Information Services Center (GES DISC)* , *16*(1), 1–32.
Sakura, Y., Tang, C., Yoshioka, R., Ishibashi, H. , 2003. *Intensive use of groundwater in some areas of China and Japan. Intensive Use of Groundwater: Challenges and Opportunities*.
Save, H., Bettadpur, S., Tapley, B.D., 2016. High-resolution CSR GRACE RL05 mascons. J. Geophys. Res.: Solid Earth 121, 7547–7569. https://doi.org/10.1002/2015JB012608.Received.
Scanlon, B.R., Zhang, Z., Save, H., Wiese, D.N., Landerer, F.W., Long, D., et al., 2016. Global evaluation of new GRACE mascon products for hydrologic applications. Water Resour. Res. 12, 9412–9429. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x.
Scanlon, B.R., Zhang, Z., Save, H., Sun, A.Y., Schmied, H.M., Van Beek, L.P.H., et al., 2018. Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. Proc. Natl. Acad. Sci. USA 115 (6), E1080–E1089. https://doi.org/10.1073/pnas.1704665115.
Shen, C., Appling, A.P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., et al., 2023. Differentiable modelling to unify machine learning and physical models for geosciences. Nature Reviews Earth & Environment 4 (8), 552–567. https://doi.org/10.1038/s43017-023-00450-9.
Snepvangers, J.J.J.C., Heuvelink, G.B.M., Huisman, J.A., 2003. Soil water content interpolation using spatio-temporal kriging with external drift. Geoderma 112 (3–4), 253–271. https://doi.org/10.1016/S0016-7061(02)00310-5.
Sun, A.Y., Scanlon, B.R., Zhang, Z., Walling, D., Bhanja, S.N., Mukherjee, A., Zhong, Z., 2019. Combining physically based modeling and deep learning for fusing GRACE satellite data: can we learn from mismatch? Water Resour. Res. 55 (2), 1179–1195. https://doi.org/10.1029/2018WR023333.
Sun, Z., Long, D., Yang, W., Li, X., Pan, Y., 2020. Reconstruction of GRACE data on changes in total water storage over the global land surface and 60 basins. Water Resour. Res. 56 (4), 1–21. https://doi.org/10.1029/2019WR026250.
Tapley, B.D., Bettadpur, S., Ries, J.C., Thompson, P.F., Watkins, M.M., 2004. GRACE measurements of mass variability in the Earth system. Science 305 (5683), 503–505. https://doi.org/10.1126/science.1099192.
Tapley, B.D., Watkins, M.M., Flechtner, F., Reigber, C., Bettadpur, S., Rodell, M., et al., 2019. Contributions of GRACE to understanding climate change. Nat. Clim. Change 9 (5), 358–369. https://doi.org/10.1038/s41558-019-0456-2.
Trabucco, A., & Zomer, R. (2019). Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2. *Figshare. Dataset.* ⟨Https://Doi.Org/10.6084/M9.Figshare.7504448.V3⟩.
Tsai, J.P., Hsiao, C.T., 2020. Spatiotemporal analysis of the groundwater head variation caused by natural stimuli using independent component analysis and continuous wavelet transform. J. Hydrol. 590 (August), 125405 https://doi.org/10.1016/j.jhydrol.2020.125405.
van Zoest, V., Osei, F.B., Hoek, G., Stein, A., 2020. Spatio-temporal regression kriging for modelling urban NO2 concentrations. Int. J. Geogr. Inf. Sci. 34 (5), 851–865. https://doi.org/10.1080/13658816.2019.1667501.
Varouchakis, E.A., Theodoridou, P.G., Karatzas, G.P., 2019. Spatiotemporal geostatistical modeling of groundwater levels under a Bayesian framework using means of physical background. J. Hydrol. 575 (May), 487–498. https://doi.org/10.1016/j.jhydrol.2019.05.055.
Wang, F., Shen, Y., Chen, Q., Wang, W., 2021. Bridging the gap between GRACE and GRACE follow-on monthly gravity field solutions using improved multichannel singular spectrum analysis. J. Hydrol. 594, 125972 https://doi.org/10.1016/j.jhydrol.2021.125972.
Waseem, M., Mani, N., Andiego, G., Usman, M., 2017. A review of criteria of fit for hydrological models. Int. Res. J. Eng. Technol. (IRJET) 4, 1765–1772.
Xiang, L., Wang, H., Steffen, H., Qiao, B., Feng, W., Jia, L., Gao, P., 2022. Determination of weak terrestrial water storage changes from GRACE in the interior of the Tibetan plateau. Remote Sens. 14 (3) https://doi.org/10.3390/rs14030544.
Xu, T., Yan, D., Weng, B., Bi, W., Do, P., Liu, F., et al., 2018. The effect evaluation of comprehensive treatment for groundwater overdraft in Quzhou County, China. Water 10 (7), 1–18. https://doi.org/10.3390/w10070874.

Yang, T., Wang, C., Yu, Z., Xu, F., 2013. Characterization of spatio-temporal patterns for various GRACE- and GLDAS-born estimates for changes of global terrestrial water storage. Glob. Planet. Change 109, 30–37. https://doi.org/10.1016/j.gloplacha.2013.07.005.

Yang, X., Tian, S., You, W., Jiang, Z., 2021. Reconstruction of continuous GRACE/GRACE-FO terrestrial water storage anomalies based on time series decomposition. J. Hydrol. 603, 127018 https://doi.org/10.1016/j.jhydrol.2021.127018.

Yi, S., Sneeuw, N., 2021. Filling the data gaps within GRACE missions using singular spectrum analysis. J. Geophys. Res.: Solid Earth 126 (5), 1–22. https://doi.org/10.1029/2020JB021227.

Zhang, C., Duan, Q., Yeh, P.J. -F., Pan, Y., Gong, H., Gong, W., et al., 2020. The effectiveness of the South-to-North water diversion middle route project on water delivery and groundwater recovery in North China Plain. Water Resour. Res. 56, e2019WR026759 https://doi.org/10.1029/2019wr026759.

Zhang, C., Duan, Q., J.-F. Yeh, P., Pan, Y., Gong, H., Moradkhani, H., et al., 2021. Sub-regional groundwater storage recovery in North China Plain after the South-to-North water diversion project. J. Hydrol. 597, 126156 https://doi.org/10.1016/j.jhydrol.2021.126156.

Zhang, T., Hoell, A., Perlwitz, J., Eischeid, J., Murray, D., Hoerling, M., Hamill, T.M., 2019. Towards probabilistic multivariate ENSO monitoring. Geophys. Res. Lett. 46 (17–18), 10532–10540. https://doi.org/10.1029/2019GL083946.

Zhang, X., Li, J., Dong, Q., Wang, Z., Zhang, H., Liu, X., 2022. Bridging the gap between GRACE and GRACE-FO using a hydrological model. Sci. Total Environ. 822, 153659 https://doi.org/10.1016/j.scitotenv.2022.153659.

Zhang, Z., Fei, Y., Chen, Z., Zhao, Z. , 2009. Survey and Evaluation of Ground-Water Sustainable Utilization in North China Plain (in Chinese). Geological House, Beijing.

Zhang, Zizhan, Chao, B.F., Chen, J., Wilson, C.R., 2015. Terrestrial water storage anomalies of yangtze river basin droughts observed by GRACE and connections with ENSO. Glob. Planet. Change 126, 35–45. https://doi.org/10.1016/j.gloplacha.2015.01.002.

Zhao, Q., Zhang, B., Yao, Y., Wu, W., Meng, G., Chen, Q., 2019. Geodetic and hydrological measurements reveal the recent acceleration of groundwater depletion in North China Plain. J. Hydrol. 575, 1065–1072. https://doi.org/10.1016/j.jhydrol.2019.06.016.