

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/131753/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Li, Fupeng, Kusche, Jürgen, Rietbroek, Roelof, Wang, Zhengtao, Forootan, Ehsan , Schulze, Kerstin and Lück, Christina 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 Resources Research* 56 (5) , e2019WR026551. 10.1029/2019WR026551

Publishers page: <http://dx.doi.org/10.1029/2019WR026551>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Li Fupeng (Orcid ID: 0000-0002-3329-1989)
Rietbroek Roelof (Orcid ID: 0000-0001-5276-5943)
Wang Zhengtao (Orcid ID: 0000-0002-3858-0640)
Forootan Ehsan (Orcid ID: 0000-0003-3055-041X)
Schulze Kerstin (Orcid ID: 0000-0002-7300-565X)

Comparison of Data-driven Techniques to Reconstruct (1992-2002) and Predict (2017-2018) GRACE-like Gridded Total Water Storage Changes using Climate Inputs

Fupeng Li^{1,2*}, Jürgen Kusche², Roelof Rietbroek², Zhengtao Wang¹, Ehsan Forootan^{3,4}, Kerstin Schulze², Christina Lück²

¹ School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China

² Institute of Geodesy and Geoinformation, University of Bonn, Bonn 53115, Germany

³ Robert Bosch GmbH, Abstatt, Germany

⁴ School of Earth and Ocean Sciences, Cardiff University, Cardiff CF10 3AT, UK

*Corresponding author: Fupeng Li (fpli@whu.edu.cn).

Key points:

- We develop a unified methodological framework to compare viable approaches for reconstructing and predicting globally gridded GRACE fields.
- Predicted total water storage change fields fit better to the observations than those simulated by hydrological models.
- Reconstructed total water storage change correctly reproduces a strong anomalous signal in the tropical river basin during the El Niño periods.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019WR026551

Abstract: The Gravity Recovery and Climate Experiment (GRACE) mission ended its operation in October 2017 and the GRACE Follow-On mission was launched only in May 2018, leading to approximately one year of data gap. Given that GRACE-type observations are exclusively providing direct estimates of Total Water Storage Change (TWSC), it would be very important to bridge the gap between these two missions. Furthermore, for many climate-related applications, it is also desirable to reconstruct TWSC prior to the GRACE period. In this study, we aim at comparing different data-driven methods and identifying the more robust alternatives for predicting GRACE-like gridded TWSC during the gap and reconstructing them to 1992 using climate inputs. To this end, we first develop a methodological framework to compare different methods such as the Multiple Linear Regression (MLR), Artificial Neural Network (ANN), and Autoregressive eXogenous (ARX) approaches. Second, metrics are developed to measure the robustness of the predictions. Finally, gridded TWSC within twenty-six regions are predicted and reconstructed using the identified methods. Test computations suggest that the correlation of predicted TWSC maps with observed ones is more than 0.3 higher than TWSC simulated by hydrological models, at the grid scale of 1° resolution. Furthermore, the reconstructed TWSC correctly reproduce the El Niño-Southern Oscillation (ENSO) signals. In general, while MLR does not perform best in the training process, it is more robust and could thus be a viable approach both for filling the GRACE gap and for reconstructing long-period TWSC fields globally when combined with statistical decomposition techniques.

1. Introduction

The Gravity Recovery and Climate Experiment (GRACE) mission, launched by the National Aeronautical and Spatial Administration (NASA) and the German Aerospace Centre (DLR) and flown from March 2002 to October 2017, was dedicated to observe temporal changes in the Earth's gravity field (Tapley et al., 2004). Changes in gravity detected by GRACE can be used to derive estimates of Total Water Storage Change (TWSC) (Syed et al., 2008; Landerer & Swenson, 2012), for hydrology studies (Chandan & Nagesh, 2018), drought or flood detection (Yirdaw et al., 2008; Chen et al., 2009; Leblanc et al., 2009; Frappart et al., 2012; Long et al., 2014; Thomas et al., 2014; Chao et al., 2016; Forootan et al., 2019) and for constraining water storage in hydrological models (Reichle et al., 2008; Van et al., 2014; Eicker et al., 2014; Tangdamrongsub et al., 2015; Khaki et al., 2017). Quantifying the total water budget, i.e. the balance of precipitation (P), evapotranspiration (E), runoff (Q) and the changes in total water storage at the Earth's surface, is key to understanding the global water cycle among the Earth's land, ocean, and atmosphere (Sheffield et al., 2009). Several studies also apply GRACE data to the measurement of ice mass loss (Velicogna, 2009; Khan et al., 2010; Velicogna et al., 2014; Mnhajeran et al., 2018; Ran et al., 2018; Tapley et al., 2019) and to the ocean mass balance (Hsu & Velicogna, 2017; Peralta & Woodgate, 2017; Chen et al., 2018; Jeon et al., 2018; Uebbing et al., 2019).

However, after more than fifteen years in orbit, the GRACE mission ended its operation in October 2017 and its successor - the GRACE Follow-On (GRACE-FO) mission - was only launched in May 2018 leading to approximately one year of missing data. Although several alternative sensors and data processing techniques have been proposed to derive surface mass change maps prior to the GRACE period and during the gap between the two generations of GRACE missions, e.g., from satellite laser ranging (Nerem et al., 2012; Talpe et al., 2017), global GNSS inversions (Wu, 2003; Rietbroek et al., 2013), or from the Swarm satellite

mission (Jäggi et al., 2016; Bezděk et al., 2016; Lück et al., 2018; Teixeira Encarnação et al., 2019), none of these appear to be able to provide a spatial resolution or accuracy comparable to that of GRACE.

Several studies have introduced approaches in reconstructing TWSC prior to the GRACE era (i.e. before April 2002) by constructing empirical relationships between GRACE TWSC and related climatic and hydrological variables (e.g., rainfall, temperature, sea surface temperature, soil moisture, etc.). For example, a simple approach was to extend basin-mean GRACE total water storage change time series beyond the GRACE observation period based on an Artificial Neural Network (ANN, Rao, 2000) model, in which the ANN learns the relationship between GRACE TWSC and climatic or hydrological variables and this was then used to reconstruct basin-averaged TWSC time series over the past decades (Long et al., 2014). Another approach was introduced in Forootan et al. (2014) to reconstruct gridded TWSC over a specific region. In their formulation, spatial patterns and temporal modes of the GRACE-derived TWSC were firstly separated, then TWSC fields were reconstructed by deriving the relations between the GRACE temporal modes and related predictors. The Independent Component Analysis (ICA, Forootan & Kusche, 2012) method was suggested to separate the GRACE signal and the AutoRegressive eXogenous (ARX, Ljung, 1987) method was applied to derive the relations and produce the reconstructions. Another approach, proposed by Humphrey et al. (2017), reconstructs total water storage changes at each grid point globally using the main climate indicators that were selected to be precipitation and temperature fields. They firstly decomposed the gridded GRACE TWSC time series, as well as precipitation and temperature, into a linear trend, an inter-annual component, a seasonal component, and a high-frequency residual component. They then reconstructed the de-seasoned (i.e. inter-annual and residual) component of TWSC at the global scale by deriving relations between the de-seasoned component of GRACE TWSC and precipitation and temperature (the linear trend and seasonal components were not reconstructed). In their approach, they employed the Seasonal-Trend decomposition based on Loess (STL, Cleveland et al., 1990) procedure to decompose the GRACE and climate signals and use the Multiple Linear Regression (MLR, Myers, 1986) approach to derive the relationships between TWSC and its predictors. Recently, these authors have advanced their method to generate global de-seasoned total water storage changes at a spatial resolution of 0.5° , at both daily and monthly scales over the period 1901 to present (Humphrey & Gudmundsson, 2019). Finally, a Deep Convolutional Neural Network (CNN) was recently applied in Sun et al. (2019) to predict spatial and temporal modes of mismatch between GRACE TWSC and water storage change as simulated by hydrological models, and to continue the correction of model-simulated TWSC fields. In their approach, the spatial representation of mismatch between GRACE TWSC and model-simulated TWSC was firstly predicted using maps of model-simulated TWSC, temperature, and precipitation at each epoch based on the CNN, then, total water storage changes was reconstructed by removing the predicted mismatch from model-simulated TWSC map.

Though all the mentioned studies are categorized as data-driven techniques and appear useful for reconstructing GRACE-like TWSC fields, to the best of our knowledge, no studies so far have compared their characteristics in a unified framework - i.e. for the same target region, with the same input climate data, the same temporal extension period (e.g., focusing on long-term trends or rather at seasonal scales) and spatial data resolution. Existing studies also did not yet assess the skills of these methods under identical conditions - e.g., length of

training and evaluation periods. However, with the GRACE data gap it is now of great concern to identify a reliable and repeatable approach to reconstruct a long and uninterrupted time series, and possibly also to reconstruct the TWSC prior to the GRACE period. The primary objective of this paper is therefore to provide a comparison of different methods for TWSC prediction or reconstruction. In this study, we place different methods in a unified methodological framework (Figure 1) to assess their skills based on identical climate input data, and finally identify a robust combination of them for the GRACE gap filling or for long term total water storage change reconstructions.

As a case study, we then assess the skills of combinations of data-driven methods such as Principal Component Analysis (PCA, Wold 1987), ICA, Least Squares (LS, Durbin & Watson, 1992), STL, ANN, ARX, and MLR methods for extrapolating the GRACE gridded TWSC time series outside the GRACE period over twenty-six river basins using precipitation, land surface temperature, climate indices, and Sea Surface Temperature (SST) data as indicators. We identify the most robust combination of these methods for all study regions, and our tests indicate that the extrapolated (up to six years in our case) gridded total water storage change maps, in all study regions, have much higher correlation with the observed GRACE data as compared to simulated water storage changes from hydrological models.

Following this introduction, in Section 2, the unified framework and the details of the applied data-driven methods are described. In Section 3, we introduce the GRACE, Swarm, climate, and hydrological data, and in Section 4, experiment results and discussions are provided for twenty-six study regions. We provide conclusions in Section 5.

2. Methods

2.1. Unified Methodological Framework

For an unbiased comparison, we deem it necessary to place methods in a unified framework first, to assess their prediction skills with identical input and validation data (see Figure 1), for the same region and using the same metrics.

In general terms, all methods utilize decomposition methods (usually ICA or PCA, Forootan et al., 2014) to partition the GRACE TWSC maps over a specific region and the suspected climate drivers (we employ precipitation, land surface temperature, climate indices, and SST) into spatial patterns and temporal modes. Then, time series analysis methods such as the STL procedure (Cleveland et al., 1990) or a simple Least Squares (LS) fitting method are used to further separate the individual modes of GRACE and climate data into typically a linear trend, seasonal, inter-annual, and the residual part. Third, the seasonal and de-seasoned (i.e. inter-annual and residual) components of the GRACE temporal modes are then reconstructed or predicted from empirical relationships between the temporal modes of GRACE and climate data as identified for a training period from either ANN, ARX or MLR method. Eventually, the GRACE-derived linear trend is commonly added to the reconstructed seasonal and de-seasoned components to extrapolate the full GRACE temporal modes. Here we would like to mention that linear and other long-term (e.g., accelerated) trends in GRACE data are often caused by ice and glacier melt, dam management, human water abstractions, and these factors could vary over time, so it may lead to misinterpretation when one simply extrapolates GRACE trends. Furthermore, the long-term trend (estimated over 10 years of GRACE data) could be affected by inter-annual and decadal variability, which might also bias the trends estimation. A focus on the reconstruction of total water storage trends would

require a specific regional treatment of all factors mentioned above, which is out of the scope of this study. Finally, GRACE-like gridded TWSC maps are reconstructed and predicted by combining the GRACE-derived spatial patterns with the reconstructed temporal modes. It is, however, unclear how the choice of the particular spatial and temporal decomposition methods affects the skills of the forecasted maps. Moreover, no systematic study exists which compares the sensitivity of the predictions or reconstructions to the type of empirical relationship which is trained from GRACE and climate data residuals. Therefore, in what follows we derive different combinations of methods and assess them within our unified framework. All possible method combinations are abbreviated as PCA-STL-ANN, PCA-STL-ARX, PCA-STL-MLR, PCA-LS-ANN, PCA-LS-ARX, PCA-LS-MLR, ICA-STL-ANN, ICA-STL-ARX, ICA-STL-MLR, ICA-LS-ANN, ICA-LS-ARX, and ICA-LS-MLR. The nomenclature of the abbreviations follows the 3-tier pattern XXX (Spatial-temporal decomposition method, Section 2.2.1) - YYY (time series decomposition method, Section 2.2.2) - ZZZ (predictive method, Section 2.3).

Here, we will focus on isolating the sensitivity of the prediction with respect to one group of techniques while keeping the other two groups consistent, i.e. we will firstly employ the method combinations PCA-LS-ANN, PCA-LS-ARX, and PCA-LS-MLR to compare the performances of three predictive models (i.e. ANN, ARX, and MLR) and identify the most robust method for the TWSC prediction, then this strategy will be applied to compare the 1) two spatio-temporal decomposition techniques ICA and PCA or 2) two time-series decomposition techniques LS and STL. Figure 1 visualizes the flow of computation via these various combinations, here we will select several (e.g., m) climate predictors for each decomposed component (i.e. seasonal, inter-annual, or residual) of detrended GRACE total water storage change temporal modes. For instance, if we identify six dominant (as to reconstruct a given percentage of the signal energy, e.g., 95%) GRACE temporal modes for a specific region, then we will reconstruct eighteen (6 modes multiplied by 3 components) GRACE components using $m \times 18$ relevant predictors.

2.2. Signal Separation Methods

2.2.1. Spatio-temporal Decomposition

The GRACE maps have a resolution of approximately 300 km, but it would be computationally expensive and not justifiable to try to predict or reconstruct unobserved TWSC at each (e.g., 100 km) grid point globally. Thus, for dimensionality reduction, we assume one will use a statistical decomposition method to identify the main patterns and modes of GRACE-observed TWSC, and continue with predicting only their temporal evolution. In the end, gridded GRACE signals will be reconstructed assuming that the dominant spatial patterns do not change over time. This is a caveat of all methods; however it is a common assumption in the reconstruction of unobserved climate data which e.g. has been used for GRACE in Becker et al. (2011) and Forootan et al. (2014), or for sea level reconstruction from tide gauges in Church et al. (2004).

The PCA method seeks to separate the original data (i.e. GRACE and climate signals) into orthogonal spatial patterns (EOFs) and their associated temporal modes (EOF modes) as follows (Wold, 1987):

$$X_{n \times t} = E_{n \times n} T_{n \times t} \quad (1)$$

The data matrix $X_{n \times t}$, with n rows for each spatial grid cell and t columns for each

epoch, represents the mean-centered original data. Columns of $E_{n \times n}$ contain the separated spatial patterns and rows of $T_{n \times t}$ the related temporal evolution. The first r dominant modes will contribute to the majority of the original signal (Wold, 1987). In this case, we choose r as to retain 95% of the total signal variance. The original matrix can be approximately restored by:

$$\hat{X}_{n \times t} = \hat{E}_{n \times r} \hat{T}_{r \times t} \quad (2)$$

where $\hat{X}_{n \times t}$ represents the restored signal, columns of $\hat{E}_{n \times r}$ are the r dominant EOFs and rows of $\hat{T}_{r \times t}$ are the associated EOF modes.

Forootan & Kusche (2012) suggested to replace the PCA method by Independent Component Analysis (ICA) in GRACE signal decomposition, motivated by the assumption that physically independent real-world processes will more likely exhibit statistical independence than orthogonality (Forootan et al., 2014). In the ICA method one additionally rotates the dominant EOFs and the temporal modes with a rotation matrix R to maximize their statistical independence:

$$\hat{X}_{n \times t} = \hat{E}_{n \times r} \hat{R}_{r \times r} \hat{R}_{r \times r}^T \hat{T}_{r \times t} \quad (3)$$

where the rotated $\hat{E}_{n \times r} \hat{R}_{r \times r}$ are then denoted as independent components (ICs) in the context, and the $\hat{R}_{r \times r}^T \hat{T}_{r \times t}$ represent the temporal modes (IC modes) identified by the ICA method. Several methods have been proposed to determine the rotation matrix R , based on different procedures to define statistical independences (e.g., minimizing third-order or fourth order statistical cumulants). Here we employ the method introduced by Cardoso (1999) and implemented in Forootan & Kusche (2012).

2.2.2. Time-series Decomposition

In order to retain and extrapolate the observed trends, and to apply statistical prediction techniques on the anomalous signals only, one commonly partitions the temporal evolution of observed EOF/IC modes into 1) a linear trend, 2) inter-annual, 3) seasonal, and 4) residual signals. These components are then considered individually. When decomposing time series into seasonal, inter-annual, etc. components, observation errors are typically not taken into account and they are not considered in this study as well. Drawing on Figure 1, we consider the Least-Squares (LS) and Seasonal-Trend decomposition based on Loess (STL) methods to estimate these deterministic signals in the observed modes. In the LS method, linear trend, inter-annual, and seasonal components are estimated e.g. by linear regression, segmented cubic polynomial function, and annual sine-waves, and the residual or anomalous signal is obtained by removing these. In the segmented cubic polynomial, the first step is to partition the total time series into several shorter time series (or sub-series), then the second step is to fit each sub-series using a cubic polynomial function:

$$y(t) = a + bt + ct^2 + dt^3, t = 1, 2, 3, \dots, n \quad (4)$$

where $y(t)$, $t=1, 2, 3, \dots, n$ represent the segmented sub-series, t is the time and n represents the length of the sub-series. In this case, we set $n=19$ months corresponding to the smoothing parameter that was used in the STL procedure for decomposing the inter-annual component as described in Cleveland et al. (1990).

STL is a filtering procedure, which was introduced by Cleveland et al. (1990) and applied by e.g. Baur (2012); Frappart et al. (2013); Hassan & Jin (2014); and Humphrey et al. (2016) to GRACE data, it allows decomposing a time series into three components trend (i.e.

linear trend and inter-annual in this study), seasonal, and residual:

$$Y = T^c + S^c + R^c \quad (5)$$

where Y is the original time series; T^c , S^c , and R^c represent the trend, seasonal, and residual components separated from the original time series, respectively; c is the cycle-index in the inner loop of the STL procedure (Cleveland et al., 1990). The trend decomposed by STL comprised of linear trend and inter-annual components, so we then separate the linear trend using linear regression and the inter-annual is obtained by removing the linear trend from the STL-derived trend. Thus, equation (5) can be expressed as:

$$Y = L^c + I^c + S^c + R^c \quad (6)$$

where L^c and I^c are the linear trend and inter-annual components of the original time series. In this study, we implemented STL as in Cleveland et al. (1990).

STL, which consists of applying a sequence of smoothing operations, is computationally more intensive and generally retains more detailed features of the acquired time series when fitting seasonal components as compared to the LS method. Such difference between STL and LS may also lead to different prediction results when applying them in the prediction or reconstruction of GRACE total water storage changes as described in Section 2.1. In this study, we employ and compare both LS and STL methods and assess their performances for the TWSC predictions.

2.3. Three Predictive Models

Predictive models seek to learn a relationship between a group of predictors (here, precipitation, temperature, sea surface temperature fields, and climate indices) and the target variable (gridded GRACE TWSC) (Bishop, 2006). Previous studies have successfully employed Artificial Neural Network (ANN), AutoRegressive eXogenous (ARX), and Multiple Linear Regression (MLR) models to predict and reconstruct GRACE TWSC time series. In this study, we place these three models in a unified framework and compare them with the focus on prediction of GRACE total water storage changes.

2.3.1. Artificial Neural Network (ANN) Model

The Multi-Layer Perceptron (MLP) ANN model has been proposed in the past for predicting GRACE time series (Sun & Alexander, 2013). We implement this ANN model therefore for learning relations between decomposed components (i.e. seasonal, inter annual, and residual) of detrended temporal modes in GRACE data and the supposed climate predictors. In the most simple MLP model there are three layers i.e. the input, hidden, and output layers (Bishop, 2006; Long et al., 2014). In this study, the output layer of the ANN model is separately chosen to represent each decomposed component of detrended GRACE temporal modes and the inputs (predictors) comprise selected m sensitive components of climate temporal modes. The selected climate components (predictors) are determined based on their correlations as related to each target GRACE component - i.e. we select the predictors by retaining the most correlated climate components. The number of input 'channels' is set to $m=3$ because we find no obvious improvement when using a larger number of predictors. The hidden layer consists of u artificial neurons and each neuron represents a sum of weighted predictors. We set the number of artificial neuron to $u=7$ in the hidden layer based on the criterion as described in Forman et al. (2014). It is difficult to reasonably

initialize the weights of the artificial neurons in the ANN training process; such that we randomly choose start weights hundreds of times and repeat the training process, in order to finally generate the final prediction as a mean over several ANNs with different start weights. In this case, we write the ANN codes based on the Matlab_R2014b neural network function to develop all MLP networks for this study.

2.3.2. AutoRegressive eXogenous (ARX) Model

The ARX model, which formulates another type of relationships between a group of inputs and the output, is governed by a system of linear equations (Ljung, 1987):

$$y(t) + \sum_{i=1}^{n_a} a_i y(t-i) = \sum_{q=1}^m \sum_{l=1}^{n_b} b_{q,l} x_q(t-l-1) + \varepsilon(t) \quad (7)$$

where $y(t)$, $t=1,2,\dots,n$ represent the target variables, t is the time epoch and n is the length of the time series; $x_q(t)$, $q=1,2,\dots,m$ represent m channels of inputs ($m=3$ in this case); n_a and n_b are the orders of the autoregressive exogenous model with respect to the output and input respectively. $\varepsilon(t)$ allows for a random Gaussian-noise input. Here, a_i and $b_{q,l}$ are the coefficients that need to be estimated in the training step using both inputs and output data; thus they play a role similar to the weights as in the artificial neural network approach. In this case, we set both n_a and n_b to 3 as discussed in Forootan et al. (2014). After obtaining \hat{a}_i , $\hat{b}_{q,l}$, \hat{n}_a , and \hat{n}_b one can predict the target variable beyond the training period based on these coefficients and parameters:

$$\hat{y}(t_n) = -\sum_{i=1}^{\hat{n}_a} \hat{a}_i \hat{y}(t_n - i) + \sum_{q=1}^m \sum_{l=1}^{\hat{n}_b} \hat{b}_{q,l} x_q(t_n - l - 1) \quad (8)$$

where $\hat{y}(t_n)$ represents the predictand of the ARX model at the epoch t_n . More details about the application of the ARX model to predict GRACE temporal modes can be found in Forootan et al. (2014). In this case, we write the ARX codes by referring to the equations as described in Forootan et al. (2014) and we use identical inputs and output data as employed in the ANN process (see Section 2.3.1) to train the ARX model and to extrapolate the GRACE TWSC time series - i.e. we choose three sensitive components of climate temporal modes to predict and reconstruct each decomposed component of detrended temporal modes of GRACE TWSC using the autoregressive exogenous model (equation (7) for training and equation (8) for testing/predicting).

2.3.3. Multiple Linear Regression (MLR) Model

The MLR model prescribes linear relationships between multiple input and one output variables (Sousa et al., 2007). In this case, we use the multiple linear regression function from Matlab_R2014b. The inputs will be three selected components of climate temporal modes while the output will be the decomposed component of detrended GRACE temporal modes. Again, for a fair comparison, we will employ identical input and output data as used in the ANN and ARX models to train the MLR representation. Here, we choose the least squares method for the estimation of the MLR coefficients, and then we predict the target variables based on the estimated coefficients.

2.3.4. Comparison

The artificial neural network model can fully derive nonlinear relationship between the

input and output data, but it is difficult to optimally determine the number of artificial neurons and it may lead to overfitting the problem. Furthermore, it is computationally intensive to repeat the ANN training process for improving the predictand. The ARX and MLR models both employ linear relations, so they cannot be expected to predict nonlinear relationships too well. Within the ARX model, each predicted value depends on the nearest former predictand - i.e. $\hat{y}(t_n)$ in equation (8) is predicted by using $\hat{y}(t_n - i)$ $i=1,2,\dots, n_a$ - so the predicting error of the autoregressive exogenous model is easily accumulated over time. Therefore, the multiple linear regression method will be likely a better choice on the condition that there are no nonlinear relationships between the input and output data.

2.3.5. Error Perturbations

In order to study the error propagation characteristics in the three predictive models, we generate a series of Gaussian-like uncertainties to perturb each target variable (i.e. GRACE seasonal, inter-annual, or residual component) based on the Monte Carlo approach (Challa & Hetherington 1988). The perturbed target variables could be expressed by the equation as follows:

$$P_i(t) = G(t) + \xi_i(t), t = 1, 2, 3, \dots, n \quad (9)$$

where $P_i(t)$ ($t=1,2,3,\dots,n$) is the i -th perturbed GRACE TWSC, i is the count of iteration times (or random sample number); t is the time epoch and n is the length of the GRACE time series; $G(t)$ represents unperturbed GRACE TWSC; and $\xi_i(t)$ is the Gaussian-like uncertainties generated by the Monte Carlo approach. Here, we first predict the GRACE-like gridded total water storage changes using the unperturbed GRACE TWSC as target variable, then we predict another group of GRACE-like gridded TWSC using the perturbed GRACE TWSC as target variables, and the error propagation bars are estimated by the mismatch between the unperturbed and perturbed GRACE-like gridded TWSC:

$$B(t) = \sqrt{\sum_{i=1}^m (PT_i(t) - UT(t))^2}, t = 1, 2, 3, \dots, n \quad (10)$$

where $B(t)$ is the error propagations in the predictive model, t is the time epoch; $PT_i(t)$ is the GRACE-like gridded TWSC predicted using the i -th perturbed GRACE TWSC as target variable and $UT(t)$ is the GRACE-like gridded TWSC predicted by the unperturbed GRACE TWSC; m represents the sum of iteration times of error perturbations.

3. Data and Processing

3.1. Total Water Storage Change Data

We use the RL06 GRACE monthly mascons, developed with a 1° resolution using Tikhonov regularization in a geodesic grid domain (Save et al., 2016; Save & Himanshu, 2019) by the Center for Space Research (CSR), between April 2002 and June 2017 as the target variables. The storage anomalies, which capture all the signals observed by GRACE within the measurement noise level, are given in equivalent water thickness units (cm) and the correlated error has been intrinsically removed, thus, these products do not need to be additionally destriped or smoothed. In this study, we unify the spatial resolution of all input data to $1^\circ \times 1^\circ$ (corresponding to the CSR mascons) to eliminate inconsistent resolutions between the input and output data.

The GRACE Follow-On (GRACE-FO) mission has been operated since May 2018 and we use the GRACE-FO TWSC as a criterion to evaluate the accuracy of the predicted TWSC.

The GRACE-FO temporal gravity field models, from June 2018 to December 2018, derived by the CSR are employed to estimate the GRACE-FO TWSC over all study regions based on the methods and strategies as described in Li et al. (2018).

During the gap of the two GRACE missions, the Swarm satellites may serve as an alternative to derive Earth's gravity field models albeit at much lower resolution. As a verification to our predicted TWSC, the RL06 Swarm time-variable gravity field models from December 2013 to December 2018 calculated by Lück et al. (2018), with a max degree of 40, are also employed to estimate the Swarm total water storage changes. Here, we only use the Swarm gravity field models complete to degree and order ten, to suppress excessive noise at higher orders.

3.2. Climate Data

3.2.1. Global Precipitation

The Climate Prediction Center (CPC) global daily unified gauge-based analysis of precipitation (Chen et al., 2008), with a spatial resolution of 0.5° , are employed in this study. The monthly precipitation is obtained by averaging the CPC global daily precipitation corresponding to the GRACE time interval, and both daily and monthly precipitation from October 1991 to March 2019 are used to reconstruct and predict the GRACE-like TWSC out of the GRACE period. For improving the correlation between precipitation and GRACE temporal mode de-seasoned terms (i.e. inter-annual and residual), we reconstruct the de-seasoned components of monthly precipitation temporal mode using the daily precipitation temporal mode based on a time delay parameter as introduced by Humphrey et al. (2016). The spatial resolutions of both daily and monthly precipitation are made consistent to 1° .

3.2.2. Global Land Surface Temperature

Global Historical Climatology Network version 2 and the Climate Anomaly Monitoring System (GHCN CAMS), developed at CPC, National Centers for Environmental Prediction (NCEP), is a monthly global land surface temperature dataset ($0.5^\circ \times 0.5^\circ$) from 1948 to near present (Fan & Dool 2004). We use this dataset between October 1991 and March 2019 as one of the input climate data to reconstruct and predict GRACE-like TWSC time series. As discussed before, the spatial resolution of this dataset is sampled onto 1° cells.

3.2.3. Sea Surface Temperature (SST)

Sea surface temperature drives ocean evaporation, which in turn affects atmospheric moisture transport and, eventually, rainfall over land areas and water storage. For example, the temporal evolution of GRACE TWSC in west Africa is highly correlated with the SST anomalies in the Pacific, Atlantic, and Indian oceans (Forootan et al., 2014). To take advantage of this kind of climate data, we use the SST in several oceans and seas that are located near the study regions (Figure 2) as one kind of input data. In our case, the monthly Optimum Interpolation (OI) sea surface temperature, with a resolution of 1° , provided by National Oceanic and Atmospheric Administration (NOAA) is employed (Reynolds et al., 2002).

3.2.4. Climate Indices

Climate indices are by definition constructed to capture large-scale variability in fields such as SST or surface pressure, which are related to land precipitation and temperature through atmospheric teleconnections. Therefore, several publications (e.g., Anyah et al., 2018) have shown that they play a key role in representing inter-annual GRACE water storage variations. In this study, seventeen climate indices - i.e. Multivariate ENSO Index (MEI), North Atlantic Oscillation (NAO), Extreme Eastern Tropical Pacific SST (Niño 1+2), Eastern Tropical Pacific SST (Niño 3), Central Tropical Pacific SST (Niño 4), East Central Tropical Pacific SST (Niño 3.4), North Tropical Atlantic SST Index (NTA), Oceanic Niño Index (ONI), Pacific Decadal Oscillation (PDO), Pacific North American Index (PNA), Quasi-Biennial Oscillation (QBO), Southern Oscillation Index (SOI), Tropical Northern Atlantic Index (TNA), Trans-Niño Index (TNI), Tropical Southern Atlantic Index (TSA), Western Hemisphere Warm Pool (WHWP), and Western Pacific Index (WP) - are involved as another kind of input data.

3.3. Hydrological Models

Hydrological models simulate soil moisture, near surface air temperature, accumulated snow, water/energy flux, and other hydrological components on land. Several studies have shown that model outputs relate well with GRACE TWSC although model schemes do not include, e.g., all water reservoirs, because the temporal evolution of the different water reservoirs is often highly connected (Humphrey et al., 2017). We use model outputs, including the NASA Global Land Data Assimilation System (GLDAS) NOAH 10M series model (Rodell et al., 2004) and CPC soil moisture (Fan et al., 2004) from January 1992 to December 2018, to evaluate the reliability of the reconstructed and predicted total water storage change. Moreover, we also employ the newest WaterGAP Global Hydrology Model (WGHM) version 2.2d results, over January 1992 to December 2016, in this study. Compared to the 2.2a version (Döll et al., 2014) the water balance is closed, the calibration routine is changed, and the human water use values and the groundwater recharge algorithm is improved.

4. Results

We choose twenty-six major river basins of the world, as delineated in <https://www.grida.no/resources/5782>, as the study regions (Figure 2). For representing the amplitude of GRACE TWSC, we calculate the Root-Mean-Square (RMS) of each gridded CSR mascon over the study regions (Figure A1). To make full use of the SST data in the TWSC reconstruction and prediction, we divide the global sea surface temperature data into fourteen patches as shown in Figure 2 and use each of them as one of input data.

4.1. Signals Separated from the GRACE Data

4.1.1. Dominant Temporal Modes of GRACE Total Water Storage Change

As discussed in Section 2.2.1, the dominant modes identified by the PCA or ICA methods contribute to the majority of the GRACE signal. We predict each selected GRACE EOF/IC temporal mode individually based on the methods as described in Section 2 and find that forecasting modes with less energy (i.e. less variance explained of the GRACE signal) tend to have relatively higher standard errors as estimated by the CSR mascons, thus, while

choosing more dominant EOF/IC modes in the TWSC prediction will reduce the signal leakage (i.e. signals from discarded modes), this will also increase the prediction uncertainties. Consequently, it is difficult to choose the optimal number of modes to be selected. An optional approach could be that one first predicts the TWSC based on a different number (e.g., from 3 to 10) of retained modes, and then uses a GRACE solution to test the uncertainties of TWSC predicted by these different numbers of modes to identify a best number for each study region. It is important to highlight that although the GRACE solution cannot be viewed as an unambiguous reference, it could be used to estimate the accuracy of the prediction because the reference (i.e. GRACE solution in this study) should be accompanied by a 'conservative error estimate'. It is clear that further tuning the methods would improve the reconstruction/prediction skills for a specific basin. But for applying our approach globally, i.e. to a large number of basins, tuning for each basin would not be in the interest of repeatability, and it is not clear how robust such over-tuned approaches would be. Thus, in what follows we rather define unified criteria for all study regions based on extensive testing on only a few representative basins. Here, we set a unified criterion - i.e. the number of modes that jointly explain at least 95% of the total variance - to identify the number of dominant modes automatically by the algorithm. We note that we set the number of selected climate modes equal to the number of selected GRACE modes just to minimize the inconsistency between the input and output data, but this is not strictly required for the algorithm.

4.1.2. Linear trend, Seasonal, Inter-annual, and Residual Components of the Dominant GRACE Temporal Modes

After identifying the dominant EOF/IC modes, we decompose each temporal mode of both GRACE EOFs and ICs using the LS or STL method as described in Section 2.2.2. The decomposed components for a case region (i.e. the Amazon basin) are shown in Figure A3. These results indicate that there is no large difference between the LS and STL methods for decomposing the linear trend and inter-annual components, and the two methods perform almost the same in separating the seasonal component from the temporal mode which has a strong periodicity (e.g., temporal modes of GRACE EOF1 and EOF2 in Figure A3). When decomposing a high frequency time series, the LS and STL methods exhibit some differences in separating seasonal signals - i.e. seasonal signals separated by STL method show more detailed features and larger oscillations (e.g., the temporal mode of GRACE EOF6 in Figure A3).

4.2. Prediction and Reconstruction of Total Water Storage Change for Twenty-six Study Regions

4.2.1. Selection of Climate Inputs for the Total Water Storage Change Prediction

In this study, we predict and reconstruct the inter-annual, seasonal, and residual components of significant GRACE TWSC temporal modes for each study region based on the predictive models that was introduced in Section 2.3. These predictive models seek to derive the relationship between the input and output data. Typically, there are a few months of lag time between the climate variations and GRACE TWSC, thus, our algorithm is designed to automatically move each climate driver (i.e. the inter-annual, seasonal, or residual of climate temporal modes) time series for a few months (i.e. 0 to 3 months) to

search for the strongest correlation, and for this we also interpolate the GRACE TWSC time series to fill the missing data. In addition, we reconstruct the de-seasoned components of precipitation temporal modes based on a time decay parameter that was developed by Humphrey et al. (2016) to further improve the correlation between the precipitation and GRACE components.

For selecting the ‘best’ input data, the correlation coefficients between each target variable (i.e. the inter-annual, seasonal, or residual of GRACE EOF and IC temporal modes) and related climate drivers are automatically computed and sorted by our algorithm, e.g., for a specific basin we calculate the correlation coefficients between the inter-annual component of GRACE EOF1 mode and inter-annuals from precipitation, land temperature, SST (in fourteen different oceans and seas) EOF modes, and seventeen climate indices and sort them by size, then this process is successively applied to GRACE EOF1 mode seasonal, GRACE EOF1 mode residual, GRACE EOF2 mode inter-annual, and so on. Here, the sensitive input data is sorted only by correlation coefficients, and for the selection this method may reject a predictor that is not very highly correlated but brings new information compared to other highly correlated predictors. So, before the selection of inputs, we use the stepwise regression method (Summers, 1985) to remove the highly correlated predictors which do not bring sufficient new information. In addition, we would like to make a cautionary note that correlation between climate input and the GRACE data does not necessarily represent causation, and in this case our (like any other similar) techniques may derive a ‘right answer due to the wrong reasons’, but of course may fail in extrapolating well.

As discussed in Section 2.3, we choose three sensitive climate drivers as predictors to extrapolate each target variable. Furthermore, we identify one most sensitive climate driver for each target variable as listed in Table 1. We find that the temporal evolution of GRACE total water storage change seasonal component is highly related to the seasonal component of SST (see the third column in Table 1), and the time series of GRACE TWSC inter-annual and residual components are strongly correlated with the inter-annual and residual of both sea surface temperature and precipitation variations (see the fourth and fifth columns in Table 1).

4.2.2. Metrics for Methods Comparison and Estimation of Prediction Uncertainties

4.2.2.1. Criteria for Identifying the Most Robust Method

There are more than fifteen years of GRACE data altogether (April 2002 - June 2017). In the data processing, we set the training section to April 2002 - June 2011 and set the testing period to July 2011 - June 2017, i.e. we use the GRACE data from April 2002 to June 2011 to train the predictive models and to test the uncertainty of the next six years (i.e. July 2011 - June 2017) of prediction. Then, we use the CSR mascons to calculate the standard error of both training and testing TWSC. We note that one can determine ‘absolute’ errors, including the systematic and random errors, only in the training phase. In the computations, we set the testing section within the GRACE period just for assessing the uncertainty of predicted TWSC. As discussed in Section 3.1, the CSR mascon contains the GRACE measurement noises. Several studies (e.g., Landerer & Swenson, 2012) have assessed the measurement errors that are contained in the GRACE-derived total water storage changes. These measurement errors may falsify the validation between the predicted/reconstructed TWSC and the CSR mascons, but it is difficult to exactly determine the exact influences of GRACE measurement errors on the TWSC prediction. Through we do not assess this influence, one

can estimate to what extent the GRACE measurement error may affect the predicted TWSC by computing the root-sum-square of the GRACE measurement errors and the prediction errors.

As discussed in Section 3, three groups of data-driven techniques - i.e. 1) Spatio-temporal decomposition techniques ICA and PCA; 2) time-series decomposition techniques LS and STL; and 3) machine learning methods ANN, ARX, and MLR - are employed in this study. Here, we firstly fix the Spatio-temporal decomposition and time-series decomposition techniques to PCA and LS and compare the robustness of three machine learning techniques ARX, ANN, and MLR in twenty-six river basins.

In this case, we used three criteria - i.e. 1) standard error of TWSC, 2) correlation coefficients of TWSC, and 3) correlation coefficients of de-seasoned TWSC as used in Reichle et al. (2004) - to identify the most (or more) robust method. Table 2 lists the standard errors of training and testing TWSC by using the three machine learning methods as evaluated by the CSR mascons at both grid and basin scales. The correlation coefficients of training/testing TWSC and TWSC anomaly (i.e. de-seasoned TWSC) as compared to the CSR mascons are listed in Table 3 and Table 4 respectively. We here summarize the optimal methods - i.e. methods with minimal standard error or maximal correlation coefficients at the grid scale - for each river basin and highlight them in blue fonts (see Table 2 - Table 4).

Within the training section, the ANN model simulates the TWSC best in 18 river basins as estimated by the criterion of standard error (see fourth column in Table 2) and simulates the TWSC best in 19 regions and 20 regions assessed by the other two criteria as shown in Table 3 and Table 4. ARX performs best in 12 basins, 16 basins, and 10 basins within the training phase as assessed by the criteria standard error of TWSC, correlation coefficients of TWSC, and correlation coefficients of de-seasoned TWSC respectively, and MLR simulates the TWSC worse in all river basins than ARX or ANN (see column 2 to column 7 in Table 2 - Table 4). These results indicate that MLR performs worst and ANN performs best within the training period.

The MLR method, in the testing period, shows the best skill in 19 regions, 18 regions, and 19 regions as evaluated based on the three criteria, respectively. Obviously, the ANN and ARX models perform worse than MLR in the testing period as listed through column 8 to column 13 in Table 2 - Table 4. Here, we also calculate the average standard error or correlation coefficient over 26 river basins for each predictive method and in both training and testing period as listed in last rows of Table 2 - Table 4. All these results indicate that though MLR perform worst within the training phase, it is the most robust method for the prediction.

For ANN and ARX, we use a group of unified empirical parameters for predicting the TWSC grids. With these empirical parameters (e.g., number of input predictors), ANN and ARX perform better than MLR in some regions while they perform worse in the other regions, indicating that these parameters are not optimal in all study regions and might lead to some overfitting problems. One alternative is to try all different combinations of these parameters for ANN and ARX and find a best combination for each study region, this may suppress the overfitting problem but will also dramatically expand the testing works and it is computationally expensive to apply them to all river basins globally. Therefore, we suggest one to try ANN and ARX if he/she aims at predicting only a few basins, and the MLR method is a more robust alternative if one wants to predict the TWSC over a large number of river basins. After our tests with three predictive methods, we then choose the spatio-

temporal decomposition and machine learning techniques to PCA and MLR to compare the performances of STL and LS in all study regions. We use the CSR mascons and the three criteria to evaluate the precision of testing/training TWSC from STL and LS methods (see Table B1-Table B3) and we find that the LS method performs better in most of the cases compared to STL in those study regions. Finally, we use the same metric to compare the PCA and ICA methods. As listed in Table B4 - Table B6, the standard errors of testing TWSC from PCA are smaller than those from ICA and the correlation coefficients of both TWSC and de-seasoned TWSC from PCA are higher in most river basins, indicating that PCA is more robust than ICA for such application.

4.2.2.2. Error Propagation in Three Predictive Models

In this case, for studying the error propagation characteristics in three predictive models, we first repeat the error perturbations and TWSC prediction and calculate the error bars for the testing TWSC at grid cell scale as described in Section 2.3.5. Then we divide the study period into different sections, i.e. training period, the first year past the training period, the second year past the training period, and etc. Figure A4 shows the propagated errors in each divided time section for the three predictive models at the grid scale. As expected, we find that the error bars, from both ANN and ARX methods, in the testing sections (first yr. - sixth yr.) are larger than those in the training section (the second and third columns in Figure A4) and the error bars from the MLR method are almost constant in both training and testing sections (the first column in Figure A4), indicating that the multiple linear regression method is more stable than the ANN and ARX methods in coping with random input uncertainties. This enhances our confidence to choose the MLR method for the prediction and reconstruction. In addition, the results also show that the distribution of propagated uncertainties in predictive methods depend on the amplitude of gridded GRACE TWSC, such as the GRACE total water storage changes with larger amplitudes (e.g., TWSC in the middle of the Amazon basin) also show larger uncertainties (Figure A4).

4.2.2.3. Prediction Uncertainties of the Identified Methods

Based on the testing results in Table 2 - Table 4, Table B1 - Table B6, and Figure A4, the combination of PCA, LS, and MLR methods are identified to be the most robust combination for extrapolating the GRACE TWSC map.

Figure 3 shows the standard errors of both training and testing GRACE-like gridded total water storage changes as evaluated by CSR mascons at the grid scale, and we find that the regions with larger amplitude of TWSC also show larger standard errors.

Figure 4 shows the correlation coefficients of training TWSC, testing TWSC, and TWSC simulated by the hydrological models as related to the GRACE mascons over twenty-six river basins. The averaged correlation coefficients (i.e. averaging the correlation coefficients of all grids in twenty-six river basins) of training TWSC, testing TWSC, GLDAS TWSC, CPC TWSC at the grid scale are 0.91, 0.86, 0.51, and 0.45 as compared to the CSR mascons. We also remove the seasonal cycles of the related TWSC at the grid scale as described in Reichle et al. (2004) and obtain the correlation coefficients of de-seasoned/anomaly TWSC signals (see

Figure 4e -

Figure 4h), the correlation coefficients of anomaly signals are 0.77, 0.68, 0.44, and 0.46,

respectively. Results indicate that both training and testing GRACE-like gridded total water storage changes have much stronger correlations with the GRACE mascons than the model-simulated TWSC. Furthermore, both training and testing TWSC time series at the basin scale fit the CSR mascons well in almost all study regions as shown in Figure 5.

4.2.2.4. Identification of the Optimal Region Size

In this study, we apply our methods to all river basins respectively. For understanding at which basin size the predicting methods work best, we firstly divide the Europe-Asia continent into different sizes of sub-continent (see Figure 6), then the identified methods combination is applied for the TWSC prediction in each sub-continent. Again, standard errors are derived by comparing to the CSR mascon solution as shown in Figure 6. The averaged standard errors in Figure 6 (a) - Figure 6 (d) are 2.86, 2.88, 2.74, and 2.84 cm respectively, indicating that dividing the Europe-Asia continent into four parts (~10-15 million km² per sub-continent) will be an optimal option for the prediction of TWSC based on our identified methods. Thus, we suggest an optimal region size 10-15 million km² for readers who want to predict the TWSC using our method.

4.2.3. Extrapolating the GRACE Total Water Storage Change Outside the GRACE Period

As discussed in Section 4.2.2, the PCA, LS, and MLR methods are identified to predict and reconstruct the GRACE-like gridded total water storage change over twenty-six river basins, i.e. in each river basin 1) we use the PCA method to identify significant modes of the GRACE and climate signal; 2) we use the LS method to separate inter-annual, seasonal, and residual components of GRACE and climate temporal modes; 3) we use the separated components from GRACE and climate data between April 2002 and June 2017 to train the multiple linear regression model, and then predict the GRACE-like gridded TWSC over July 2017 to December 2018 and reconstruct TWSC from January 1992 to March 2002. El Niño-Southern Oscillation (ENSO) represents natural variability in the climate system, which may cause some climate extremes especially in the tropical regions (Juan et al., 2016). Isolating the de-seasoned signal in total water storage change enables one to detect climate extreme events such as hydrological drought (Thomas et al., 2014). In an attempt to investigate the response of water storage to ENSO, we show the predicted, training, and reconstructed TWSC time series of four tropical river basins - i.e. Amazon, Parana, Congo, and Zambezi basins - at the basin scale (see

Figure 7). For a comparison, we also plot the GRACE TWSC (April 2002 - June 2017), Swarm TWSC (December 2013 - December 2018), GRACE-FO TWSC (June 2018 - December 2018), and the TWSC (January 1992 - December 2018) simulated by hydrological models in

Figure 7. Within our analysis period, there were two significant El Niño events, i.e. the years 1997/1998 and 2015/2016 that were significant can be derived from ENSO indicators. We find that, after removing the seasonal cycle, the reconstructed total water storage change shows strong abnormal signals in four tropical regions during the first significant El Niño periods (i.e. 1997/1998), which is consistent with the de-seasoned GRACE TWSC during another significant El Niño period (i.e. 2015/2016) (see the dark area of

Figure 7). As shown in

Figure 7, the GRACE-FO TWSC fits well with the predicted TWSC in four tropical river basins. The hydrological model does not explicitly simulate groundwater redistribution and its skills in representing anthropogenic water withdrawals are limited, this may explain most differences between TWSC from the hydrological model and those from GRACE, GRACE-FO, or Swarm mission seen in

Figure 7. The Swarm TWSC fits well with the GRACE total water storage change in the Amazon basin, but shows larger deviations and uncertainties in the other regions, this shows some potential of the Swarm mission to detect large amplitude water storage changes. We also find that the Swarm TWSC has larger uncertainty in 2013 and 2014 than later, this is mainly caused by the more active ionosphere during these two years as discussed in Schreiter et al. (2019). In addition, the predicted and reconstructed TWSC in the other river basins could be found in Figure A5 and Figure A6. The results clearly suggest that the predicted TWSC over June 2018 to December 2018 fit well with the GRACE-FO TWSC in almost all study regions.

4.2.4. Comparison to Previous Studies

Except Humphrey & Gudmundsson (2019), we do not know other studies that aim at the reconstruction of GRACE-like gridded total water storage change for the global land surface. Thus, we restrict the comparison of our method to the method as used in Humphrey & Gudmundsson (2019).

In their TWSC reconstruction, Humphrey & Gudmundsson (2019) focused on a grid cell representation using precipitation and temperature as inputs. However, it appears difficult with their approach to make use of additional input datasets originating from outside the study regions (e.g., the SST data or climate indices), and they did not reconstruct the seasonal signal of the GRACE TWSC as described in Humphrey & Gudmundsson (2019).

In this study, the decay filter, MLR, and STL techniques employed in Humphrey et al. 2017 are included in our unified framework, but we had to combine these in a somewhat different way. Different from Humphrey's original method, we focused our attention on the reconstruction of dominant GRACE modes, and we feel it is beneficial to involve additional data which have been shown before to be highly related to the evolution of dominant GRACE temporal modes as inputs (e.g., the SST, climate indices, etc.). Therefore, our implementation is able to assimilate more information to support the TWSC reconstruction. Furthermore, besides aiming at the de-seasoned anomalous TWSC signals, we also reconstruct the seasonal signals based on their high correlations with the seasonal signals of SST (see Table 1). Thus, our implementation seems able to reconstruct a more complete picture of the GRACE TWSC record as compared to the method originally developed by Humphrey & Gudmundsson (2019).

In addition, we would like to mention that it is the first time that three groups of data-driven techniques - i.e. spatio-temporal decomposition, time series decomposition, and machine learning - are formulated in a unified way for the reconstruction or prediction of leading modes identified in GRACE-derived total water storage grids.

5. Conclusions

In this study, a unified methodology framework is developed to compare different data-driven techniques for predicting and reconstructing gridded GRACE-like total water storage variations outside the GRACE period. We find that both ARX and ANN methods simulate

the target variable better than the MLR method, but they are not robust enough for the prediction on account of some overfitting problems. The PCA-LS-MLR methods combination is identified as the most robust alternative through our framework for predicting and reconstructing the gridded TWSC over all river basins. One encouraging result is that our testing TWSC (six years lead time with regard to the GRACE period) from the identified methods show much stronger correlation with the CSR mascons compared to the model-simulated TWSC at almost each grid of the study regions, thus, our results could be an alternative for the mass balance constraint for hydrological models beyond the GRACE period (e.g., Eicker et al., 2014). We also find the temporal evolution of the seasonal component of the GRACE total water storage change in the study regions are closely related to the seasonal variation of SST and the de-seasoned (i.e. inter-annual and residual) components of GRACE TWSC have strong correlation with the de-seasoned changes of both sea surface temperature and precipitation. These results may improve our understanding of roughly relationships between the TWSC and the related climate drivers.

We study the error propagation in the adopted predictive models, and the results indicate the MLR method is more stable and robust than the ANN and ARX methods in coping with error perturbations. Finally, we predict 1.5 years (i.e. July 2017.7 - December 2018) of gridded TWSC past the GRACE period and reconstruct more than ten years (i.e. January 1992 - March 2002) of gridded TWSC. At the basin scale, the de-seasoned signal of reconstructed TWSC exhibits a strong abnormal signal in the tropical basin during significant El Niño periods. The total water storage change derived from the GRACE Follow-on mission fits well with the predicted TWSC in almost all study regions and the Swarm TWSC shows potential to detect extreme climate events but it contains large uncertainties.

The approach identified from our framework presents a viable alternative for bridging the data gap of the GRACE missions and can also be used for extrapolating the global GRACE gridded TWSC time series outside the GRACE period for a longer time.

Acknowledgments: The GRACE RL06 mascon solution used in this study is available from CSR (http://www2.csr.utexas.edu/grace/RL06_mascons.html); GRACE-FO data is available from GFZ (<https://isdc.gfz-potsdam.de/grace-fo-isdc>); Climate data and CPC soil moisture are available from NOAA (<https://www.esrl.noaa.gov/psd/>); GLDAS data is available from NASA (<https://disc.sci.gsfc.nasa.gov/uui/datasets?keywords=GLDAS&start=1920-01-01&end=2017-02-13>). This study is supported by the NSFC (China), under Grants 41274032, 41474018, 41429401, 41774019, and 41704011, and by the Ph.D. Short-time Mobility Program of Wuhan University. We also acknowledge funding by the German Research Foundation DFG under grants KU1207/26-1 and RI2657/2-2. We thank Bart Forman and an anonymous reviewer for their constructive feedback during the review of this paper.

Codes and output dataset

The related codes and TWSC dataset produced in this study are available at: https://github.com/strawpants/twsc_recon. Noting that the Github is not an appropriate repository and we put them in Github just for review purposes, thus, the data archiving is underway and we plan to move the codes/data to the repository Open Science Framework (OSF).

Appendix C: Robustness Tests of the Identified Methods

In the test computations, we had to choose some options, such as setting the training section to April 2002 - June 2011, setting the input number of climate drivers to three, etc.,

and identified an optimal methods combination i.e. PCA-LS-MLR. But the prediction skill may vary with different options. Therefore, in this section we will tell how robust this optimal combination is with respect to different meta-parameters, predictors, and reference data periods. Since it is impossible to test all possible applications, we will consider a few typical choices here and show for a few basins how robust our results are. We choose five river basins located within different continents and climate zones - i.e. the Amazon, Congo, Yangtze, Mississippi, and Danube basins - for the robustness tests.

We chose three sensitive climate drivers as the input for the prediction previously. For understanding how robust the number of selected sensitive climate drivers affects the prediction results, we now set this number from 1 to 15 in the prediction and use the CSR mascons to assess the standard error of testing total water storage change respectively. The standard error based on different numbers of inputs are shown in Figure C1. We find that the standard error of testing TWSC based on less than 8 (i.e. from 1 to 8) of inputs do not show large differences, but the standard error could increase when we use a larger number of inputs, e.g., more than 12 inputs in Congo basin. Thus, we do not suggest the users to set the number of inputs larger than 8 for the prediction when using our method. In the previous experiments we had chosen the number of selected modes by using a unified criterion, such as to retain 95% of the total energy; in what follows we will vary the number of modes between 3 and 10 to find how robust this criterion is. Figure C2 shows the standard errors of testing TWSC as compared to the CSR mascons and the results indicate that the uncertainty of predictions will be stable when choosing more than 5 modes for the prediction. As discussed in Humphrey et al. (2016), there is a few months of time delay between the climate changes and (affect) water storage changes. Our algorithm has been designed to maximize lag-correlation within a window of three months (0...3), in what follows we will extend this window successively to six months, e.g., if the window is set to four months then we will move the climate driver from 0 to 4 months to search for its highest correlation as related to the GRACE data. We predict the TWSC in five river basins based on different values of window (i.e. from 0-0 to 0-6) and then we estimate the standard errors of prediction using the CSR mascons as shown in Figure C3. The results indicate that the prediction uncertainties are reduced when we turning the time window from 0 to 3, and there is no obvious improvement when the window is increased from 3 to 6 months. For testing the robustness of the reference data period, we firstly fix the length of training section to eight years and we fix the length of the testing section to six years, i.e. we firstly use the GRACE data from April 2002 to March 2010 (totaling eight years) to train the predictive models and to predict the next six years (i.e. April 2010 - March 2016) of TWSC. Then we move the training and testing sections over time (e.g., move the training section from April 2002 – March 2010 to May 2002 – April 2010) and predict the TWSC using different periods of reference data respectively. Finally, we use the CSR mascons to evaluate the standard errors of the testing total water storage change in all moving time sections and we show them in Figure C4. The results indicate that the standard error of predictions from different time sections do not show large differences in five study regions, which demonstrating the robustness of our method for this variation.

References

- Anyah, R. O., Forootan, E., Awange, J., & Khaki, M. (2018). Understanding linkages between global climate indices and terrestrial water storage changes over africa using grace products. *Science of The Total Environment*, 635, 1405-1416.
- Baur, O. (2012). On the computation of mass-change trends from GRACE gravity field time-series. *Journal of Geodynamics*, 61, 120-128.
- Becker, M., Meyssignac, B., Xavier, L., Cazenave, A., & 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, 533-546.
- Berry, M. J., & Linoff, G. S. (2004). Data mining techniques: For marketing, sales, and customer relationship management. Wiley Computer Publishing.
- Bishop, C. M. (2006). Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York.
- Cardoso, J. F. (1992). Fourth-Order Cumulant Structure Forcing. Application to Blind Array Processing. *Proc. IEEE SP Workshop*, SSAP-92,136-139.
- Cardoso, JF. (1999) High-order contrasts for independent component analysis. *Neural Compute*, 11:157–192.
- Chandan, B., & Nagesh, K. D. (2018). Analyzing large-scale hydrologic processes using grace and hydrometeorological datasets. *Water Resources Management*, 32(13), 4409–4423.
- Chen, J. L., Tapley, B., Save, H., Tamisiea, M. E., Bettadpur, S., & Ries, J. (2018). Quantification of Ocean Mass Change Using Gravity Recovery and Climate Experiment, Satellite Altimeter, and Argo Floats Observat. *Journal of Geophysical Research Solid Earth*, 123(11), 10212-10225.
- Chen, J. L., Wilson, C. R., Tapley, B. D., Yang, Z. L., & Niu, G. Y. (2009). 2005 drought event in the amazon river basin as measured by grace and estimated by climate models. *Journal of Geophysical Research Solid Earth*, 114(B5), -.
- Chen, M., Shi, W., Xie, P., Silva, V. B. S., Kousky, V. E., & Higgins, R. W. (2008). Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research Atmospheres*, 113(D4), -.
- Church, J. A., White, N. J., Coleman, R., Lambeck, K., & Mitrovica, J. X. (2004). Estimates of the regional distribution of sea level rise over the 1950–2000 period. *Journal of Climate*, 17(13), 2609-2625.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E. (1990). Terpenning I. STL: a seasonal trend decomposition procedure based on loess. *J. Off. Stat.*, 6, 3-73.
- Challa, M. S. S., & Hetherington, J. H. (1988). Gaussian ensemble: an alternate monte carlo scheme. *Physical Review A*, 38(12), 6324-6337.
- Durbin, J., & Watson, G. S. (1992). Testing for Serial Correlation in Least Squares Regression. Breakthroughs in Statistics.
- Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., & Eicker, A. (2014), Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, *Water Resources Research*, 50, 5698–5720.
- Feng J., Han B., Zhao., & Wang Z. (2019). A New Global Total Electron Content Empirical Model. *Remote Sens.*, 11, 706.
- Frappart, Frédéric, Ramillien, G., & Ronchail, J. (2013). Changes in terrestrial water storage, versus rainfall and discharges in the amazon basin. *International Journal of Climatology*, 33(14), 3029-3046.
- Fan, Y., & Dool, H. V. D. (2004). Climate prediction center global monthly soil moisture data set at 0.5° resolution for 1948 to present. *Journal of Geophysical Research: Atmospheres*, 109(10), D10102.
- Frédéric, Frappart, Ramillien, G., Leblanc, M., Tweed, S. O., Bonnet, M. P., & Maisongrande, P. (2011). An independent component analysis filtering approach for estimating continental hydrology in the grace gravity data. *Remote Sensing of Environment*, 115(1), 187-204.
- Forman, B. A., Reichle, R. H., & Derksen, C.. (2014). Estimating passive microwave brightness temperature over snow-covered land in north america using a land surface

- model and an artificial neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1), 235-248.
- Forootan, E & Kusche, J. (2012). Separation of global time-variable gravity signals into maximally independent components. *Journal of Geodesy*, 86(7), 477-497.
- Forootan, E., Khaki, M., Schumacher, M., Wulfmeyer, V., Mehrnegar, N., & Van Dijk, A. I. J. M., et al. (2019). Understanding the global hydrological droughts of 2003–2016 and their relationships with teleconnections. *Science of The Total Environment*, 650, 2587-2604.
- Forootan, E., Kusche, J., Loth, I., Schuh, W.-D., Eicker, A., Awange, J., Longuevergne, L., Diekkrueger, B., Schmidt, M. and Shum, C.K. (2014) Multivariate prediction of total water storage anomalies over West Africa from multi-satellite data. *Surveys in Geophysics*, 35(4), 913–940.
- Haario, H., Laine, M., Mira, A., & Saksman, E. (2006). Dram: efficient adaptive MCMC. *Statistics & Computing*, 16(4), 339-354.
- Hannachi, A., Unkel, S., Trendafilov, N. T., & Jolliffe, I. T. (2009). Independent component analysis of climate data: a new look at eof rotation. *Journal of Climate*, 22(11), 2797-2812.
- Hassan, A. A., & Jin, S. (2014). Lake level change and total water discharge in east africa rift valley from satellite-based observations. *Global and Planetary Change*, 117, 79-90.
- Hsu K. L., Gupta H V , Sorooshian S . Artificial Neural Network Modeling of the Rainfall-Runoff Process[J]. *Water Resources Research*, 1995, 31(10):2517-2530.
- Hsu, C. W., & Velicogna, I. (2017). Detection of sea level fingerprints derived from grace gravity data. *Geophysical Research Letters*, 44, 8953–8961.
- Humphrey V., Gudmundsson L., Seneviratne S. I. (2016). Assessing Global Water Storage Variability from GRACE: Trends, Seasonal Cycle, Subseasonal Anomalies and Extremes. *Surveys in Geophysics*, 37(2), 357-395.
- Humphrey, V., Gudmundsson, L. (2019). GRACE-REC: a reconstruction of climate-driven water storage changes over the last century. *Earth Syst. Sci. Data Discuss*, 25.
- Humphrey, V., Gudmundsson, L., Seneviratne, S. I. (2017). A global reconstruction of climate-driven subdecadal water storage variability. *Geophysical Research Letters*, 44(3):2300–2309.
- Jäggi, A., Dahle, C., Arnold, D., Bock, H., Meyer, U., & Beutler, G., et al. (2016). Swarm kinematic orbits and gravity fields from 18months of gps data. *Advances in Space Research*, 57(1):218-233.
- Jeon, T., Seo, K. W., Youm K., Chen J., & Wilson C. R. (2019). Global sea level change signatures observed by GRACE satellite gravimetry. *Nature*, 8,13519.
- Juan, C., Jimenez, M., Jonathan, B., Andres, S. A., Ken, T., Tadvinder, M., Jose, A. S., & Gerard, V. D. S. (2016) Record-breaking warming and extreme drought in the Amazon rainforest during the course of El Niño 2015–2016. *Scientific reports*, 10, 1-6.
- Khaki, M., Hoteit, I., Kuhn, M., Awange, J., Forootan, E., & Van Dijk, A. I. J. M., et al. (2017). Assessing sequential data assimilation techniques for integrating grace data into a hydrological model. *Advances in Water Resources*, 107, 301-316.
- Khan, S. A., Wahr, J., Bevis, M., Velicogna, I., & Kendrick, E. (2010). Spread of ice mass loss into northwest greenland observed by GRACE and GPS. *Geophysical Research Letters*, 37, L06501.
- Lück, C., Kusche, J., Rietbroek, R., Löcher, A. (2018). Time-variable gravity fields and ocean mass change from 37 months of kinematic swarm orbits. *Solid Earth*, 9(2), 323-339.
- Landerer, F. W., Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, 48, W04531.
- Leblanc, M. J., Tregoning, P., Ramillien, G., Tweed, S. O., & Fakes, A. (2009). Basin-scale, integrated observations of the early 21st century multiyear drought in southeast Australia. *Water Resources Research*, 45, W04408.
- Li F., Wang Z., Chao N., & Song, Q. (2018). Assessing the Influence of the Three Gorges Dam on Hydrological Drought Using GRACE Data. *Water*, 10, 669.
- Ljung, L. (2002). System Identification: Theory for The User. Tsinghua University Press.
- 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 Sensing of Environment*, 155, 145-160.
- Lucas Schreiter, Daniel Arnold, Veerle Sterken, and Adrian Jäggi. (2019). Mitigation of ionospheric signatures in Swarm GPS gravity field estimation using weighting strategies.

- Ann. Geophys.*, (37), 111-127.
- Mnhajeran, Y., Velicogna, Rignot, E. (2018). Mass Loss of Totten and Moscow University Glaciers, East Antarctica, Using Regionally Optimized GRACE Mascons. *Geophysical Research Letters*, 45(14), 7010-2018.
- Myers, R. H. (1986). Classical and modern regression with application. Classical and modern regression with applications. Duxbury Press.
- Nerem, R., Talpe, M., & Pilinski, E. B., et al. (2012). Reconstruction of Temporal Gravity Variations Prior to the GRACE Mission. *AGU Fall Meeting*. AGU Fall Meeting Abstracts.
- Peralta-Ferriz, C., Woodgate, R. A. (2017). The Dominant Role of the East Siberian Sea in Driving the Oceanic Flow Through the Bering Strait-Conclusions from GRACE Ocean Mass Satellite Data and In Situ Mooring Observations Between 2002 and 2016. *Geophysical Research Letters*, 44, 11472–11481
- Ran, J., Ditmar, P., Klees, R., & Farahani, H. H. (2018). Statistically optimal estimation of greenland ice sheet mass variations from grace monthly solutions using an improved mascon approach. *Journal of Geodesy*, 92(3), 299-319.
- Rao, S. G., Rao, A. R. (2000). Artificial Neural Networks in Hydrology. *Journal of Hydrologic Engineering*, 5(2), 124-137.
- Reichle, R. H., Koster, R. D., Dong, J., & Berg, A. A.. (2004). Global soil moisture from satellite observations, land surface models, and ground data: implications for data assimilation. *Journal of Hydrometeorology*, 5(3), 430-442.
- Reichle, R. H., Rodell, M., Zaitchik, B. F. (2008). Assimilation of GRACE Terrestrial Water Storage Data into a Land Surface Model: Results for the Mississippi River Basin[J]. *Journal of Hydrometeorology*, 9(3), 535-548.
- Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes D.C., & Wang, W. (2002) An improved in situ and satellite SST analysis for climate. *J. Climate*, 15, 1609-1625.
- Rietbroek, R., Fritsche, M., Dahle, C., Brunnabend, S.-E., Behnisch, M., Kusche, J., Flechtner, F., Schröter, J., Dietrich, R. (2014) Can GPS-Derived Surface Loading Bridge a GRACE Mission Gap?, *Surveys in Geophysics*, 1–17.
- Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D., Toll, D. (2004). The global land data assimilation system. *Bull. Am. Meteorol. Soc.*, 85, 381–394.
- Frappart, Frédéric, Papa, F., Joecila, S. D. S., Ramillien, G., Prigent, C., & Seyler, Frédérique, et al. (2012). Surface freshwater storage and dynamics in the amazon basin during the 2005 exceptional drought. *Environmental Research Letters*, 7(4), 044010.
- Save, H., Bettadpur, S., & Tapley, B. D. (2016), High resolution CSR GRACE RL05 mascons, *J. Geophys. Res. Solid Earth*, 121, 7547–7569.
- Save, Himanshu, 2019, "CSR GRACE RL06 Mascon Solutions", <https://doi.org/10.18738/T8/UN91VR>, Texas Data Repository Dataverse, V1.
- Sheffield, J., Ferguson, C. R., Troy, T. J., Wood, E. F., & McCabe, M. F. (2009). Closing the terrestrial water budget from satellite remote sensing. *Geophysical Research Letters*, 36(7), L07403.
- Bezděk, A., Sebera, J., Encarnação J. T. D., & Jaroslav, K. (2016). Time-variable gravity fields derived from GPS tracking of Swarm. *Geophysical Journal International*, 205(3), ggw094.
- Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, 22(1), 97-103.
- 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 Resources Research*, 55, 1179–1195.
- Sun, Alexander, Y. (2013). Predicting groundwater level changes using GRACE data[J]. *Water Resources Research*, 49(9), 5900-5912.
- Summers, D. J. (1985). Reconstruction of a strip geometry calorimeter using stepwise regression. *Nuclear Instruments & Methods in Physics Research, Section A, (Accelerators, Spectrometers, Detectors and Associated Equipment)*, 228(2-3), 290-293.
- Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J., & Wilson, C. R. (2008). Analysis of terrestrial water storage changes from grace and gldas. *Water Resources Research*, 44(2),

339-356.

- Eicker, A., Schumacher, M., Kusche, J., Döll, P., Müller Schmied, H. (2014) Calibration/Data Assimilation Approach for Integrating GRACE Data into the WaterGAP Global Hydrology Model (WGHM) Using an Ensemble Kalman Filter: First Results. *Surveys in Geophysics*, 35(6), 1285–1309.
- Talpe, M. J., Nerem, R. S., Forootan, E., Schmidt, M., Lemoine, F. G., & Enderlin, E. M., et al. (2017). Ice mass change in greenland and antarctica between 1993 and 2013 from satellite gravity measurements. *Journal of Geodesy*, 91(11), 1283-1298.
- Tangdamrongsub, N., Steele-Dunne, S. C., Gunter, B. C., Ditmar, P. G., & Weerts, A. H. (2015). Data assimilation of grace terrestrial water storage estimates into a regional hydrological model of the rhine river basin. *Hydrology and Earth System Sciences*, 19(4), 2079-2100.
- Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The gravity recovery and climate experiment: mission overview and early results. *Geophysical Research Letters*, 31, L09607.
- Tapley, B. D., Watkins, M. M., & Flechtner, F., et al. (2019). Contributions of GRACE to understanding climate change. *Nature*, 9, 358–369.
- Thomas, A. C., Reager, J. T., Famiglietti, J. S., & Rodell, M. (2014). A grace-based water storage deficit approach for hydrological drought characterization. *Geophysical Research Letters*, 41(5), 1537-1545.
- Teixeira Encarnação, J., Visser, P., Arnold, D., Bezdek, A., Doornbos, E., Ellmer, M., Guo, J., van den IJssel, J., Iorfida, E., Jaggi, A., Klokocnik, J., Krauss, S., Mao, X., Mayer-Gurr, T., Meyer, U., Sebera, J., Shum, C. K., Zhang, C., Zhang, Y., and Dahle, C.: Multi-approach gravity field models from Swarm GPS data. *Earth Syst. Sci. Data Discuss.* 2019.
- Uebbing B., Kusche J., Rietbroek R., & Landerer F. W. (2019). Processing Choices Affect Ocean Mass Estimates from GRACE. *Journal of Geophysical Research: Oceans*, 124(2), 1029-1044.
- Van, D., Renzullo, L. J., Wada, Y., & Tregoning, P. (2014). A global water cycle reanalysis (2003–2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble. *Hydrology and Earth System Sciences*, 18(8), 2955-2973.
- Velicogna, I. (2009). Increasing rates of ice mass loss from the Greenland and Antarctic ice sheets revealed by GRACE. *Geophysical Research Letters*, 36(19), 158-168.
- Velicogna, I., Sutterley T.C., & van den Broeke, M. R. (2014), Regional acceleration in ice mass loss from Greenland and Antarctica using GRACE time-variable gravity data. *Geophysical Research Letters*, 41, 8130–8137
- Wold, S. (1987). Principal component analysis. *Chemometrics & Intelligent Laboratory Systems*, 2(1), 37-52.
- Wu, Xiaoping. (2003). Large-scale global surface mass variations inferred from GPS measurements of load-induced deformation. *Geophysical Research Letters*, 30(14), 1742.
- Yirdaw, S. Z., Snelgrove, K. R., Agboma, C. O. (2008). GRACE satellite observations of terrestrial moisture changes for drought characterization in the Canadian Prairie. *Journal of Hydrology*, 356, 84-92.

Table 1

Classification of the Most Sensitive Climate Drivers for GRACE Seasonal, Inter-annual, and Residual Components in Twenty-six River Basins. SST=sea surface temperature; P=precipitation; T=land surface temperature

Basin ID	Name	Most sensitive climate drivers for GRACE seasonal component	Most sensitive climate drivers for GRACE inter-annual component	Most sensitive climate drivers for GRACE residual component
1	Yukon	SST	SST	SST
2	Mackenzie	SST	SST	SST
3	Nelson	SST	p	SST
4	Mississippi	SST	P	P
5	St Lawrence	SST	SST	P
6	Amazon	SST	SST	P
7	Parana	SST	P	P
8	Niger	SST	SST	SST
9	Lake Chad	SST	SST	SST
10	Nile	SST	SST	SST
11	Congo	SST	SST	SST
12	Zambezi	SST	SST	P
13	Orange	SST	SST	SST
14	Danube	SST	P	P
15	Euphrates	SST	P	P
16	Volga	SST	SST	P
17	Ob	SST	P	P
18	Yenisey	SST	SST	SST
19	Lena	SST	SST	SST
20	Kolyma	SST	SST	SST
21	Amur	SST	P	P
22	Huang He	SST	SST	P
23	Yangtze	SST	P	P
24	Ganges	SST	T	P
25	Indus	SST	SST	P
26	Murray Darling	SST	P	P

Table 2

Standard Errors of Training and Testing TWSC at Both Grid and Basin Scales by Using Three Predictive Models in Twenty-six River Basins as Compared to CSR Mascons. Noting that All Results Listed in This Table are Calculated by Fixing The Spatio-temporal Decomposition and Time-series Decomposition Techniques to PCA and LS.

Basin	Training						Testing					
	MLR(cm)		ANN(cm)		ARX(cm)		MLR(cm)		ANN(cm)		ARX(cm)	
	<i>gri</i> <i>d</i>	<i>basi</i> <i>n</i>										
Yukon	2.0	1.2	1.7	1.0	1.8	1.0	2.2	1.3	2.3	1.4	2.8	1.6
Mackenzie	1.5	0.7	1.2	0.5	1.3	0.7	1.9	1.0	2.1	1.0	2.2	1.2
Nelson	2.4	1.8	1.7	1.0	1.4	0.7	2.5	1.7	3.0	2.3	2.9	2.2
Mississippi	2.5	1.1	2.4	0.7	2.3	1.3	2.8	1.7	3.2	1.6	3.0	1.9
St Lawrence	2.4	2.0	1.8	0.9	2.1	1.5	3.7	3.0	4.1	3.4	4.8	5.2
Amazon	5.4	2.0	4.8	1.5	5.4	2.9	7.1	3.9	7.8	4.2	7.8	4.6
Parana	3.7	1.5	3.1	1.2	3.1	1.5	4.9	2.0	5.4	2.1	4.0	1.9
Niger	1.4	0.9	1.2	0.7	1.5	1.0	1.8	1.0	1.7	1.0	2.1	1.3
Lake Chad	1.1	0.7	0.9	0.5	1.0	0.5	1.5	1.0	1.4	0.9	2.0	1.6
Nile	1.9	0.9	1.4	0.6	1.7	0.6	2.2	1.2	2.7	1.6	3.1	2.3
Congo	3.1	1.5	2.5	1.0	3.0	1.2	3.4	1.7	3.6	1.7	5.0	3.6
Zambezi	4.8	3.2	3.3	1.2	3.3	1.8	5.2	3.4	6.3	4.2	5.9	4.0
Orange	1.4	1.1	1.1	0.8	0.9	0.6	1.5	1.1	1.5	1.1	1.4	1.0
Danube	2.4	1.6	2.0	1.3	1.8	0.9	2.8	2.0	3.8	2.7	3.0	1.9
Euphrates	2.1	1.7	1.7	1.3	1.3	1.0	3.3	2.7	3.0	2.3	3.2	2.5
Volga	2.2	1.4	1.7	0.9	2.0	1.4	3.8	3.2	3.9	3.0	4.5	2.9
Ob	2.0	1.0	1.8	1.0	2.3	1.8	3.1	2.4	3.5	2.7	3.6	2.8
Yenisey	2.5	1.6	2.0	1.2	2.2	1.5	2.9	1.9	3.2	2.0	2.9	1.6
Lena	1.6	1.0	1.1	0.5	1.3	0.8	2.5	2.1	2.5	1.9	2.2	1.6
Kolyma	2.0	1.6	1.1	0.7	2.0	1.8	2.1	1.6	2.7	2.3	3.3	3.1
Amur	1.6	0.8	1.4	0.7	1.3	0.6	2.0	1.3	2.2	1.3	2.7	1.4
Huang He	1.3	0.8	1.1	0.5	1.0	0.4	1.8	1.1	2.1	1.4	1.7	1.0
Yangtze	2.2	0.9	1.9	0.8	1.9	0.7	2.6	1.3	2.9	1.4	3.1	1.6
Ganges	3.5	1.6	3.0	1.1	3.2	1.5	4.2	2.0	4.7	2.4	5.1	2.6
Indus	2.2	1.4	1.8	1.0	1.6	0.9	2.3	1.4	2.8	1.7	2.6	1.7
Murray	2.1	1.8	1.2	0.8	1.2	0.8	2.7	2.4	3.3	3.1	3.2	2.7
Darling												
Average	2.3	1.3	1.8	0.9	1.9	1.1	3.0	1.9	3.3	2.1	3.4	2.3

Table 3

Correlation Coefficients of Training and Testing TWSC at Both Grid and Basin Scales by Using Three Predictive Models in Twenty-six River Basins as Compared to CSR Mascons. Noting that All Results Listed in This Table are Calculated by Fixing The Spatio-temporal Decomposition and Time-series Decomposition Techniques to PCA and LS.

Basin	Training						Testing					
	MLR(cm)		ANN(cm)		ARX(cm)		MLR(cm)		ANN(cm)		ARX(cm)	
	<i>grid</i>	<i>basin</i>										
Yukon	0.92	0.97	0.93	0.98	0.93	0.98	0.91	0.96	0.90	0.96	0.87	0.94
Mackenzie	0.93	0.98	0.96	0.99	0.95	0.98	0.90	0.96	0.87	0.96	0.86	0.95
Nelson	0.84	0.88	0.93	0.96	0.95	0.98	0.81	0.85	0.69	0.73	0.67	0.76
Mississippi	0.90	0.97	0.94	0.99	0.93	0.96	0.86	0.93	0.81	0.94	0.84	0.92
St Lawrence	0.85	0.89	0.94	0.97	0.92	0.93	0.76	0.87	0.77	0.87	0.68	0.53
Amazon	0.94	0.99	0.95	0.99	0.94	0.98	0.93	0.97	0.91	0.96	0.91	0.95
Parana	0.88	0.96	0.91	0.98	0.91	0.96	0.77	0.93	0.71	0.94	0.78	0.94
Niger	0.88	0.99	0.89	0.99	0.89	0.99	0.82	0.99	0.83	0.99	0.83	0.98
Lake Chad	0.75	0.98	0.80	0.99	0.79	0.98	0.68	0.96	0.69	0.97	0.62	0.89
Nile	0.87	0.97	0.91	0.99	0.89	0.98	0.86	0.95	0.83	0.91	0.82	0.84
Congo	0.95	0.93	0.96	0.97	0.96	0.96	0.93	0.92	0.93	0.91	0.86	0.73
Zambezi	0.90	0.96	0.95	0.99	0.96	0.99	0.92	0.96	0.90	0.96	0.88	0.94
Orange	0.83	0.85	0.89	0.92	0.94	0.96	0.80	0.88	0.77	0.84	0.79	0.87
Danube	0.94	0.96	0.96	0.97	0.96	0.99	0.91	0.95	0.83	0.91	0.88	0.95
Euphrates	0.92	0.96	0.95	0.98	0.97	0.99	0.77	0.89	0.80	0.89	0.84	0.89
Volga	0.95	0.98	0.97	0.99	0.96	0.98	0.87	0.90	0.88	0.93	0.83	0.91
Ob	0.93	0.98	0.94	0.98	0.91	0.93	0.86	0.93	0.85	0.91	0.82	0.90
Yenisey	0.87	0.94	0.91	0.96	0.91	0.94	0.85	0.93	0.82	0.92	0.82	0.94
Lena	0.92	0.96	0.96	0.99	0.95	0.97	0.85	0.83	0.83	0.91	0.82	0.89
Kolyma	0.88	0.92	0.96	0.98	0.88	0.89	0.90	0.95	0.83	0.88	0.80	0.80
Amur	0.86	0.93	0.90	0.95	0.90	0.96	0.85	0.90	0.81	0.88	0.67	0.88
Huang He	0.82	0.93	0.87	0.96	0.88	0.96	0.62	0.79	0.51	0.69	0.66	0.81
Yangtze	0.88	0.96	0.91	0.96	0.91	0.97	0.81	0.90	0.78	0.86	0.75	0.87
Ganges	0.94	0.99	0.95	0.99	0.95	0.99	0.92	0.98	0.90	0.97	0.89	0.97
Indus	0.87	0.93	0.91	0.96	0.93	0.97	0.87	0.91	0.83	0.91	0.82	0.85
Murray Darling	0.86	0.88	0.95	0.98	0.96	0.97	0.66	0.81	0.53	0.62	0.72	0.83
Average	0.89	0.95	0.93	0.98	0.92	0.97	0.84	0.92	0.80	0.89	0.80	0.87

Table 4

Correlation Coefficients of Training and Testing TWSC after Removing the Seasonal Cycle at Both Grid and Basin Scales by Using Three Predictive Models in Twenty-six River Basins as Compared to CSR Mascons. Noting that All Results Listed in This Table are Calculated by Fixing The Spatio-temporal Decomposition and Time-series Decomposition Techniques to PCA and LS.

Basin	Training						Testing					
	MLR(cm)		ANN(cm)		ARX(cm)		MLR(cm)		ANN(cm)		ARX(cm)	
	<i>grid</i>	<i>basin</i>										
Yukon	0.65	0.85	0.75	0.89	0.74	0.86	0.60	0.65	0.57	0.59	0.49	0.52
Mackenzie	0.74	0.84	0.85	0.94	0.81	0.86	0.60	0.64	0.47	0.62	0.48	0.48
Nelson	0.78	0.81	0.91	0.94	0.93	0.97	0.65	0.67	0.39	0.27	0.42	0.49
Mississippi	0.80	0.90	0.89	0.96	0.85	0.88	0.75	0.82	0.66	0.86	0.71	0.77
St Lawrence	0.71	0.78	0.88	0.95	0.83	0.88	0.43	0.79	0.47	0.80	0.41	0.19
Amazon	0.71	0.84	0.74	0.91	0.70	0.75	0.66	0.72	0.60	0.73	0.57	0.63
Parana	0.75	0.89	0.84	0.94	0.83	0.90	0.65	0.90	0.54	0.92	0.64	0.91
Niger	0.75	0.90	0.79	0.93	0.76	0.87	0.64	0.87	0.65	0.87	0.61	0.79
Lake Chad	0.68	0.89	0.75	0.94	0.72	0.93	0.56	0.80	0.69	0.97	0.62	0.89
Nile	0.70	0.87	0.80	0.94	0.75	0.95	0.66	0.76	0.62	0.66	0.52	0.40
Congo	0.75	0.86	0.84	0.95	0.79	0.92	0.70	0.80	0.70	0.80	0.53	0.54
Zambezi	0.71	0.86	0.86	0.98	0.86	0.96	0.68	0.88	0.58	0.85	0.59	0.82
Orange	0.78	0.84	0.86	0.91	0.92	0.96	0.66	0.88	0.65	0.83	0.68	0.87
Danube	0.84	0.90	0.90	0.94	0.90	0.96	0.75	0.85	0.55	0.73	0.67	0.83
Euphrates	0.84	0.89	0.89	0.92	0.93	0.98	0.49	0.54	0.45	0.35	0.49	0.38
Volga	0.84	0.90	0.91	0.95	0.86	0.91	0.63	0.61	0.67	0.70	0.46	0.60
Ob	0.76	0.91	0.81	0.92	0.71	0.72	0.60	0.83	0.51	0.80	0.50	0.76
Yenisey	0.72	0.67	0.81	0.82	0.79	0.73	0.73	0.74	0.66	0.67	0.69	0.75
Lena	0.83	0.90	0.92	0.97	0.90	0.94	0.59	0.39	0.56	0.63	0.53	0.51
Kolyma	0.69	0.75	0.91	0.95	0.71	0.69	0.76	0.87	0.58	0.64	0.47	0.33
Amur	0.82	0.92	0.87	0.94	0.87	0.95	0.82	0.89	0.77	0.87	0.63	0.86
Huang He	0.77	0.91	0.83	0.95	0.85	0.95	0.47	0.70	0.37	0.60	0.48	0.71
Yangtze	0.71	0.87	0.79	0.89	0.80	0.89	0.51	0.77	0.45	0.68	0.45	0.74
Ganges	0.79	0.91	0.84	0.95	0.83	0.93	0.68	0.88	0.65	0.85	0.62	0.83
Indus	0.79	0.87	0.86	0.94	0.88	0.94	0.72	0.78	0.64	0.77	0.62	0.67
Murray Darling	0.84	0.87	0.95	0.97	0.95	0.97	0.49	0.74	0.35	0.51	0.56	0.81
Average	0.76	0.86	0.85	0.93	0.83	0.89	0.63	0.76	0.57	0.71	0.56	0.66

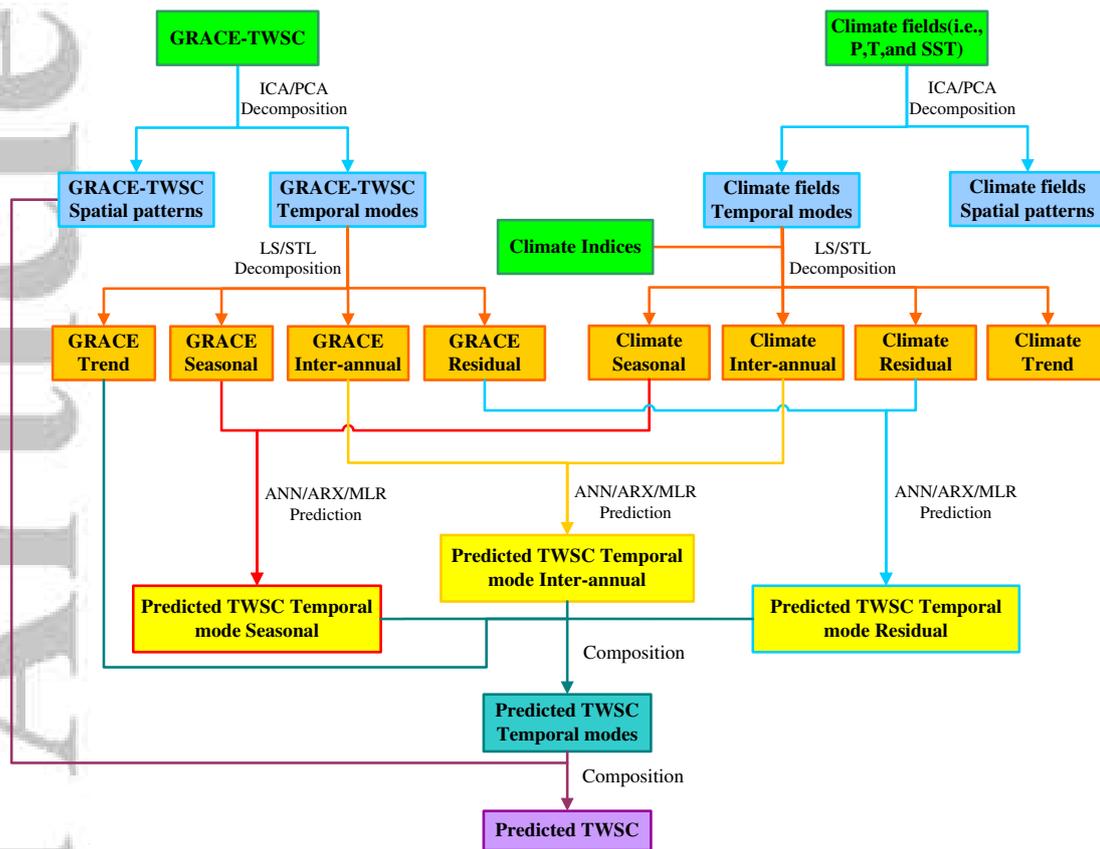


Figure 1. Illustration of the data flow in our unified framework for comparing different data-driven methods. TWSC means total water storage change; P for precipitation; T for land surface temperature; and SST for sea surface temperature. ICA/PCA are independent and principal component analysis techniques; LS/STL are least squares and seasonal-trend decomposition based on loess procedure; and ANN/ARX/MLR are artificial neural network, autoregressive exogenous, and multiple linear regression models.

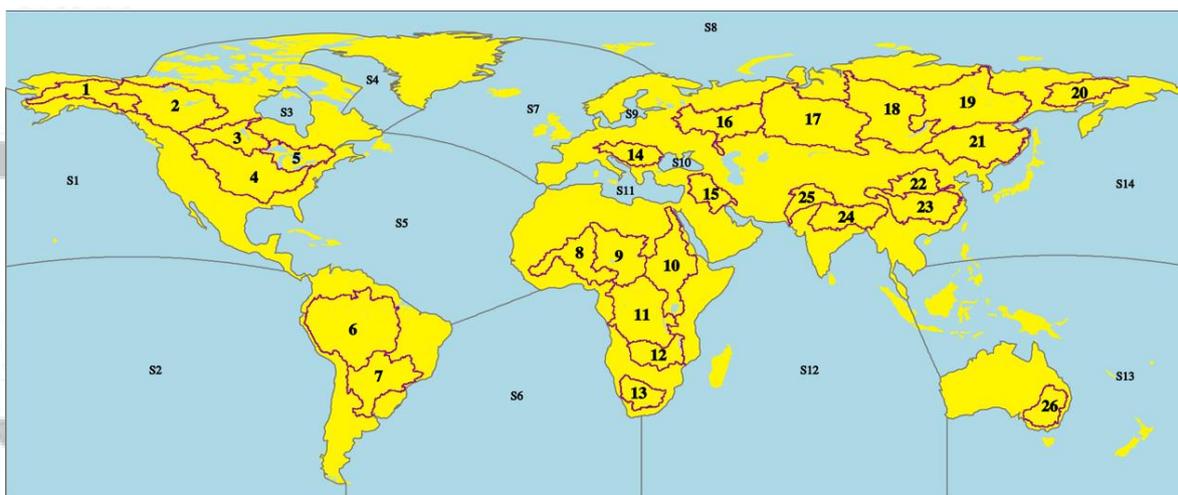


Figure 2. The study regions (i.e. twenty-six river basins) and divided oceans and seas. The river basins are numbered from 1 to 26 and the divided oceans or seas are numbered from S1 to S14 as shown in this figure.

Accepted Article

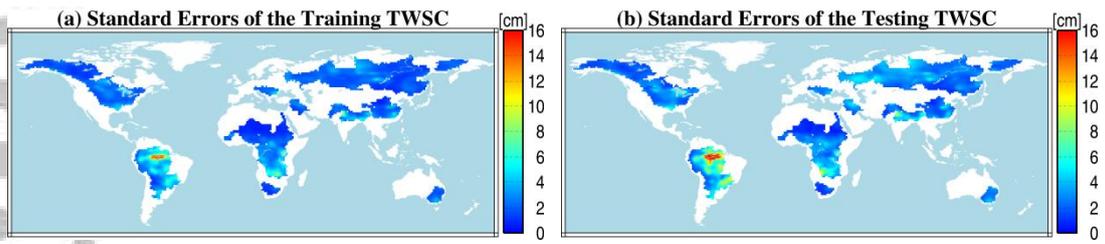


Figure 3. The standard errors of (a) training TWSC (i.e. April 2002 - June 2011), and (b) testing TWSC (over July 2011 - June 2017) based on the PCA, LS, and MLR methods at each grid as evaluated by the CSR mascons.

Accepted Article

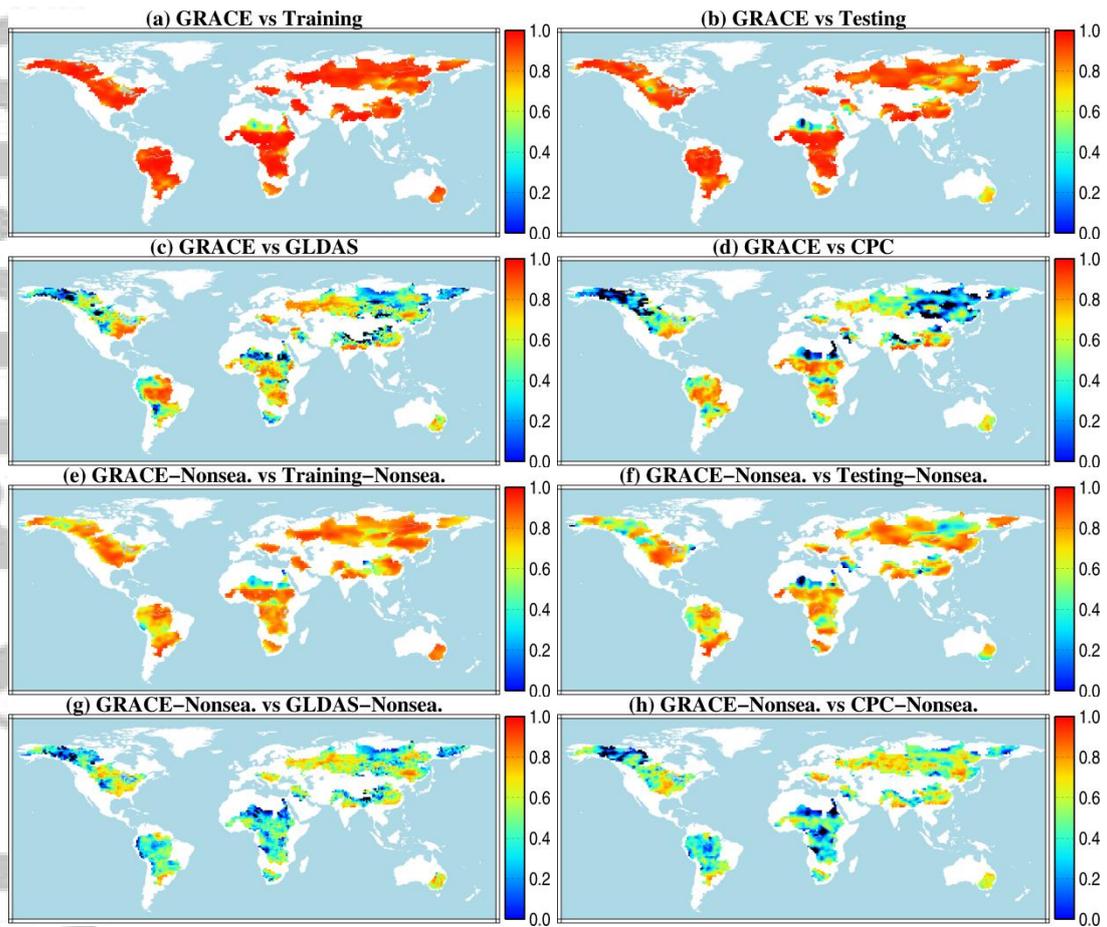


Figure 4. Correlations between the CSR mascons and (a) training TWSC, (b) testing TWSC, (c) GLDAS TWSC, and (d) CPC TWSC over twenty-six river basins. (e)-(h) are the correlations computed by the related de-seasoned signals.

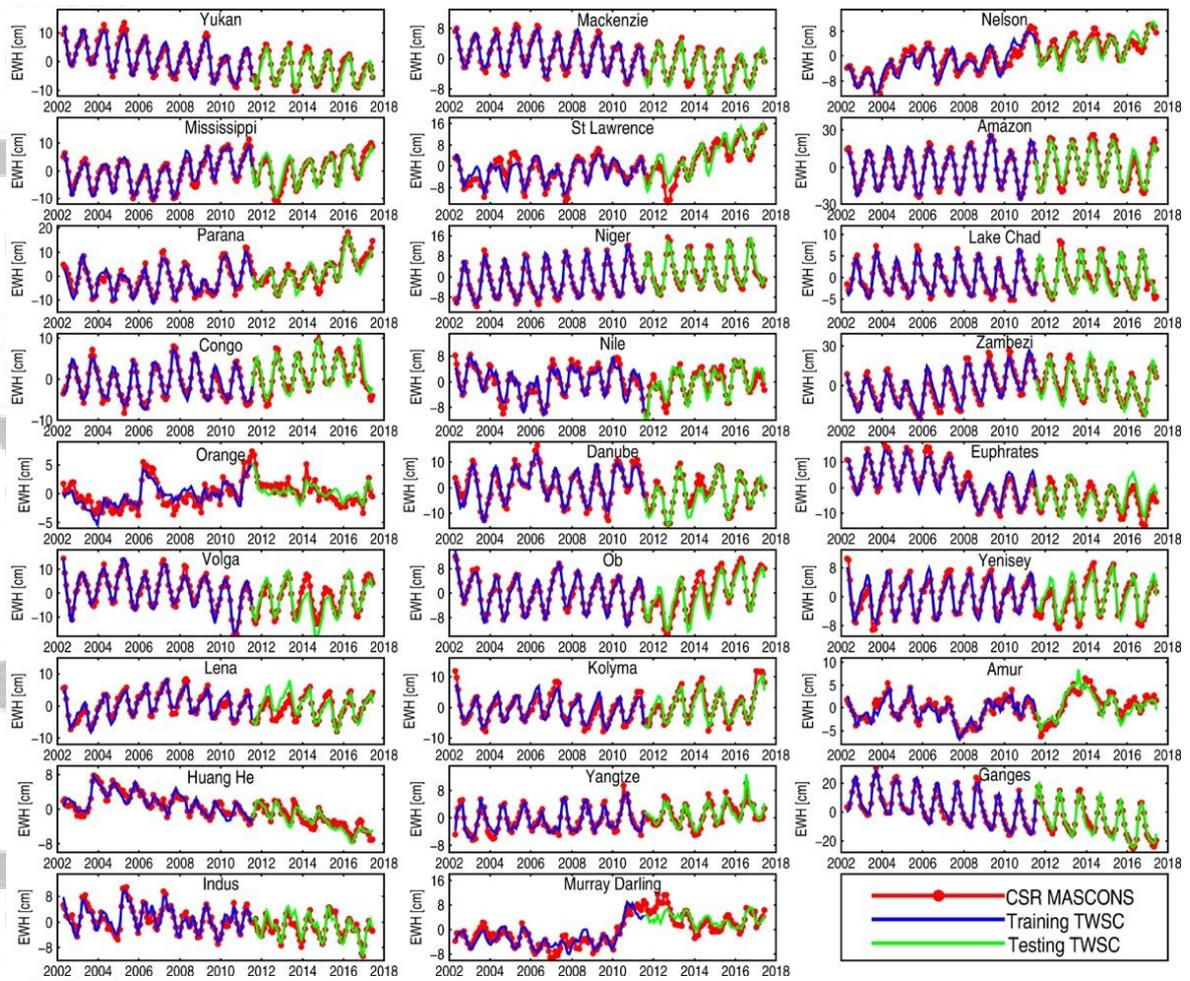


Figure 5. The training TWSC (blue line) and testing TWSC (green line) relative to the CSR mascons (red line) at the basin scale for twenty-six river basins.

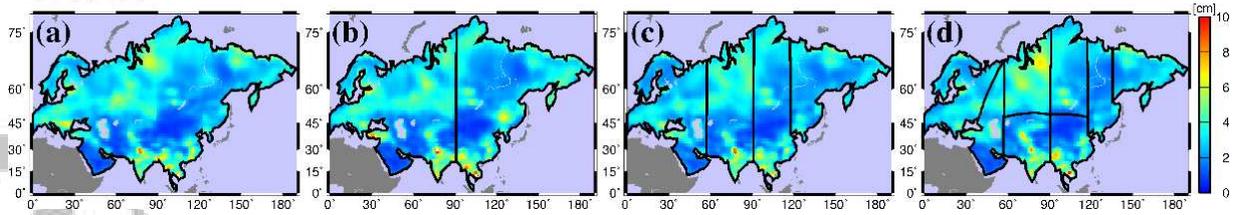


Figure 6. Standard errors of the testing TWSC (over July 2011 - June 2017) by dividing the Asia-Europe continent into (a) one, (b) two, (c) four, and (d) eight parts as evaluated by the CSR mascons. Here we first divide the continent into one, two, four, and eight parts respectively, and then we apply the identified methods to each divided part and use the CSR mascons to estimate the prediction uncertainties as shown in this figure.

Accepted Article

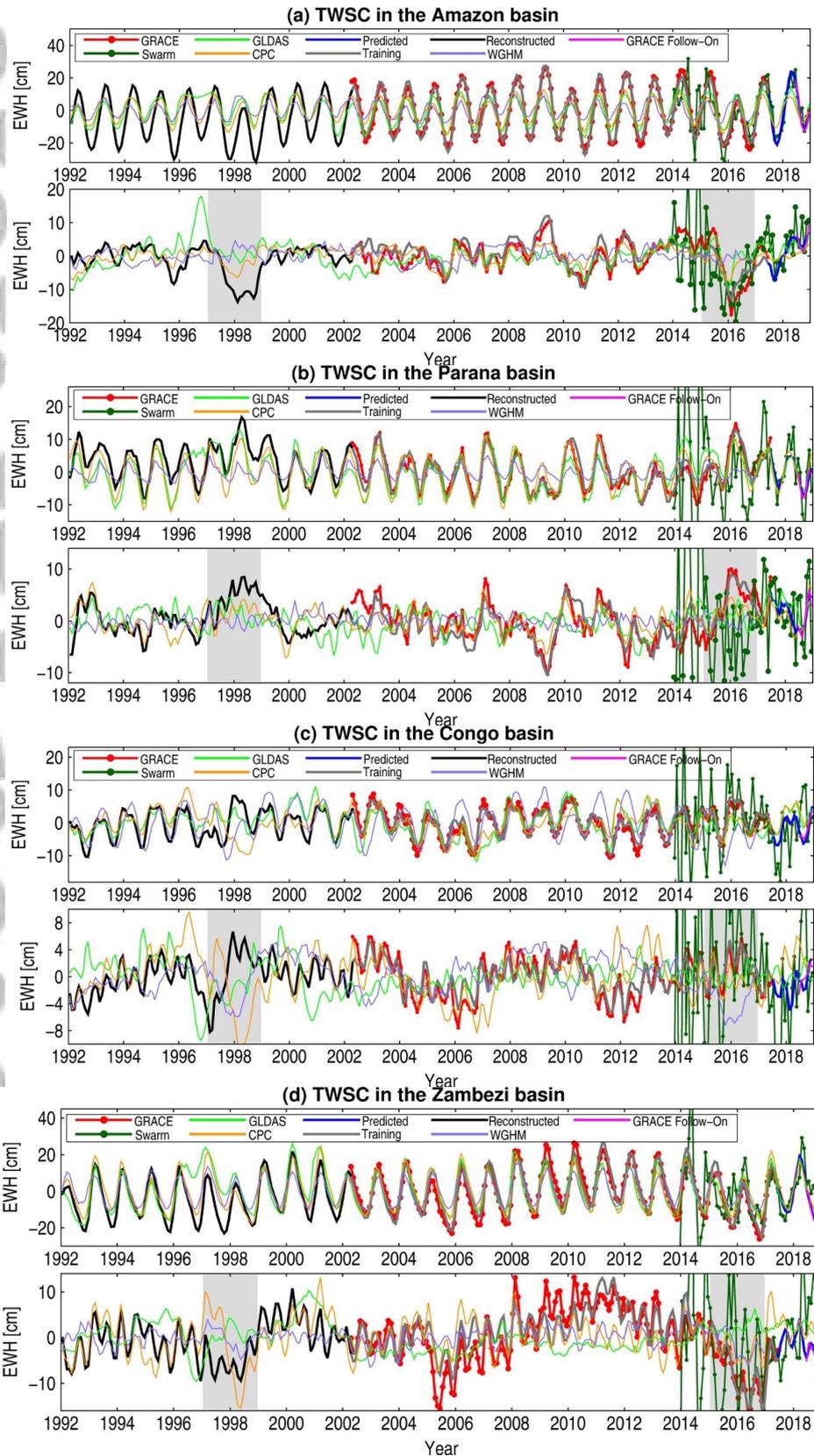


Figure 7. The TWSC (up) and de-seasoned TWSC (down) time series at the basin scale in the (a) Amazon, (b) Parana, (c) Congo, and (d) Zambezi basins. Gray phases represent the strong El Niño years (i.e. 1997/1998 and 2015/2016). For a fair comparison, the linear trend of each time series (except the GRACE-FO) has been removed.

Appendix A: Figures

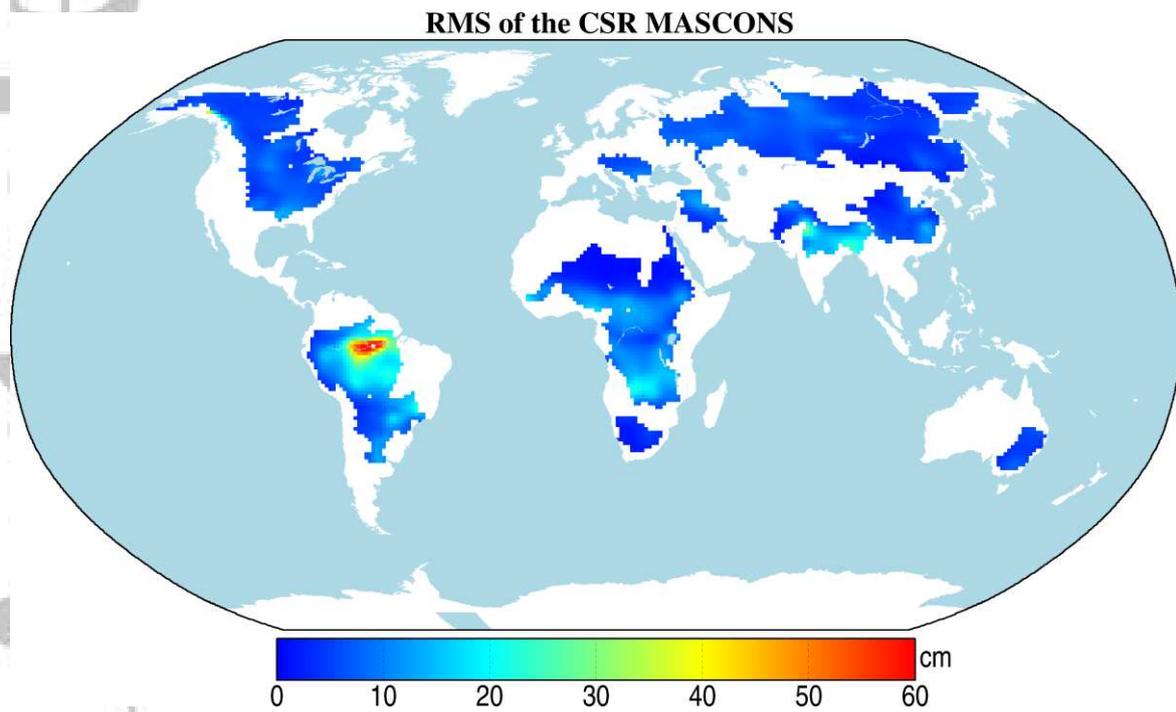


Figure A1. The RMS of mean-centered CSR mascons (from April 2002 to June 2017) over twenty-six river basins.

Accepted

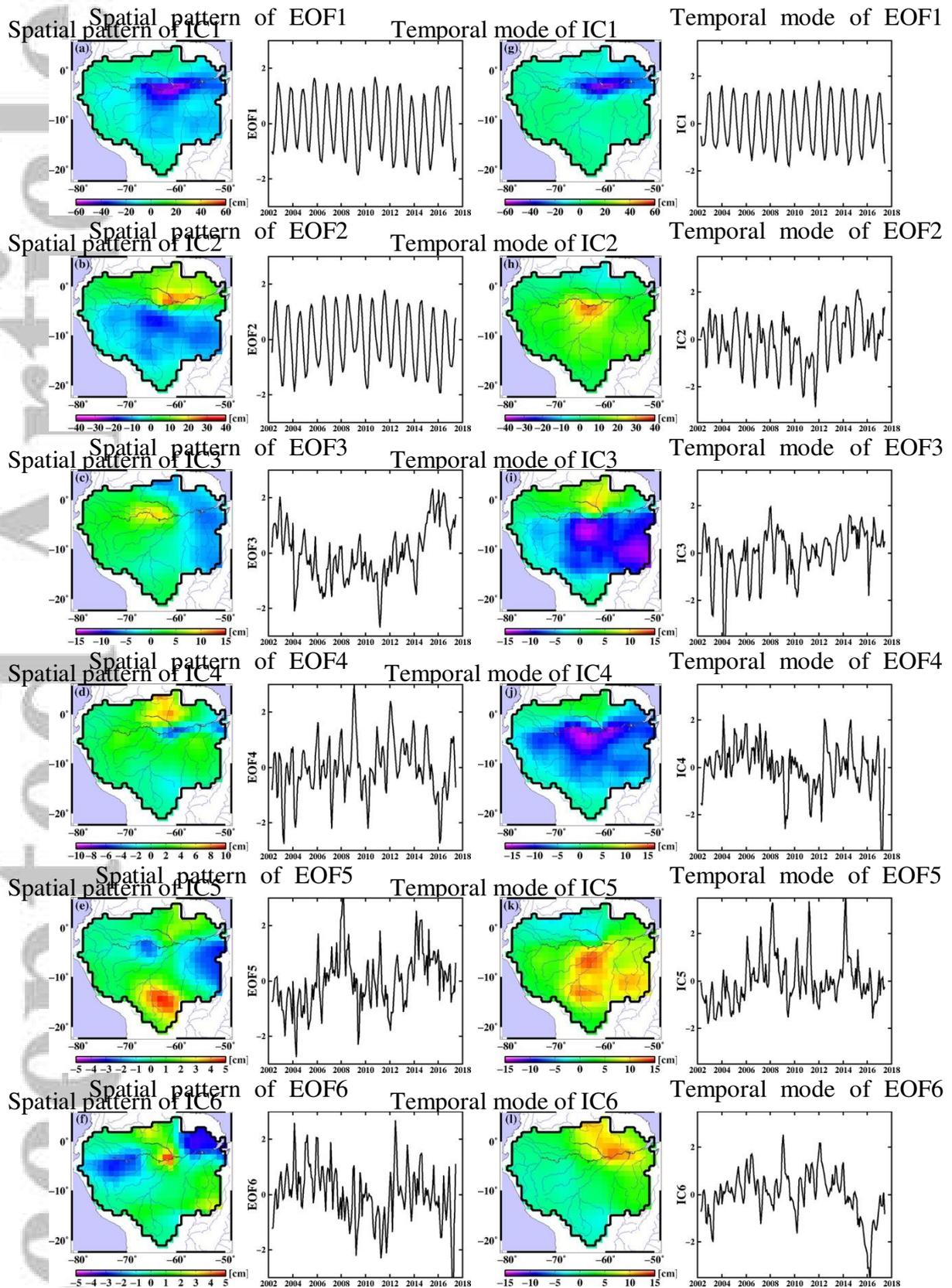


Figure A2. Spatial patterns and temporal modes of identified GRACE TWSC EOFs and ICs in the Amazon basin. Results on the maps represent the spatial modes and the time series on the right side of each map represents the corresponding temporal mode.

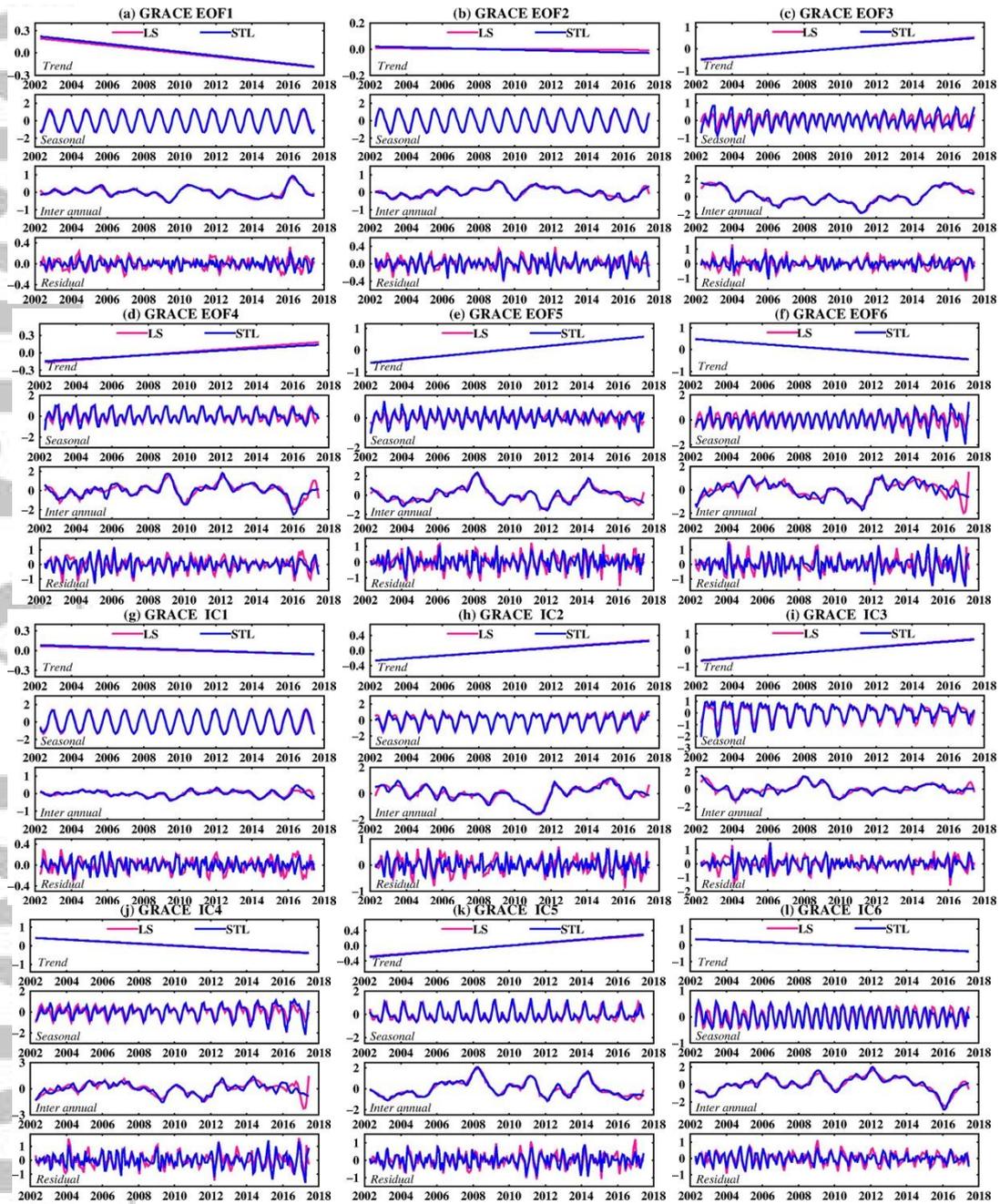


Figure A3. Decomposed components (i.e. linear, seasonal, inter-annual, and residual) of identified GRACE TWSC EOF and IC temporal modes in the Amazon basin based on the LS and STL methods. The red and blue lines represent the components decomposed by LS and STL methods, respectively.

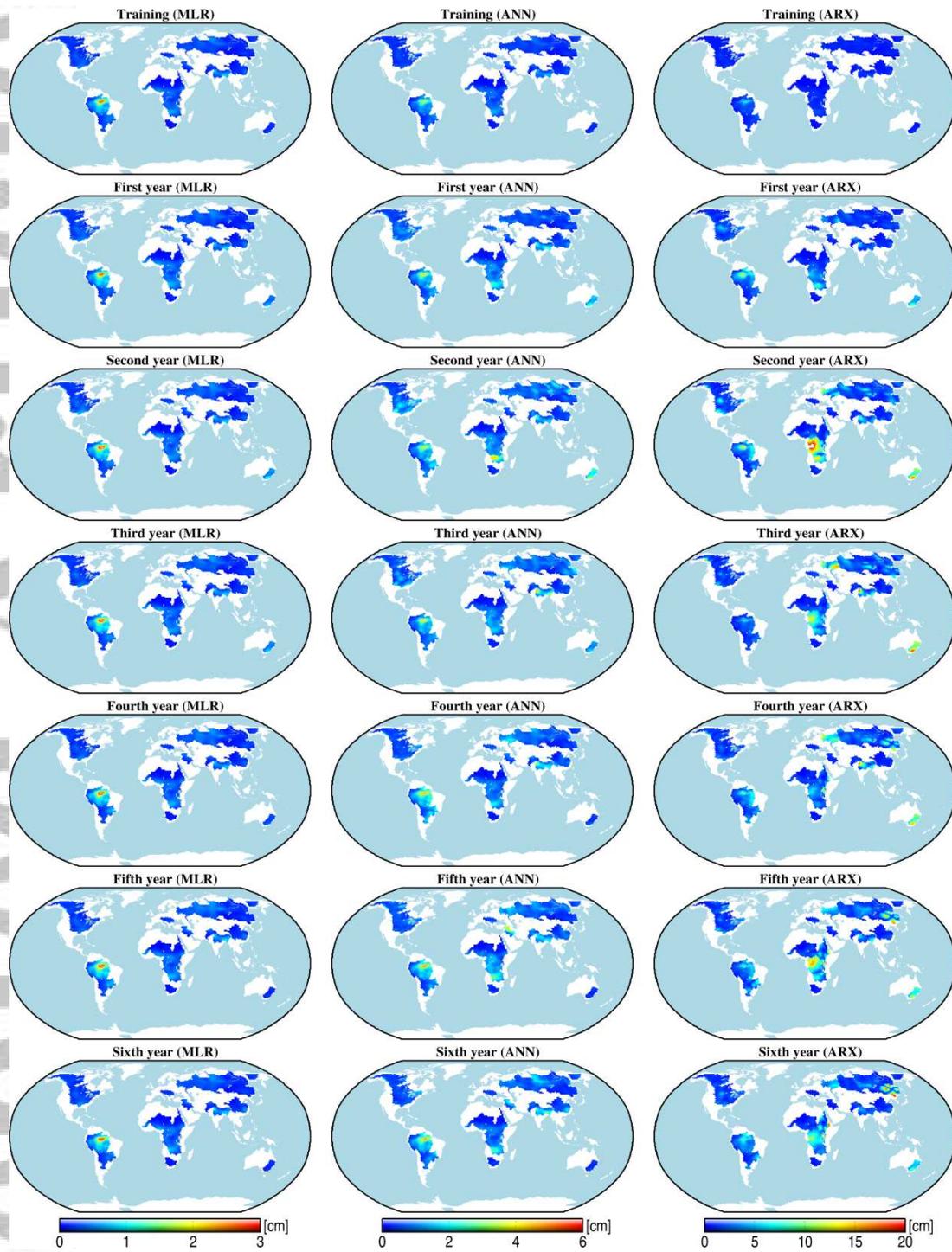


Figure A4. Error propagation in the three predictive models assessed using Monte Carlo uncertainties in twenty-six river basins at the grid scale. We derive the uncertainties from the simulating period to the sixth year past the training phase individually, e.g., the gridded uncertainties of the third year are estimated by using only the third year of predictand.

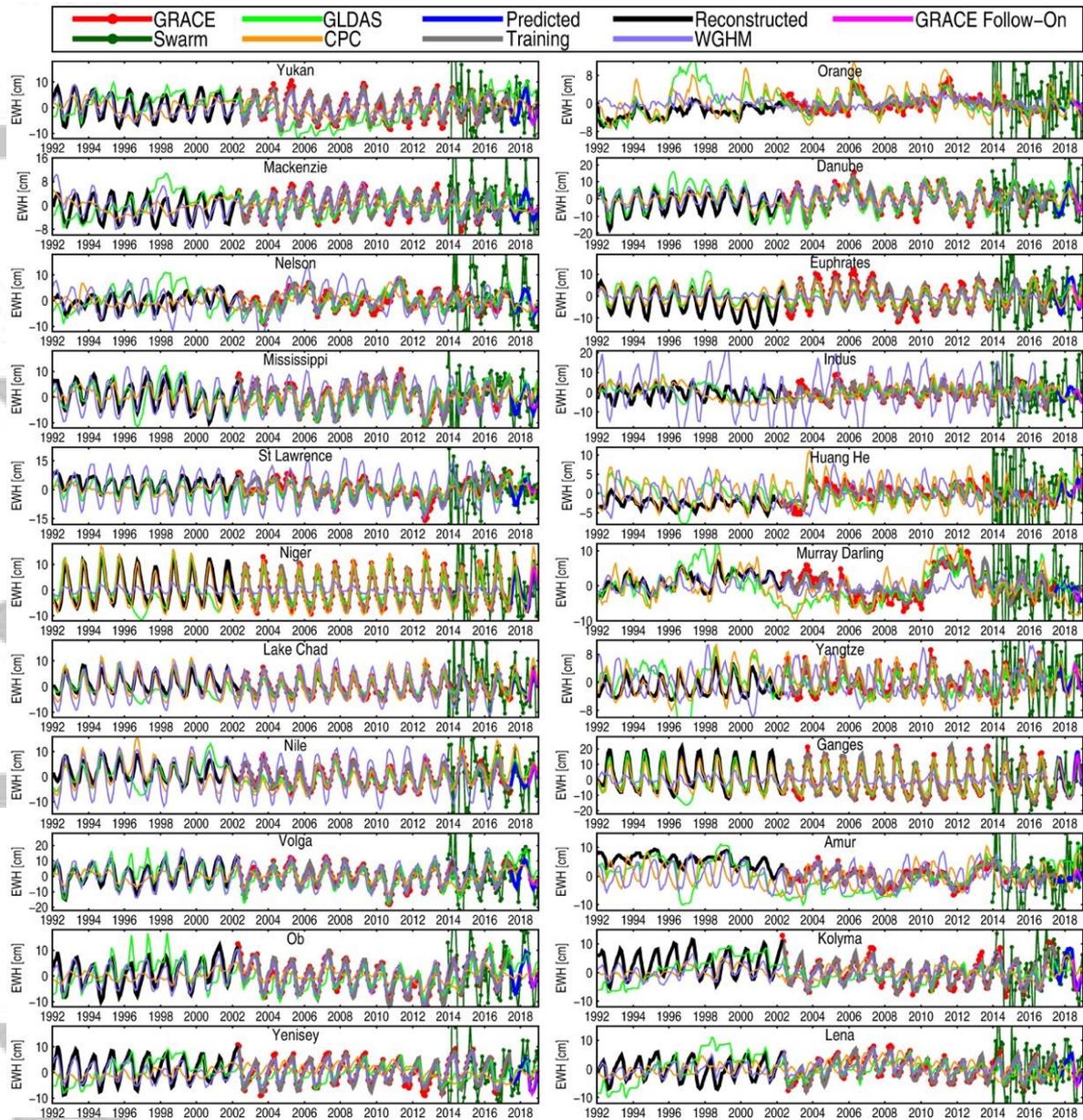


Figure A5. Reconstructed, training, and predicted TWSC relative to the TWSC from GRACE, GRACE-FO, Swarm, and hydrological data at the basin scale for twenty-two river basins.

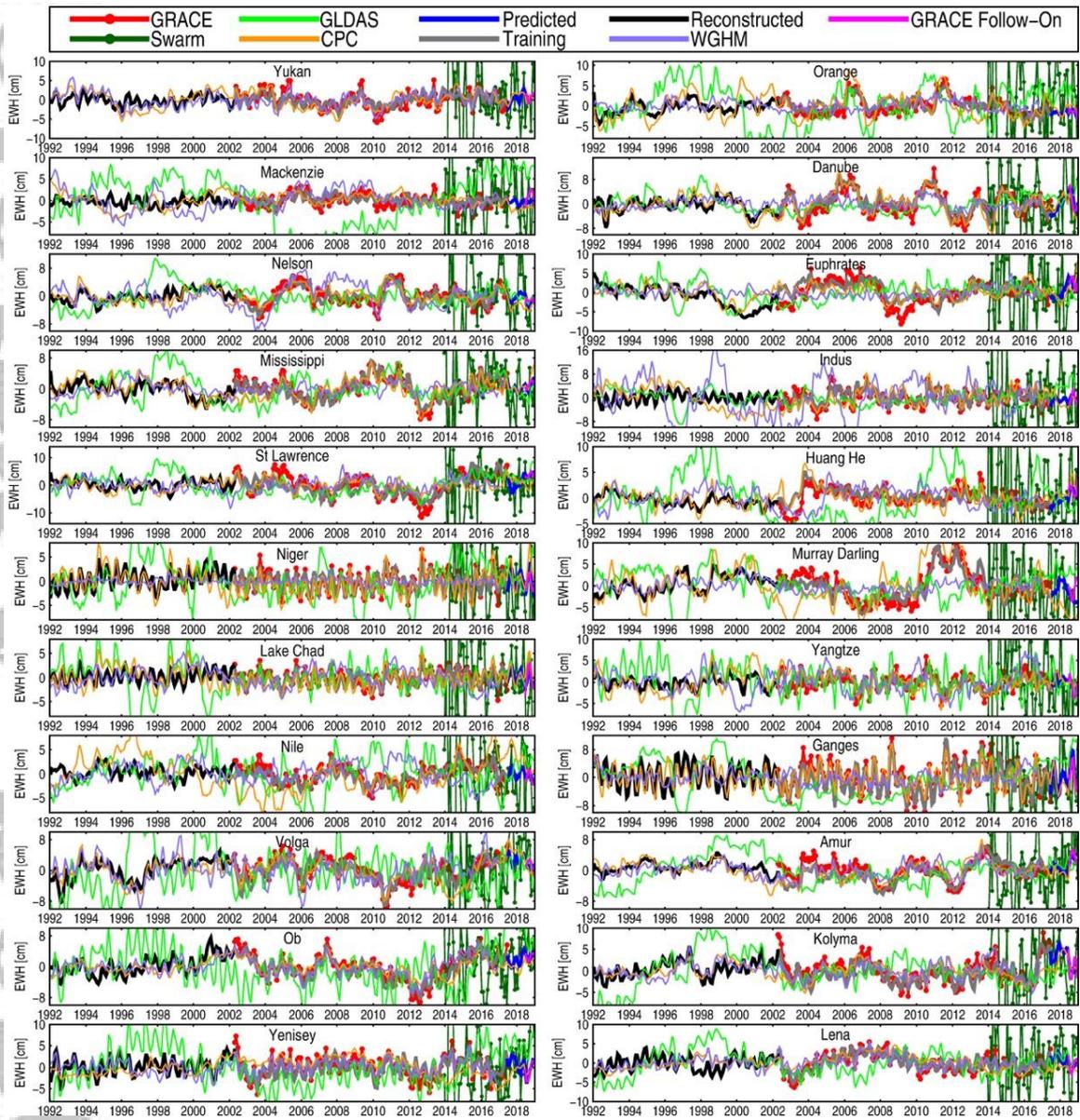


Figure A6. Same as Figure A5 about for the de-seasoned signal.

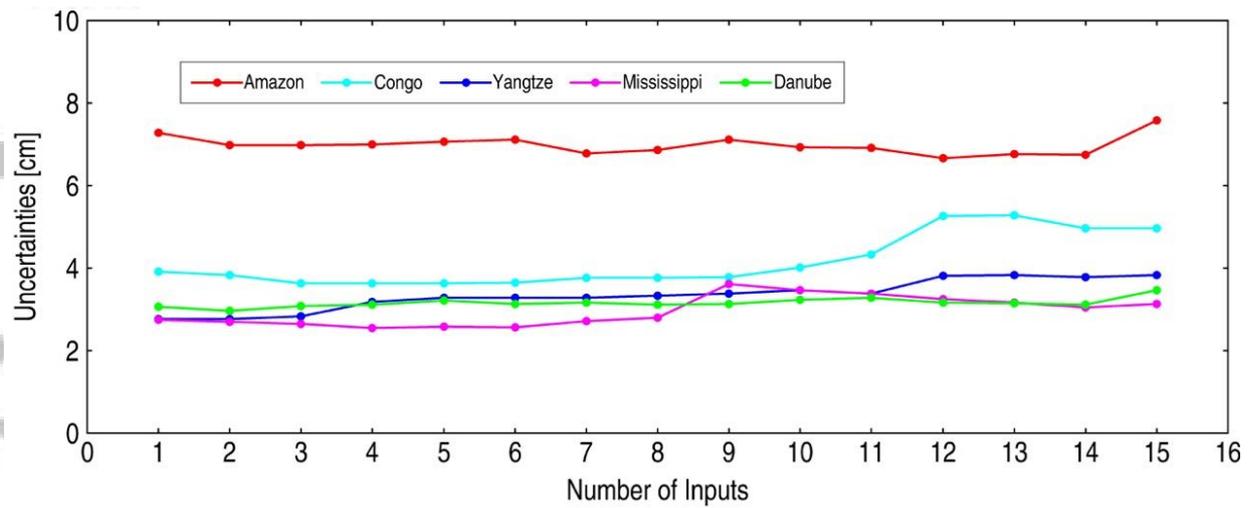


Figure C1. Standard errors of testing TWSC at the grid scale by turning the number of input climate drivers from 1 to 15 in five river basins as evaluated from the CSR Mascons.

Accepted Article

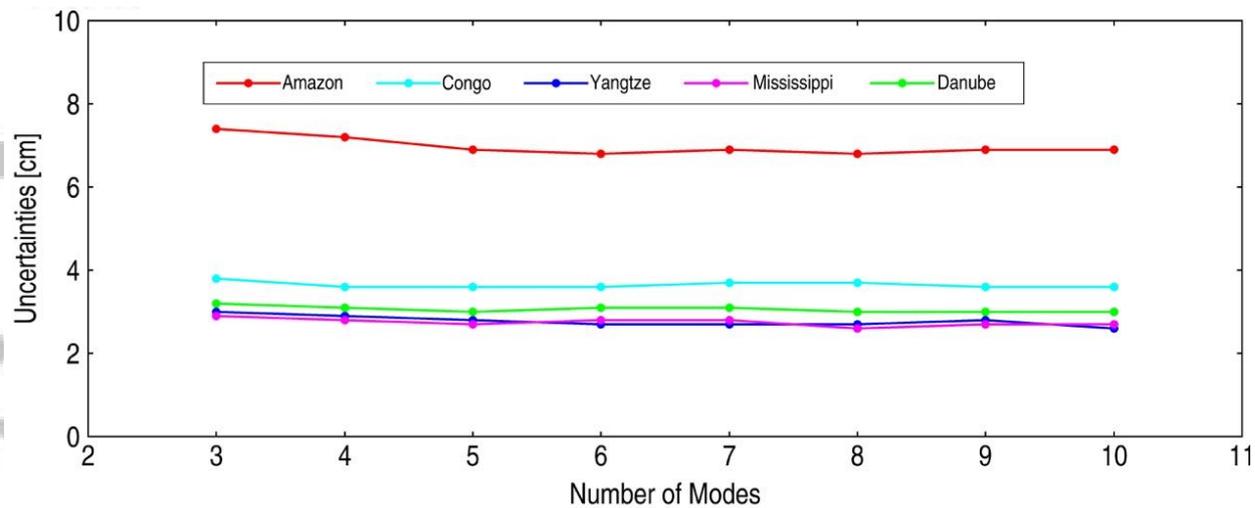


Figure C2. Standard errors of testing TWSC at the grid scales by turning the number of selected modes from 3 to 10 in five river basins as evaluated from the CSR Mascons.

Accepted Article

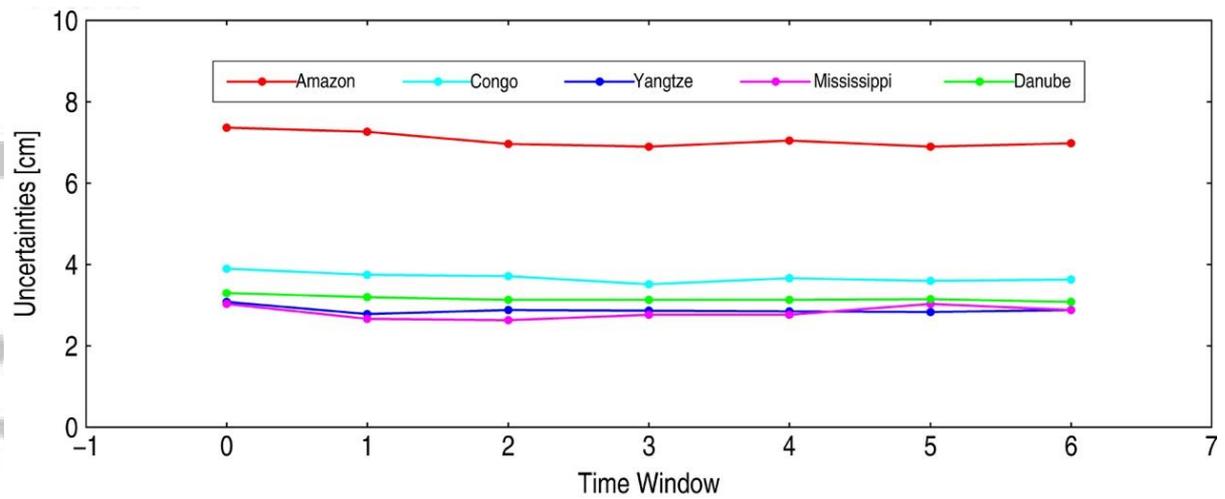


Figure C3. Standard errors of testing TWSC at the grid scales by turning the time window from (0...0) to (0...6) in five river basins as evaluated by the CSR Mascons.

Accepted Article

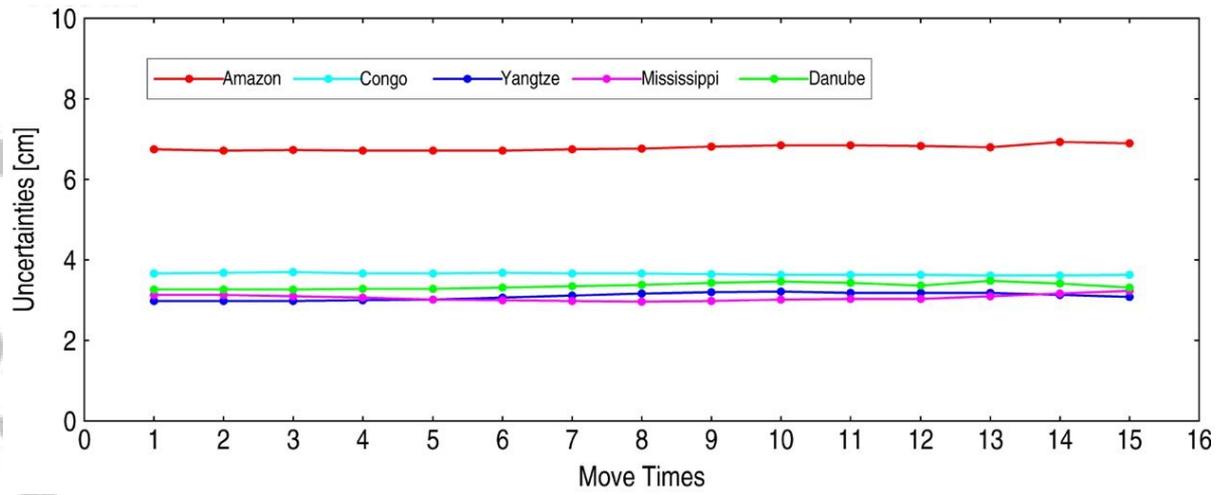


Figure C4. Standard errors of testing TWSC at the grid scales by moving the training and testing periods from 1 to 15 times in five river basins as evaluated by the CSR Mascons.

Accepted Article

Appendix B: Tables

Table B1

Standard Errors of Training and Testing TWSC at Both Grid and Basin Scales by Using STL and LS Methods in Twenty-six River Basins as Compared to CSR Mascons. Noting that All Results Listed in This Table are Calculated by Fixing The Spatio-temporal Decomposition and Predictive Techniques to PCA and MLR.

Basin	Training				Testing			
	STL		LS		STL		LS	
	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>
Yukon	1.9	1.1	2.0	1.2	2.3	1.4	2.2	1.3
Mackenzie	1.8	0.9	1.5	0.7	1.8	1.0	1.9	1.0
Nelson	2.5	1.7	2.4	1.8	2.5	1.7	2.5	1.7
Mississippi	2.3	1.0	2.5	1.1	3.4	2.3	2.8	1.7
St Lawrence	2.9	1.8	2.4	2.0	3.9	3.1	3.7	3.0
Amazon	5.5	2.3	5.4	2.0	7.2	4.1	7.1	3.9
Parana	3.5	1.6	3.7	1.5	5.5	2.4	4.9	2.0
Niger	1.6	1.0	1.4	0.9	1.5	1.1	1.8	1.0
Lake Chad	1.2	0.8	1.1	0.7	1.6	1.0	1.5	1.0
Nile	2.0	1.0	1.9	0.9	2.3	1.2	2.2	1.2
Congo	3.2	1.6	3.1	1.5	3.6	1.9	3.4	1.7
Zambezi	4.9	3.4	4.8	3.2	5.7	3.6	5.2	3.4
Orange	1.4	1.0	1.4	1.1	1.3	1.0	1.5	1.1
Danube	2.5	1.6	2.4	1.6	3.0	2.1	2.8	2.0
Euphrates	2.7	2.0	2.1	1.7	3.1	2.6	3.3	2.7
Volga	2.2	1.4	2.2	1.4	3.6	3.1	3.8	3.2
Ob	2.2	1.0	2.0	1.0	3.1	2.4	3.1	2.4
Yenisey	2.3	1.5	2.5	1.6	3.0	1.9	2.9	1.9
Lena	1.7	1.1	1.6	1.0	2.4	2.1	2.5	2.1
Kolyma	2.0	1.6	2.0	1.6	2.2	1.7	2.1	1.6
Amur	2.0	1.0	1.6	0.8	2.2	1.4	2.0	1.3
Huang He	1.3	0.9	1.3	0.8	1.8	1.2	1.8	1.1
Yangtze	2.2	0.9	2.2	0.9	2.8	1.5	2.6	1.3
Ganges	3.6	1.7	3.5	1.6	4.6	2.2	4.2	2.0
Indus	2.3	1.6	2.2	1.4	2.5	1.6	2.3	1.4
Murray Darling	2.2	1.8	2.1	1.8	3.3	2.6	2.7	2.4
Average	2.5	1.4	2.4	1.4	3.0	2.0	2.9	1.9

Table B2

Correlation Coefficients of Training and Testing TWSC at Both Grid and Basin Scales by Using STL and LS Methods in Twenty-six River Basins as Compared to CSR Mascons. Noting that All Results Listed in This Table are Calculated by Fixing The Spatio-temporal Decomposition and Predictive Techniques to PCA and MLR.

Basin	Training				Testing			
	STL		LS		STL		LS	
	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>
Yukon	0.92	0.97	0.92	0.97	0.89	0.95	0.91	0.96
Mackenzie	0.91	0.97	0.93	0.98	0.90	0.96	0.90	0.96
Nelson	0.86	0.89	0.84	0.88	0.78	0.85	0.81	0.85
Mississippi	0.90	0.97	0.90	0.97	0.79	0.88	0.86	0.93
St Lawrence	0.82	0.86	0.85	0.89	0.74	0.85	0.76	0.87
Amazon	0.93	0.98	0.94	0.99	0.92	0.96	0.93	0.97
Parana	0.88	0.97	0.88	0.96	0.74	0.90	0.77	0.93
Niger	0.85	0.96	0.88	0.99	0.84	0.97	0.82	0.99
Lake Chad	0.74	0.97	0.75	0.98	0.70	0.94	0.68	0.96
Nile	0.86	0.96	0.87	0.97	0.86	0.94	0.86	0.95
Congo	0.94	0.93	0.95	0.93	0.92	0.90	0.93	0.92
Zambezi	0.90	0.96	0.90	0.96	0.90	0.94	0.92	0.96
Orange	0.83	0.85	0.83	0.85	0.83	0.89	0.80	0.88
Danube	0.93	0.96	0.94	0.96	0.91	0.94	0.91	0.95
Euphrates	0.87	0.91	0.92	0.96	0.81	0.90	0.77	0.89
Volga	0.95	0.97	0.95	0.98	0.88	0.90	0.87	0.90
Ob	0.92	0.97	0.93	0.98	0.86	0.93	0.86	0.93
Yenisey	0.88	0.95	0.87	0.94	0.83	0.92	0.85	0.93
Lena	0.91	0.95	0.92	0.96	0.83	0.85	0.85	0.83
Kolyma	0.87	0.91	0.88	0.92	0.88	0.93	0.90	0.95
Amur	0.83	0.86	0.86	0.93	0.82	0.81	0.85	0.90
Huang He	0.81	0.91	0.82	0.93	0.63	0.83	0.62	0.79
Yangtze	0.87	0.94	0.88	0.96	0.82	0.88	0.81	0.90
Ganges	0.93	0.98	0.94	0.99	0.90	0.97	0.92	0.98
Indus	0.85	0.92	0.87	0.93	0.84	0.87	0.87	0.91
Murray Darling	0.85	0.88	0.86	0.88	0.70	0.84	0.66	0.81
Average	0.88	0.94	0.89	0.95	0.83	0.90	0.84	0.92

Table B3

Correlation Coefficients of Training and Testing TWSC after Removing the Seasonal Cycle at Both Grid and Basin Scales by Using STL and LS Methods in Twenty-six River Basins as Compared to CSR Mascons. Noting that All Results Listed in This Table are Calculated by Fixing The Spatio-temporal Decomposition and Predictive Techniques to PCA and MLR.

Basin	Training				Testing			
	STL		LS		STL		LS	
	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>
Yukon	0.68	0.85	0.65	0.85	0.53	0.56	0.60	0.65
Mackenzie	0.70	0.78	0.74	0.84	0.59	0.60	0.60	0.64
Nelson	0.78	0.83	0.78	0.81	0.61	0.68	0.65	0.67
Mississippi	0.80	0.91	0.80	0.90	0.64	0.65	0.75	0.82
St Lawrence	0.68	0.73	0.71	0.78	0.38	0.77	0.43	0.79
Amazon	0.68	0.84	0.71	0.84	0.65	0.76	0.66	0.72
Parana	0.76	0.92	0.75	0.89	0.61	0.86	0.65	0.90
Niger	0.73	0.86	0.75	0.90	0.69	0.85	0.64	0.87
Lake Chad	0.66	0.86	0.68	0.89	0.55	0.69	0.56	0.80
Nile	0.67	0.85	0.70	0.87	0.63	0.74	0.66	0.76
Congo	0.72	0.87	0.75	0.86	0.68	0.77	0.70	0.80
Zambezi	0.69	0.84	0.71	0.86	0.61	0.81	0.68	0.88
Orange	0.79	0.83	0.78	0.84	0.72	0.89	0.66	0.88
Danube	0.84	0.90	0.84	0.90	0.74	0.82	0.75	0.85
Euphrates	0.72	0.71	0.84	0.89	0.46	0.46	0.49	0.54
Volga	0.83	0.90	0.84	0.90	0.63	0.60	0.63	0.61
Ob	0.74	0.88	0.76	0.91	0.58	0.83	0.60	0.83
Yenisey	0.76	0.80	0.72	0.67	0.70	0.69	0.73	0.74
Lena	0.82	0.89	0.83	0.90	0.57	0.43	0.59	0.39
Kolyma	0.68	0.73	0.69	0.75	0.74	0.83	0.76	0.87
Amur	0.78	0.83	0.82	0.92	0.78	0.78	0.82	0.89
Huang He	0.74	0.88	0.77	0.91	0.50	0.74	0.47	0.70
Yangtze	0.70	0.81	0.71	0.87	0.55	0.72	0.51	0.77
Ganges	0.77	0.91	0.79	0.91	0.66	0.85	0.68	0.88
Indus	0.76	0.85	0.79	0.87	0.67	0.70	0.72	0.78
Murray Darling	0.84	0.86	0.84	0.87	0.55	0.80	0.49	0.74
Average	0.74	0.84	0.76	0.86	0.62	0.73	0.63	0.76

Table B4

Standard Errors of Training and Testing TWSC at Both Grid and Basin Scales by Using PCA and ICA Methods in Twenty-six River Basins as Compared to CSR Mascons. Noting that All Results Listed in This Table are Calculated by Fixing The Time-series Decomposition and Predictive Techniques to LS and MLR.

Basin	Training				Testing			
	ICA		PCA		ICA		PCA	
	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>
Yukon	2.0	1.2	2.0	1.2	3.1	1.8	2.2	1.3
Mackenzie	1.6	0.8	1.5	0.7	2.1	1.3	1.9	1.0
Nelson	2.6	1.8	2.4	1.8	2.3	1.6	2.5	1.7
Mississippi	2.7	1.2	2.5	1.1	3.5	2.2	2.8	1.7
St Lawrence	2.8	2.2	2.4	2.0	4.0	3.2	3.7	3.0
Amazon	6.1	2.5	5.4	2.0	7.4	4.2	7.1	3.9
Parana	4.0	1.8	3.7	1.5	5.9	3.2	4.9	2.0
Niger	1.8	1.1	1.4	0.9	2.0	1.1	1.8	1.0
Lake Chad	1.2	0.7	1.1	0.7	1.5	1.0	1.5	1.0
Nile	2.2	1.1	1.9	0.9	2.3	1.2	2.2	1.2
Congo	3.0	1.4	3.1	1.5	3.5	1.8	3.4	1.7
Zambezi	5.1	3.5	4.8	3.2	6.6	4.0	5.2	3.4
Orange	1.5	1.1	1.4	1.1	1.5	1.0	1.5	1.1
Danube	2.5	1.8	2.4	1.6	2.5	1.8	2.8	2.0
Euphrates	3.0	2.1	2.1	1.7	4.0	3.4	3.3	2.7
Volga	2.7	1.9	2.2	1.4	3.4	2.5	3.8	3.2
Ob	2.9	1.7	2.0	1.0	3.2	2.5	3.1	2.4
Yenisey	2.4	1.5	2.5	1.6	3.1	2.1	2.9	1.9
Lena	2.3	1.5	1.6	1.0	2.1	1.9	2.5	2.1
Kolyma	1.9	1.5	2.0	1.6	2.7	2.0	2.1	1.6
Amur	1.7	1.0	1.6	0.8	2.2	1.4	2.0	1.3
Huang He	1.6	1.0	1.3	0.8	1.8	1.1	1.8	1.1
Yangtze	2.5	1.1	2.2	0.9	2.8	1.4	2.6	1.3
Ganges	3.5	1.6	3.5	1.6	4.6	2.5	4.2	2.0
Indus	2.3	1.5	2.2	1.4	2.7	1.6	2.3	1.4
Murray Darling	1.9	1.6	2.1	1.8	2.0	1.6	2.7	2.4
Average	2.6	1.5	2.4	1.4	3.2	2.1	2.9	1.9

Table B5

Correlation Coefficients of Training and Testing TWSC at Both Grid and Basin Scales by Using PCA and ICA Methods in Twenty-six River Basins as Compared to CSR Mascons. Noting that All Results Listed in This Table are Calculated by Fixing The Time-series Decomposition and Predictive Techniques to LS and MLR.

Basin	Training				Testing			
	ICA		PCA		ICA		PCA	
	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>
Yukon	0.89	0.94	0.92	0.97	0.82	0.87	0.91	0.96
Mackenzie	0.91	0.96	0.93	0.98	0.87	0.94	0.90	0.96
Nelson	0.83	0.85	0.84	0.88	0.81	0.90	0.81	0.85
Mississippi	0.88	0.94	0.90	0.97	0.82	0.87	0.86	0.93
St Lawrence	0.86	0.89	0.85	0.89	0.83	0.83	0.76	0.87
Amazon	0.92	0.97	0.94	0.99	0.91	0.98	0.93	0.97
Parana	0.84	0.92	0.88	0.96	0.73	0.90	0.77	0.93
Niger	0.86	0.97	0.88	0.99	0.84	0.97	0.82	0.99
Lake Chad	0.72	0.97	0.75	0.98	0.68	0.95	0.68	0.96
Nile	0.86	0.94	0.87	0.97	0.84	0.92	0.86	0.95
Congo	0.95	0.92	0.95	0.93	0.93	0.88	0.93	0.92
Zambezi	0.89	0.95	0.90	0.96	0.83	0.90	0.92	0.96
Orange	0.80	0.83	0.83	0.85	0.73	0.86	0.80	0.88
Danube	0.93	0.95	0.94	0.96	0.92	0.96	0.91	0.95
Euphrates	0.86	0.89	0.92	0.96	0.80	0.78	0.77	0.89
Volga	0.92	0.95	0.95	0.98	0.89	0.92	0.87	0.90
Ob	0.86	0.90	0.93	0.98	0.83	0.92	0.86	0.93
Yenisey	0.86	0.95	0.87	0.94	0.83	0.87	0.85	0.93
Lena	0.90	0.93	0.92	0.96	0.86	0.89	0.85	0.83
Kolyma	0.89	0.91	0.88	0.92	0.86	0.91	0.90	0.95
Amur	0.82	0.90	0.86	0.93	0.77	0.84	0.85	0.90
Huang He	0.70	0.83	0.82	0.93	0.60	0.77	0.62	0.79
Yangtze	0.87	0.92	0.88	0.96	0.80	0.85	0.81	0.90
Ganges	0.93	0.98	0.94	0.99	0.90	0.97	0.92	0.98
Indus	0.85	0.92	0.87	0.93	0.83	0.87	0.87	0.91
Murray Darling	0.89	0.92	0.86	0.88	0.82	0.90	0.66	0.81
Average	0.86	0.92	0.89	0.95	0.82	0.89	0.84	0.92

Table B6

Correlation Coefficients of Training and Testing TWSC after Removing the Seasonal Cycle at Both Grid and Basin Scales by Using PCA and ICA Methods in Twenty-six River Basins as Compared to CSR Mascons. Noting that All Results Listed in This Table are Calculated by Fixing The Time-series Decomposition and Predictive Techniques to LS and MLR.

Basin	Training				Testing			
	ICA		PCA		ICA		PCA	
	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>	<i>grid</i>	<i>basin</i>
Yukon	0.57	0.63	0.65	0.85	0.36	0.26	0.60	0.65
Mackenzie	0.71	0.73	0.74	0.84	0.49	0.35	0.60	0.64
Nelson	0.77	0.75	0.78	0.81	0.69	0.78	0.65	0.67
Mississippi	0.75	0.79	0.80	0.90	0.67	0.60	0.75	0.82
St Lawrence	0.74	0.80	0.71	0.78	0.65	0.74	0.43	0.79
Amazon	0.63	0.66	0.71	0.84	0.68	0.86	0.66	0.72
Parana	0.73	0.81	0.75	0.89	0.63	0.85	0.65	0.90
Niger	0.68	0.77	0.75	0.90	0.63	0.73	0.64	0.87
Lake Chad	0.62	0.85	0.68	0.89	0.55	0.78	0.56	0.80
Nile	0.64	0.73	0.70	0.87	0.65	0.59	0.66	0.76
Congo	0.77	0.84	0.75	0.86	0.69	0.71	0.70	0.80
Zambezi	0.66	0.80	0.71	0.86	0.43	0.77	0.68	0.88
Orange	0.76	0.82	0.78	0.84	0.61	0.85	0.66	0.88
Danube	0.83	0.88	0.84	0.90	0.79	0.86	0.75	0.85
Euphrates	0.67	0.63	0.84	0.89	0.42	0.22	0.49	0.54
Volga	0.78	0.81	0.84	0.90	0.70	0.76	0.63	0.61
Ob	0.59	0.69	0.76	0.91	0.51	0.82	0.60	0.83
Yenisey	0.74	0.77	0.72	0.67	0.69	0.50	0.73	0.74
Lena	0.80	0.83	0.83	0.90	0.67	0.56	0.59	0.39
Kolyma	0.76	0.76	0.69	0.75	0.69	0.77	0.76	0.87
Amur	0.78	0.88	0.82	0.92	0.72	0.81	0.82	0.89
Huang He	0.61	0.78	0.77	0.91	0.46	0.65	0.47	0.70
Yangtze	0.68	0.75	0.71	0.87	0.52	0.59	0.51	0.77
Ganges	0.76	0.90	0.79	0.91	0.65	0.82	0.68	0.88
Indus	0.76	0.85	0.79	0.87	0.66	0.67	0.72	0.78
Murray Darling	0.88	0.91	0.84	0.87	0.72	0.88	0.49	0.74
Average	0.71	0.78	0.76	0.86	0.61	0.68	0.63	0.76

Table B7
The Acronyms that Involved in this Paper

GRACE	Gravity Recovery and Climate Experiment
NASA	National Aeronautical and Spatial Administration
DLR	German Aerospace Centre
TWSC	Total Water Storage Change
ANN	Artificial Neural Network
ICA	Independent Component Analysis
ARX	AutoRegressive eXogenous
STL	Seasonal-Trend decomposition based on Loess
MLR	Multiple Linear Regression
CNN	Convolutional Neural Network
PCA	Principal Component Analysis
LS	Least Squares
SST	Sea Surface Temperature
EOF	orthogonal spatial patterns
ICA	Independent Component Analysis
MLP	Multi-Layer Perceptron
MCMC	Markov Chain Monte Carlo
CSR	Center for Space Research
CPC	Climate Prediction Center
GHCN CAMS	Global Historical Climatology Network and the Climate Anomaly Monitoring System
NCEP	National Centers for Environmental Prediction
NOAA	National Oceanic and Atmospheric Administration
MEI	Multivariate ENSO Index
NAO	North Atlantic Oscillation
Niño 1+2	Extreme Eastern Tropical Pacific SST
Niño 3	Eastern Tropical Pacific SST
Niño 4	Central Tropical Pacific SST
Niño 3.4	East Central Tropical Pacific SST
NTA	North Tropical Atlantic SST Index
ONI	Oceanic Niño Index
PDO	Pacific Decadal Oscillation
PNA	Pacific North American Index
QBO	Quasi-Biennial Oscillation
SOI	Southern Oscillation Index
TNA	Tropical Northern Atlantic Index
TNI	Trans-Niño Index
TSA	Tropical Southern Atlantic Index
WHWP	Western Hemisphere Warm Pool
WP	Western Pacific Index
GLDAS	Global Land Data Assimilation System
WGHM	WaterGAP Global Hydrology Model
GAM	Generalized Additive Model
ENSO	El Nino-Southern Oscillation