



Global analysis of spatiotemporal variability in merged total water storage changes using multiple GRACE products and global hydrological models



Di Long ^{a,*}, Yun Pan ^b, Jian Zhou ^c, Yang Chen ^d, Xueyan Hou ^a, Yang Hong ^a,
Bridget R. Scanlon ^e, Laurent Longuevergne ^f

^a State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China

^b College of Resources Environment and Tourism, Capital Normal University, Beijing 100048, China

^c Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China

^d Tianjin Key Laboratory of Water Resources and Environment, Tianjin Normal University, Tianjin 300387, China

^e Bureau of Economic Geology, Jackson School of Geosciences, The University of Texas at Austin, TX 78758, United States

^f Geosciences Rennes, UMR CNRS 6118, Université de Rennes 1, Rennes, France

ARTICLE INFO

Article history:

Received 1 August 2016

Received in revised form 20 January 2017

Accepted 20 February 2017

Available online 27 February 2017

Keywords:

GRACE

Total water storage change

Three cornered hat method

Bayesian model averaging

Empirical orthogonal function

Long-term trends

Seasonal variability

Global hydrologic models

ABSTRACT

Proliferation of different total water storage (TWS) change products from the Gravity Recovery and Climate Experiment (GRACE) satellites, including the newly released mascon solution, warrants detailed analysis of their uncertainties and an urgent need to optimize different products for obtaining an elevated understanding of TWS changes globally. The three cornered hat method is used to quantify uncertainties in TWS changes from GRACE observations, land surface models, and global hydrological models, indicating that the WaterGap Global Hydrological Model (WGHM)-based TWS changes show the lowest uncertainty over sixty basins covering a range of climate settings and levels of human activities globally. Bayesian model averaging (BMA) using WGHM TWS output for training (2003–2006) is subsequently used to merge TWS changes from various GRACE products. Results indicate that the BMA-based TWS changes show the highest consistency with the WGHM output for the validation period (2007–2009) in terms of the highest medium of the Nash-Sutcliffe Efficiency (NSE) coefficient of 0.714 among all TWS change products for the sixty basins. The mascon solution shows a medium of NSE of 0.682, higher than other GRACE TWS change products. Analysis of spatiotemporal variability in BMA-based TWS changes and the mascon solution indicates that higher depletion rates for the 13-year period (Apr 2002–Mar 2015) occurred over major aquifers due to groundwater withdrawals for irrigation (e.g., Tigris, Central Valley, Ganges, upper Arkansas, and Indus), basins subject to great glacier and snow melting (e.g., Yukon, Fraser, and eastern Ganges), the north Caspian Sea (e.g., Don and Ural), and the Caspian Sea. Significant increasing trends in TWS are found over west (e.g., Gambia and Niger) and South Africa (e.g., Zambezi), South America (e.g., Essequibo), North America (e.g., Koksoak and Missouri), central India (e.g., Narmada and Godavari), the north Tibetan Plateau, and the middle Yangtze River basin. Empirical Orthogonal Function decomposition is used to investigate spatiotemporal variations in the GRACE mascon solution-based TWS changes during the study period, showing a detailed pattern of increasing and decreasing long-term trends, interannual and seasonal variations in TWS over the global land surface.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Total water storage (TWS) change integrates surface water storage (SWS, including canopy interception, reservoirs, wetlands and lakes, rivers and snow water equivalent), soil moisture storage (SMS), and groundwater storage (GWS) changes, and is a critical state variable of the hydrological cycle that reflects the net effect of all hydrological

flux variables, e.g., precipitation, evapotranspiration, runoff, and drainage. Climate change and variability and extensive anthropogenic activities (e.g., irrigation) have profoundly influenced TWS and resulted in changes in the hydrological cycle, threatening sustainable water use and agricultural production over major aquifers globally (Long et al., 2016; Richey et al., 2015a; Richey et al., 2015b). Climate extremes (e.g., drought) have also resulted in significant water storage depletion that further deteriorates surface and subsurface water resource management, e.g., the Millennium Drought in the Murray-Darling basin in Australia (Leblanc et al., 2009), the 2011 extreme drought in Texas (Long et al.,

* Corresponding author.

E-mail address: dlong@tsinghua.edu.cn (D. Long).

2013; Thomas et al., 2014), and a drought starting from 2007 in the Middle East (Voss et al., 2013). Alternatively, increases in TWS over land may act as a buffer that slows down global sea level rise (Reager et al., 2016), which may result in increased flooding potential for land (Long et al., 2015b; Reager et al., 2014). In addition, the cryosphere is highly sensitive to climate change and extremes. Glacier and snow melting over high-mountain areas will likely threaten water use in the lower reaches of rivers fed jointly by precipitation and glacier and snow meltwater (Farinotti et al., 2015; Immerzeel et al., 2010). A comprehensive and detailed investigation into TWS changes in space and time should provide a thorough understanding of how climate change and variability and anthropogenic activities increase or decrease water storage and the related consequences. This investigation can subsequently be valuable for scientists and decision-makers in formulating scientifically based approaches and policies for water resource management.

Traditionally, there was no ground-based measurement for TWS changes, limiting our understanding of how TWS of the land changed in space and time. The launch of the Gravity Recovery and Climate Experiment (GRACE) satellites in 2002 has provided an unprecedented opportunity to detect TWS changes at a spatial resolution of ~200,000 km² by making accurate measurements of changes in the distance between two identical satellites that generally reflect changes in the Earth's gravity field (Tapley et al., 2004). GRACE data require low-pass filtering (i.e., truncation of high-degree and high-order spherical harmonic (SH) coefficients of the Earth's gravity field, destriping, and smoothing for the SH coefficients) to reduce high-frequency noise. There are three main research centers using various processing strategies to derive SH solutions, including the Center for Space Research (CSR) at The University of Texas at Austin, Jet Propulsion Laboratory (JPL) at the California Institute of Technology, and the German Research Center for Geoscience (GFZ). Low-pass filtering, however, results in partial signal loss (amplification) that can be restored (dampened) using different signal restoration approaches. These approaches can generally be subdivided into two basic types (Long et al., 2016): (1) an approach that depends on TWS output from land surface models (LSMs) as a priori information to compute bias and leakage errors, e.g., the additive correction approach (Klees et al., 2007; Longuevergne et al., 2010), or to generate scaling factors to correct for bias and leakage errors from GRACE data (Landerer and Swenson, 2012) and (2) an approach that depends less on LSMs, such as the multiplicative correction approach (Swenson and Wahr, 2007; Velicogna and Wahr, 2006a) and forward modeling approach (Chen et al., 2014; Horwath and Dietrich, 2009; Long et al., 2016). In addition, recently released JPL mass concentration blocks (mascons)-based TWS changes (Watkins et al., 2015) do not require scaling and might show low uncertainties over regions where LSMs do not work well in generating scaling factors (Long et al., 2015a).

Overall, each approach for GRACE signal restoration has advantages and limitations under a certain circumstance (Long et al., 2016). LSM-based approaches may suffer from uncertainties across relatively dry areas and areas with intensive irrigation due to larger uncertainties in LSMs over these regions (Long et al., 2015a). Approaches that do not depend largely on LSMs, such as the multiplicative correction approach, may result in significant deviations once its assumption of uniform distribution of TWS changes across a study basin is not met. The mascons do not rely largely on LSMs for providing a priori information for GRACE signal restoration. Furthermore, the mascons can be applied at regional to global scales, as opposed to the global SH solutions that cannot distinguish land and ocean areas and would therefore result in a marked leakage effect along coastal areas (Landerer and Swenson, 2012; Long et al., 2015b). The gridded nature of the mascons allows non-geodesists to average over arbitrary regions of interest and compares it with other gridded output from LSMs or hydrological models as GRACE Level 3 gridded SH products do (Landerer and Swenson, 2012). However, the reliability of this product and how it compares with other traditional products, e.g., SH products using different signal restoration approaches, have not been systematically investigated.

An optimized algorithm is urgently needed to generate TWS changes that contain advantages but reduce limitations in different GRACE TWS change products. There has been increasing interest in use of Bayesian model averaging (BMA) to merge different weather and hydrological predictions by pooling information from multiple models based on a priori information (Duan and Phillips, 2010; Hoeting et al., 1999; Raftery et al., 2005) that can be obtained from in situ measurements. It is, however, impossible to obtain TWS changes from ground-based measurements as the reference to merge different GRACE TWS change estimates. Model TWS output may be an alternative of a priori information for BMA once its uncertainty is shown to be lower than GRACE-derived TWS changes. For example, the WaterGap Global Hydrological Model (WGHM), as one of well-established global hydrological models (GHMs), comprehensively simulates SWS, SMS and GWS changes caused by both natural groundwater recharge and discharge and human-induced groundwater abstractions and return flow (Alcamo et al., 2003; Doll et al., 2003; Müller Schmied et al., 2014). There are also many LSMs, e.g., Noah2.7, Mosaic, VIC, and CLM2.0 in GLDAS-1 (Rodell et al., 2004) that can provide TWS output. However, TWS changes from LSMs may be subject to large uncertainties due to limitations/ errors in forcing, representation of physical processes, and parameter calibration (Long et al., 2013). In addition, GHM-based TWS changes may be subject to more delayed release than GRACE monitoring due to large efforts required in data collection and preparation. The latency in information on GHM TWS estimates may not meet the need for water resource management. We consider that historic simulations of TWS changes using GHMs (e.g., WGHM) may incorporate good a priori knowledge on the impact of water use on surface and subsurface water storage changes and therefore provide more reliable estimates of TWS changes than LSMs that do not often have a groundwater storage component. We propose that BMA and TWS simulations from some GHM as the reference can be jointly used to merge multiple TWS change products from GRACE observations, if the GHM-based TWS is shown to have an overall lower uncertainty than GRACE and other GHMs and LSMs.

A number of regional studies have been performed to understand how GRACE-derived TWS changes over time, and in particular, long-term trends in TWS due to climate variability and/or anthropogenic activities. Significant TWS depletion over the Northwest India Aquifer (Chen et al., 2014; Long et al., 2016; Rodell et al., 2009), the North China Plain (Feng et al., 2013; Huang et al., 2015), the Middle East (Forootan et al., 2014; Joodaki et al., 2014; Voss et al., 2013), US California's Central Valley (Famiglietti et al., 2011; Scanlon et al., 2012) and the High Plains Aquifer (Breña-Naranjo et al., 2014; Strassberg et al., 2009) was attributed mostly to groundwater withdrawals for irrigation, particularly during droughts. Impacts of the 2005 exceptional drought on TWS in the Amazon basin and its ecological consequences have also attracted a lot of attentions (Chen et al., 2009; Frappart et al., 2012; Morton et al., 2014). On the cryosphere, significant TWS depletion has been found over Greenland and the Antarctic (Velicogna and Wahr, 2006a, 2006b), Alaska (Arendt et al., 2013; Chen et al., 2006), Patagonia (Chen et al., 2007), and the southeast Tibetan Plateau (Jacob et al., 2012) because of ice, glacier and snow melting. Increasing trends in TWS have also been well documented over west and South Africa (Ahmed et al., 2014; Ramillien et al., 2014). In addition to the long-term trends the above-mentioned regional studies have investigated, interannual and seasonal variability in TWS is also important for a better understanding of related hydrological processes over regions of interest. Regions without showing significant long-term trends and interannual variability are also of interest to us since these regions may be less susceptible to climate change and variability. A global assessment of spatiotemporal variability in TWS is therefore needed to provide a comprehensive picture as to how TWS changes over time under different climate settings and levels of anthropogenic activities. Investigating such changes and characteristics at basin scales will be of interest to hydrologists to understand hydrological processes

for a specific basin, and similarities and differences compared with other basins globally.

Döll et al. (2014) evaluated groundwater depletion (GWD) from an improved WGHM model considering groundwater recharge from surface water bodies over major aquifers globally and compared the GWD with GRACE and groundwater level observations. A majority of aquifers indicate the consistency between model output and GRACE observations. This indicates that joint use of GRACE and WGHM may improve our understanding of water storage changes. Reager et al. (2016) evaluated groundwater depletion globally using the JPL mascons and quantified how the depletion impacts global sea-level rise, indicating that the rate of the sea-level rise has been reduced by ~15% due to increasing groundwater storage induced mostly by increased precipitation. Humphrey et al. (2016) performed a comprehensive global analysis of spatiotemporal variability in TWS using gridded SH data without scaling. Interannual, seasonal, and sub-seasonal variability in TWS was derived by the Seasonal Trend Decomposition using Loess procedure (STL). The correlation of TWS variability with precipitation and temperature fields was also investigated in detail. The filtered gridded SH data used by Humphrey et al. (2016) may result in damped amplitudes of long-term trends and temporal variability. A more accurate and deeper understanding of TWS changes can be achieved by using optimized TWS changes that incorporate advantages but reduces limitations of different available GRACE TWS change products. Furthermore, the newly released mascons ease the issue of scaling involved in most of current TWS change products and may therefore be more valuable in unraveling spatiotemporal variability in TWS over the global land surface.

The objectives of this study are therefore to (1) evaluate uncertainties in TWS changes from GRACE observations, LSMs, and GHMs using a statistical approach (three cornered hat) without the need for the true values of TWS changes, (2) merge TWS changes from three official gridded SH solutions with leakage errors corrected by scaling factors derived from CLM4.0 (i.e., CSR, JPL, and GFZ), two widely used approaches for leakage error correction (additive and multiplicative), and the JPL mascons using BMA and TWS output from a model that shows an overall lower uncertainty than GRACE observations and other model outputs, (3) compare the BMA-based TWS changes with the mascons and other GRACE TWS change products and evaluate with the model output, and (4) examine spatiotemporal variations in TWS for sixty basins and the global land surface for a 13-year period (Apr 2002–Mar 2015). Results of this study for the first time provide a detailed picture of spatiotemporal variability in GRACE-derived TWS changes, including increases and decreases in storage over land using a merged product at the basin scale and the recent JPL mascons at the grid cell scale. The proposed approach will be valuable in generating more reliable TWS changes for studies on global and regional hydrology and water resource management for basins with varying climate settings and levels of human activities.

2. Methods and materials

2.1. Methods

2.1.1. Three cornered hat approach for uncertainty analysis of TWS changes

Traditional methods of evaluating TWS changes are based on hydrological flux variables involved in the water balance equation (e.g., Castle et al., 2014; Long et al., 2014a), i.e., $\Delta S = P - ET - R$ where ΔS is TWS change; P is precipitation, ET is evapotranspiration, and R is streamflow under the assumption of no ground flow fluxes between the study basin of interest and its surrounding basins. However, TWS changes inferred from the water balance-based approach are subject to uncertainties in these flux variables, particularly uncertainties in evapotranspiration (Long et al., 2014a) and precipitation (Wang et al., 2015). Alternatively, streamflow records are hard to obtain in most parts of the world,

particularly over underdeveloped regions and transboundary river basins with sensitive water issues.

Because of the extreme difficulty in obtaining realistic TWS changes to evaluate the reliability of different TWS products or model outputs, here we propose to use the three cornered hat (TCH) method to quantify uncertainties in TWS changes from three categories without the need for a priori information on the realistic TWS changes: (1) GRACE satellites (including three SH solutions from CSR, JPL, and GFZ), (2) LSMs in GLDAS-1 (i.e., Noah2.7, Mosaic, VIC, and CLM2.0), and (3) GHMs (including WGHM and PCR-GLOBWB) of the sixty basins (see Section 2.3) globally. It was apparently impossible to exhaust all GRACE solutions, LSMs and GHMs available in this study. We considered that these selected products and outputs are widely used in the hydrological community and the three SH solutions and four LSMs are openly accessible. Therefore, use of them seems sufficient to demonstrate the relative uncertainties in TWS changes for each basic category. The TWS change product/output of the lowest uncertainty for the sixty basins will subsequently be taken as the reference for BMA to merge different TWS changes from GRACE satellites. GRACE-observed TWS changes enjoy advantages of timeliness and fewer efforts in data preparation and processing than LSMs/GHMs. LSMs/GHMs have detailed descriptions of hydrological processes, and, specifically, GHMs consider the impacts of human activities (e.g., water withdrawals for agricultural, industrial, and domestic water use) on surface and subsurface water storage changes, but the timeliness of GHMs is often lower than GRACE satellites.

The TCH approach (Premoli and Tavella, 1993), similar to the triple collocation approach (Stoffelen, 1998), can be used to estimate relative uncertainties in different TWS change products without a priori knowledge of actual TWS changes when at least three different sets of data (e.g., GRACE, LSMs, and GHMs in this study) are available. The theory of TCH is described in the following and based on the assumption that observational errors are normally distributed. A given set of observations, obs_i consists of two components: the true value, x , and an associated measurement error, e_i :

$$obs_i = x + e_i \quad (1)$$

Given a set of three pairs of observations (i, j, k) , the difference between observations (i, j) can be written as:

$$obs_i - obs_j = x + e_i - (x + e_j) = e_i - e_j \quad (2)$$

The associated variance of the differences can be written as:

$$\sigma_{ij}^2 = \sigma_{ei}^2 + \sigma_{ej}^2 - 2\text{cov}(e_i, e_j) \quad (3)$$

If errors between estimates of i and j are independent, the $\text{cov}(e_i, e_j)$ equals zero. Finally, the individual variances σ_{ei}^2 may be separated by:

$$\sigma_{ei}^2 = \frac{1}{2} \left(\sigma_{ij}^2 + \sigma_{ik}^2 - \sigma_{jk}^2 \right) \quad (4)$$

Note that the three observations in the TCH approach could be correlated (e.g., LSMs and GHMs may have similar forcing), it is important to consider that cross correlation among the three observations (i.e., $\text{cov}(e_i, e_j)$) is not necessarily zero. When considering three signals, the standard deviation of the signal differences provides three equations. If cross correlation (uncorrelated errors) is not considered, one has only three unknowns that could be solved based on Eq. (4). If cross correlation is considered, there will be six unknowns to determine. Tavella and Premoli (1991) proposed to add mathematical characteristics of the covariance matrices, i.e., its positive definiteness that the determinant is possible, or, in other words, the determined variances should be all positive so as to solve the problem. This generalization was further developed by Premoli and Tavella (1993) and Tavella and Premoli

(1994). In this study, we used the TCH approach that considers cross correlation among observations and does not require that the data sources be entirely independent (Premoli and Tavella, 1993). The TCH approach considering cross correlation has also been used recently to quantify uncertainties in ET estimates from GRACE satellites, LSMs, and remotely sensed vegetation indices (Long et al., 2014a). It has been shown to be an effective way to quantify uncertainties in data sets/model outputs without *a priori* knowledge of the variables.

2.1.2. Bayesian Model Averaging for TWS change merging

The BMA approach was used to merge six GRACE TWS products (i.e., additive correction, multiplicative correction, three gridded SH (CSR, JPL, and GFZ), and one mascon solution (JPL)) at the basin scale. Note that the model of ‘BMA’ here means different algorithms to derive TWS changes from GRACE observations. The predictive probability density function (PDF) for TWS changes is a weighted average of the PDFs for individual TWS products, weighted by their posterior model probability (Duan and Phillips, 2010; Raftery et al., 2005). BMA may improve the accuracy of TWS changes by adjusting the predictive PDF to obtain a good fit to a product/output that shows the lowest uncertainty based on the TCH approach, which will be illustrated in Section 3.1.

BMA considers a dependent variable y , the corresponding observation data at a given time t for training y_t , and an ensemble model of all members' predictions $X\{x_1, x_2, \dots, x_f\}$ for variable y . In this study, y is the ensemble TWS change and f is the number of TWS products, i.e., 6 for the first try and other combinations of TWS products (<6) will also be considered. The training data set y_t was obtained from the product/output that shows the lowest uncertainty based on the TCH approach. Based on the law of total probability, the PDF of dependent variable y given a range of TWS products (x_1, x_2, \dots, x_f) , $p(y|x_1, x_2, \dots, x_f)$ can be expressed as (Bulmer, 1979):

$$p(y|x_1, x_2, \dots, x_f) = \sum_{i=1}^f p(x_i|y_t) \cdot p(y|x_i) \quad (5)$$

where $p(y|x_i)$ is the predictive PDF given a specific TWS product x_i and $p(x_i|y_t)$ is the posterior probability that product x_i is correct given the corresponding observation data. Probability $p(x_i|y_t)$ can be considered as a statistical weight u_i , which reflects how well x_i matches the observation data y_t , and $\sum u_i = 1$. Eq. (1) can therefore be expressed as:

$$p(y|x_1, x_2, \dots, x_f) = \sum_{i=1}^f u_i \cdot p(y|x_i) \quad (6)$$

It is assumed that $p(y|x_i)$ follows a Gaussian distribution determined by a mean \tilde{x}_i and a variance ω_i^2 (Duan and Phillips, 2010; Raftery et al., 2005). This results in the following:

$$p(y|x_i) = g(y|\theta_i) \quad (7)$$

where g denotes the Gaussian distribution and $\theta_i = \{\tilde{x}_i, \omega_i^2\}$ represents the parameter vector. Combining Eqs. (6) and (7) results in:

$$p(y|x_1, x_2, \dots, x_f) = \sum_{i=1}^f u_i \cdot g(y|\theta_i) \quad (8)$$

The conditional expectation (E) of y is the ultimate BMA-predicted TWS change by merging different TWS change products, which can be expressed as follows:

$$E(y|x_1, x_2, \dots, x_f) = \sum_{i=1}^f u_i \cdot \tilde{x}_i \quad (9)$$

where \tilde{x}_i is the given TWS change using each individual algorithm. To obtain both u_i and θ_i , a log likelihood function l from the Gaussian

function g based on training data can be used. The log likelihood function can be expressed as:

$$l(\theta_1, \theta_2, \dots, \theta_f) = \sum_{(s,t)} \log \left[\sum_{i=1}^f u_i \cdot g(y_{s,t}|\theta_i) \right] \quad (10)$$

where $\sum_{(s,t)}$ denotes the summation of observed TWS changes for a specific basin s at time t ; $y_{s,t}$ is an observed TWS change value at basin s and time t . The BMA approach will estimate weight u_i and parameter vector θ_i by maximizing the log likelihood function l using the expectation-maximization (EM) algorithm (Duan and Phillips, 2010). TWS output with the lowest uncertainty using the TCH approach (i.e., the reference data, termed TWS_R in this study) was used to derive weight u_i for the period 2003–2006 (i.e., training), and the TWS_R for the period 2007–2009 (i.e., validation) was used for validating the merged TWS changes from BMA. Uncertainties in TWS changes from the BMA approach for each month consider one standard deviation of TWS changes from different products used to merge TWS changes. We also used different TWS_R for the training and validation periods to see if the BMA approach is dependent on the reference data.

2.1.3. Statistical approaches for analysis of TWS changes

Long-term trends of TWS time series from the BMA approach and mascons were derived by linear regression after the time series were filtered using the Butterworth filter (order 4 and normalized cutoff frequency 0.05) to remove seasonal variations and high-frequency noise in these time series (Scanlon et al., 2012). Standard errors of slopes of the linear regression for the filtered TWS change time series are also given to indicate uncertainties in slopes. The significance level was taken as 0.05 in this study.

Empirical orthogonal function (EOF) analysis was used to examine the spatiotemporal variability in TWS during the study period from Apr 2002 through Mar 2015. EOF is a statistical method that can diagonalize the covariance matrix of a field into a set of eigenvalues and corresponding eigenvectors, and then project the eigenvectors onto the original field to obtain a time series (Venegas et al., 1997). In practice, EOF analysis is performed to decompose time series of variable fields into two components, i.e., the spatial function component without being changed over time and the temporal function component depending only on temporal variations (e.g., interannual, annual, and seasonal variations). The spatial function component characterizes the spatial distribution of variable fields, and the temporal function component consists of linear combinations of variables in the spatial field, i.e., the principle component (PC). The first several PCs account for the major portion of the total variances of the original field, equivalent to concentrating the primary information of the original fields on these PCs. Therefore, examining temporal variations in PCs is essentially examining the temporal variations of the original field. Before conducting the EOF analysis, the TWS change data from mascons were pre-processed by subtracting the temporal mean for each grid cell to derive the anomaly field.

2.2. Data and statistic metrics

TWS changes from four LSMs in GLDAS-1 for the period 2003–2009 at a spatial resolution of $0.25^\circ \times 0.25^\circ$ (Noah) or $1^\circ \times 1^\circ$ (Mosaic, VIC, and CLM) include part of surface water storage components (i.e., snow water equivalent and canopy interception) and soil moisture storage changes (2 m for Noah 2.7, 3.5 m for Mosaic, 1.9 m for VIC, and 3.4 m for CLM2.0) (Rodell et al., 2004). TWS changes from two GHMs, i.e., PCR-GLOBWB (van Beek et al., 2012; Wada et al., 2011; Wada et al., 2014) and WaterGAP WGHM2.2 (Doll et al., 2003; Müller Schmied et al., 2014) for the period 2003–2009 at a spatial resolution of $0.5^\circ \times 0.5^\circ$ were obtained for use in evaluating their relative

uncertainties using TCH. The gridded GHM/LSM TWS change output was spatially averaged for each study basin as reference data for BMA that merges different GRACE TWS change products at the basin scale. Note that one of the most significant differences between the LSMs and GHMs is that LSMs do not have a groundwater storage component but GHMs have. Such a difference can be manifested over major aquifers globally where groundwater storage has been rapidly depleting due to intensive groundwater withdrawals for irrigation.

GRACE SH coefficients (CSR RL05) (<ftp://podaac.jpl.nasa.gov/allData/grace/L2/CSR/RL05/>) and monthly soil moisture and snow water equivalent from Noah2.7 in GLDAS-1 at a spatial resolution of $0.25^\circ \times 0.25^\circ$ were jointly used to estimate TWS changes and to correct for bias and leakage errors using the additive and multiplicative correction approaches (Longuevergne et al., 2010). Scaling factors from CLM4.0 provided by the JPL website (<http://grace.jpl.nasa.gov/data/get-data/monthly-mass-grids-land/>) were also used to restore filtered GRACE signal of CSR, JPL, and GFZ SH coefficients using the scaling factor approach (Landerer and Swenson, 2012). JPL mascons-based TWS changes in the form of $3^\circ \times 3^\circ$ equal-area caps (but were downscaled to $0.5^\circ \times 0.5^\circ$) for the study period were obtained for use in comparison with other products and as an input for the BMA approach.

For TCH, TWS changes from GRACE, LSMs, and GHMs for the overlapping period 2003–2009 were used to quantify their relative uncertainties. For BMA, the four-year period (2003–2006) was taken as the training period and the three-year period (2007–2009) was taken as the validation period. The mean value of each GRACE-based TWS change and GHM/LSM-based TWS time series for the period 2003–2009 was subtracted before being applied to BMA. Statistic metrics include root mean square difference (RMSD), mean absolute bias, coefficient of determination (R^2), and Nash-Sutcliffe efficiency (NSE) coefficient.

2.3. Study basins

Sixty river basins globally representing a broad range of climate, land cover, and irrigation extent, examined in a previous study (Long et al., 2015a) were also selected in this study (Fig. 1 and Table 1). Characterization of climate of the selected basins is based on the mean annual aridity index (AI), the mean of annual precipitation over annual potential evapotranspiration (Trabucco and Zomer, 2009), including humid

($AI > 0.65$), sub-humid ($AI \leq 0.65$ and > 0.5), semi-arid ($AI \leq 0.5$ and > 0.2), arid ($AI \leq 0.2$ and > 0.05), and hyperarid ($AI \leq 0.05$). In this study, arid and hyperarid are grouped into arid. Based on these criteria, 57% of the river basins (34) are categorized as humid climate, 15% (9) as sub-humid climate, 25% (15) as semi-arid, and 3% (2) as arid (including the hyperarid). Alternatively, these basins are categorized as large basins (basin area $> 1,000,000 \text{ km}^2$), medium-size basins ($200,000 \text{ km}^2 < \text{basin area} \leq 1,000,000 \text{ km}^2$), and small basins ($\text{basin area} \leq 200,000 \text{ km}^2$). Based on basin size, 35% of the basins (21) are grouped into large basins, 42% (25) into medium-size basins, and 23% (14) into small basins. Note that some of the study basins have been subject to large-scale groundwater depletion from irrigation, including the upper reach of the Arkansas basin (the Central High Plains), the Sacramento-San Joaquin River basin in California (Famiglietti et al., 2011; Scanlon et al., 2012), the Tigris basin in the Middle East (Joodaki et al., 2014; Voss et al., 2013), and the upper reach of the Ganges (Rodell et al., 2009).

3. Results and discussion

3.1. Uncertainties in total water storage changes using the TCH approach

Results indicate that GHMs show appreciably lower uncertainties (medium: 22 mm) in TWS change than LSMs (medium: 28 mm) and GRACE satellites (medium: 28 mm) for the sixty basins examined globally (Fig. 2). GRACE satellites and LSMs seem to generate TWS changes with the similar order of uncertainty. More specifically, WGHM generates the lowest uncertainty of 21.5 mm in all the model outputs and GRACE-derived products compared, followed by PCR-GLOBWB of 23.1 mm, which lends the confidence to use of GHMs TWS output as the reference for BMA to optimize different GRACE TWS change products.

For the four LSMs in GLDAS-1, Noah2.7 shows the lowest uncertainty of the medium of 23.7 mm, followed by VIC of 27.3 mm, and CLM 2.0 of 29 mm (Fig. 2). Mosaic exhibits the highest uncertainty of 32 mm. These results are consistent with published studies that Noah often performs best in simulating water storage changes than other LSMs in GLDAS-1 (e.g., Long et al., 2014a; Long et al., 2014b; Longuevergne et al., 2010), and Mosaic is often excluded from analysis (e.g., Castle et al., 2014; Voss et al., 2013). But this is not necessarily true

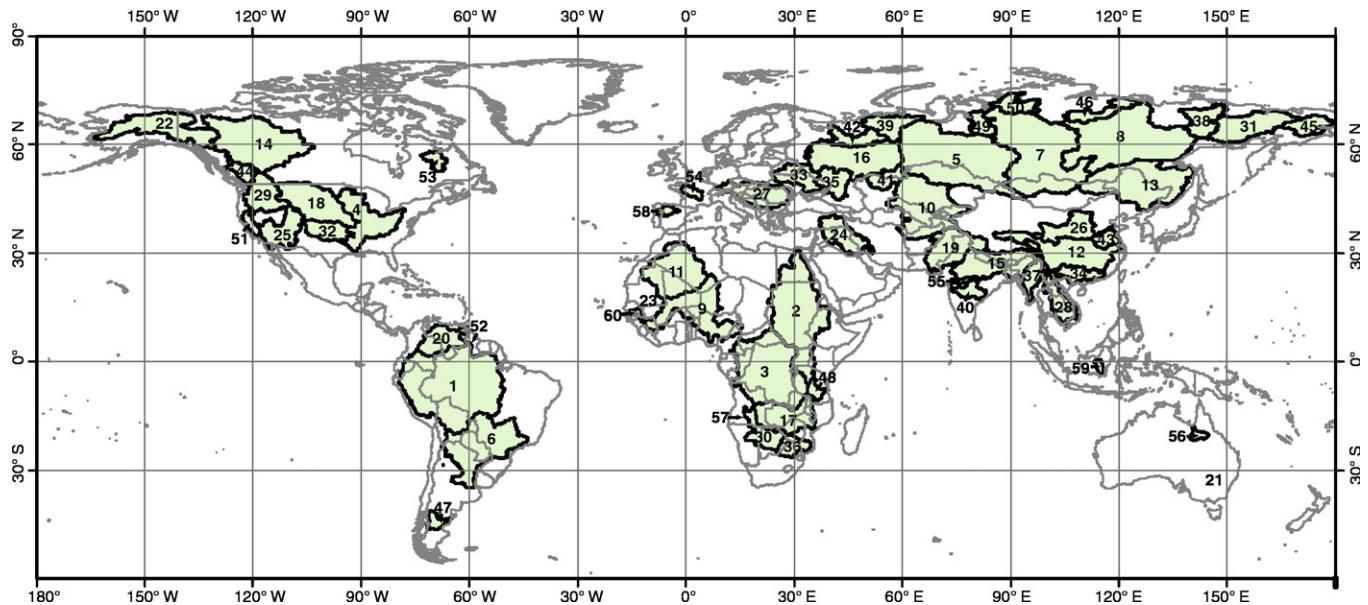


Fig. 1. Locations of 60 global river basins examined in this study. Corresponding basins' names are shown in Table 1. Note that basins 18 (Missouri) and 32 (Arkansas) are sub-basins of basin 4 (Mississippi).

This figure is adapted from (Long et al., 2015a).

Table 1

Characteristics of global river basins examined in this study. SA, SH, and H denote semi-arid, sub-humid, and humid, respectively. Total error is the monthly averaged root mean square of the measurement error and leakage error. Basins 1–21, 22–46, 47–60 are large, medium-size, and small basins, respectively. The gray lines show divisions of the large, medium-size, and small basins. This table is adapted from (Long et al., 2015a).

Basin ID	Name	Continent	Area (km ²)	Aridity index	Climate	Total error (mm)
1	Amazon	South America	5,854,000	1.27	H	12.5
2	Nile	Africa	3,826,000	0.29	SA	8.5
3	Congo	Africa	3,699,000	0.89	H	13.0
4	Mississippi	North America	3,203,000	0.68	H	8.0
5	Ob	Asia	3,026,000	0.74	H	6.5
6	Parana	South America	2,664,000	0.71	H	16.6
7	Yenisei	Asia	2,579,000	0.82	H	7.1
8	Lena	Asia	2,442,000	0.78	H	9.5
9	Niger	Africa	2,240,000	0.32	SA	12.6
10	Aral	Asia	2,148,000	0.30	SA	5.2
11	Tamanrasset	Africa	1,818,000	0.02	A	5.9
12	Yangtze	Asia	1,794,000	0.98	H	15.5
13	Amur	Asia	1,755,000	0.77	H	6.9
14	Mackenzie	North America	1,695,000	0.76	H	11.9
15	Ganges	Asia	1,628,000	0.90	H	21.6
16	Volga	Europe	1,476,000	0.94	H	14.6
17	Zambezi	Africa	1,364,000	0.54	SH	20.0
18	Missouri	North America	1,352,000	0.49	SA	8.4
19	Indus	Asia	1,143,000	0.37	SA	18.2
20	Orinoco	South America	1,039,000	1.33	H	25.2
21	Murray	Oceania	1,032,000	0.35	SA	13.5
22	Yukon	North America	856,000	0.68	H	18.4
23	Senegal	Africa	847,000	0.16	A	11.3
24	Tigris	Asia	823,000	0.28	SA	12.3
25	Colorado	North America	808,000	0.25	SA	9.4
26	Huanghe	Asia	795,000	0.50	SH	10.8
27	Danube	Europe	788,000	0.92	H	17.9
28	Mekong	Asia	759,000	1.02	H	28.5
29	Columbia	North America	732,000	0.68	H	16.1
30	Okavango	Africa	701,000	0.29	SA	20.9
31	Kolyma	Asia	666,000	0.75	H	15.1
32	Arkansas	North America	544,000	0.54	SH	15.1
33	Dnieper	Europe	500,000	0.80	H	19.7
34	Zhujiang	Asia	450,000	1.18	H	31.2
35	Don	Asia	423,000	0.62	SH	19.9
36	Limpopo	Africa	420,000	0.33	SA	16.5
37	Irrawaddy	Asia	406,000	1.24	H	41.6
38	Indigirk	Asia	334,000	0.73	H	17.1
39	Pechora	Asia	314,000	1.24	H	27.1
40	Godavari	Asia	312,000	0.64	SH	34.5
41	Ural	Asia	296,000	0.37	SA	16.4
42	N. Davina	Asia	288,000	1.11	H	31.4
43	Huai	Asia	244,000	0.73	H	28.6
44	Fraser	North America	238,000	0.97	H	32.4
45	Anadyr	Asia (Russia)	226,000	1.00	H	17.2
46	Olenek	Asia (Russia)	223,000	0.75	H	21.7
47	Chubut	South America	197,000	0.23	SA	29.1
48	Rufiji	Africa	187,000	0.62	SH	32.4
49	Taz	Asia (Russia)	171,000	1.34	H	34.4
50	Pyasina	Asia (Russia)	164,000	1.35	H	27.6
51	Central Valley	North America	156,000	0.55	SH	25.2
52	Essequibo	South America	151,000	1.17	H	62.6
53	Koksoak	North America	150,000	1.69	H	33.7
54	Loire	Europe	118,000	0.88	H	34.8
55	Narmada	Asia	114,000	0.64	SH	51.7
56	Flinders	Oceania	110,000	0.27	SA	26.2
57	Cunene	Africa	98,000	0.40	SA	45.6
58	Douro	Europe	97,000	0.64	SH	35.1
59	Barito	Asia	74,000	2.02	H	48.3
60	Gambia	Africa	72,000	0.46	SA	36.6

for some specific region, e.g., the Middle East where Mosaic performs best in simulating TWS changes (Joodaki et al., 2014). Uncertainties in TWS changes from these LSMs are comparable with those from GRACE satellite observations. This is especially important because it demonstrates that GRACE-based observations can provide valuable

TWS change estimates in an effective manner, which reduces efforts in data preparation and model implementation involved in LSM simulation. Regarding the three gridded SH solutions, CSR shows the lowest uncertainty of 26.7 mm, followed by JPL of 27.4 mm, and GFZ of 29.8 mm.

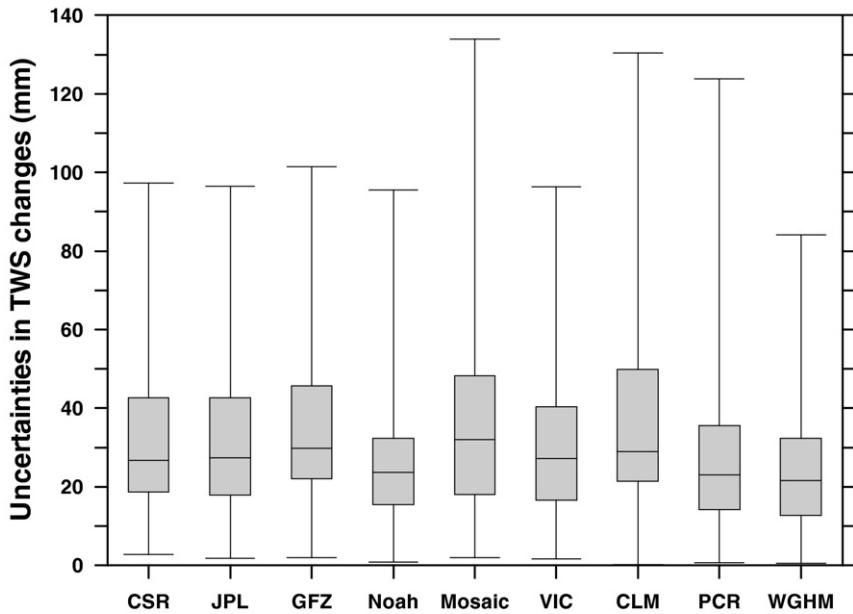


Fig. 2. Uncertainties (mm) in TWS changes from GRACE observations (CSR, JPL, and GFZ gridded SH, RL05), four LSMs in GLDAS-1 (Noah, Mosaic, VIC, and CLM), and two GHMs (PCR-GLOBWB and WGHM).

Overall, GRACE-derived TWS changes show the similar magnitude of uncertainty as the LSMs. GHMs can provide generally lower uncertainty than GRACE observations and LSM simulations. In the two GHMs, WGHM TWS output seems to be of lower uncertainty than PCR-GLOBWB. This finding reinforces the necessity of using historical simulations of TWS changes as a priori information to merge different GRACE TWS change products so that information on water storage changes from sophisticated GHMs can be carried away by GRACE observations. In doing so, advantages of satellite observations and model simulations can be complementary and incorporated in the merged TWS changes using BMA illustrated in the following section.

3.2. Evaluation of total water storage changes from GRACE and BMA

We evaluated TWS changes from GRACE and BMA using different TWS_R for training and validation, which provides a comprehensive and thorough understanding of how BMA works to merge different GRACE TWS changes. Given that WGHM, PCR-GLOBWB, and Noah show relatively lower uncertainties of TWS output (21.5 mm, 23.1 mm, and 23.7 mm, respectively) over the 60 basins globally than other LSM outputs and GRACE products, the three models were taken as references for training (2003–2006) and validating (2007–2009) the BMA algorithm under 9 scenarios (referring to Table 2).

3.2.1. Statistics for all 60 basins globally

We first discuss results of training and evaluating the BMA algorithm using WGHM TWS output (i.e., scenario 1 in Table 2). Weights of BMA (u_i in Eq. (6)) were first derived using six TWS change time series (i.e., the additive correction, multiplicative correction, three gridded SH products and mascons) from GRACE and WGHM TWS output for the period 2003–2006 (i.e., the training period) (Fig. 3a). Results show that the additive correction-based TWS changes are allocated the highest mean weight of 0.2 for all the 60 basins, followed by the mascons of 0.186. The three gridded SH products have similar weights ranging from 0.16–0.17 (CSR: 0.168, JPL: 0.172, and GFZ: 0.163). The multiplicative product is, however, allocated the lowest weight of 0.11, indicating that it deviates most from the WGHM output and other products (Fig. 3a). Given that TWS changes from the multiplicative correction approach show the largest deviation from other GRACE products and WGHM output (Long et al., 2015a) and the lowest weight (0.11) in this study, the BMA approach using only five GRACE TWS change products after excluding the multiplicative correction (i.e., the additive correction, three gridded SH products, and mascons) was examined to determine if the simulation skill of TWS changes could be improved.

Results indicate that the additive correction, again, is allocated the highest weight of 0.232, followed by the mascons of 0.207. The three gridded SH products show similar weights (CSR: 0.187; JPL: 0.192; and GFZ: 0.182) (Fig. 3b). The overall performance of BMA trained and

Table 2
Evaluation of the BMA approach trained by WGHM, PCR-GLOBWB, and Noah for the period (2003–2006) using WGHM, PCR, and Noah as references for the period (2007–2009), with showing mediums and means of NSE, R^2 , RMSD, and bias and their rankings in all products evaluated for each scenario (i.e., additive correction, CSR, JPL, GFZ gridded products, mascons, and BMA) for the 60 basins globally. Note that the rankings of NSE and R^2 are in the descending order and those for RMSD and bias are in the ascending order.

Scenario	Training	Validation	NSE medium (ranking)/mean (ranking)	R^2 medium (ranking)/mean (ranking)	RMSD (mm) medium (ranking)/mean (ranking)	Bias (mm) medium (ranking)/mean (ranking)
1	WGHM	WGHM	0.714(1)/0.656(1)	0.791 (2)/0.746(1)	30.4(3)/33.0(2)	25.0(2)/27.7(2)
2	PCR		0.691(2)/0.650(2)	0.794(1)/0.746(2)	30.3(2)/33.3(4)	25.3(4)/28.0(3)
3	Noah		0.686(3)/0.647(3)	0.790(3)/0.745(3)	30.5(4)/30.3(3)	25.3(3)/28.0(4)
4	WGHM	PCR	0.716(2)/0.525(3)	0.813(1)/0.743(1)	28.0(4)/33.1(4)	23.1(2)/27.4(4)
5	PCR		0.727(1)/0.546(1)	0.810(3)/0.743(2)	27.9(3)/32.8(2)	23.5(4)/27.2(2)
6	Noah		0.714(4)/0.524(4)	0.811(2)/0.743(3)	27.7(2)/33.0(3)	23.3(3)/27.4(3)
7	WGHM	Noah	0.734(2)/0.601(2)	0.777(2)/0.726(1)	28.7(4)/34.2(4)	23.9(3)/28.0(4)
8	PCR		0.733(3)/0.607(1)	0.776(3)/0.725(3)	28.5(3)/34.2(3)	24.1(4)/27.9(3)
9	Noah		0.734(1)/0.600(3)	0.779(1)/0.726(2)	28.3(2)/33.9(2)	23.4(2)/27.6(2)

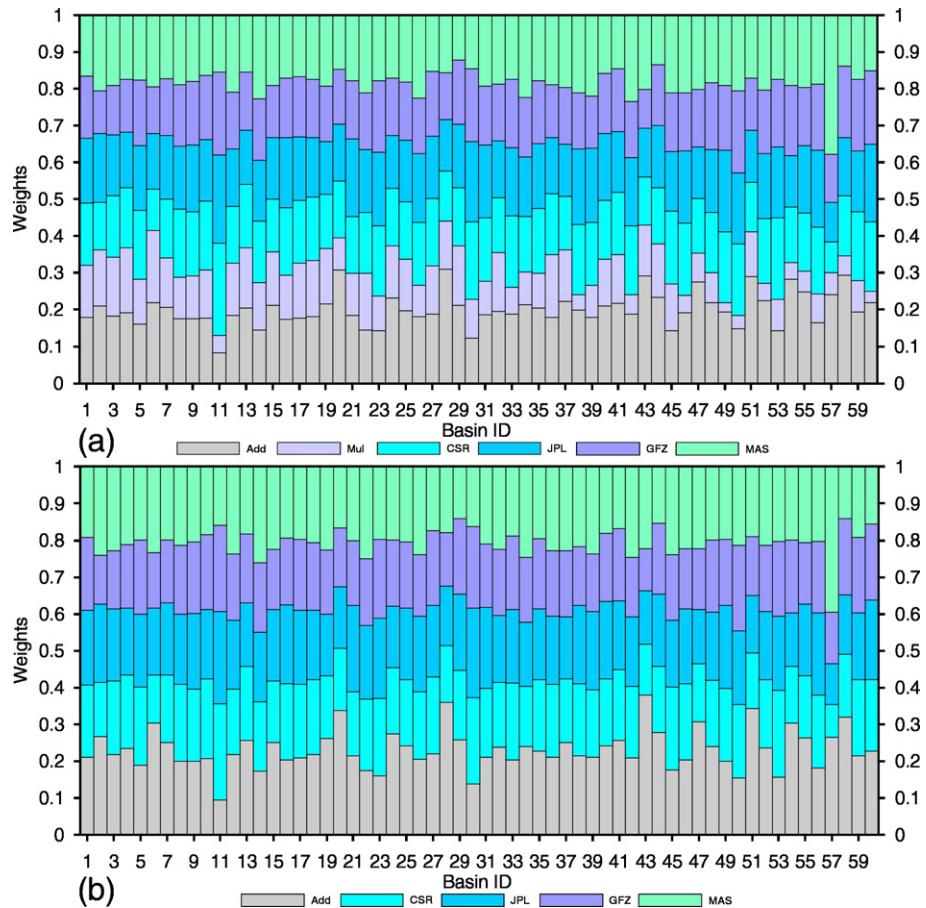


Fig. 3. Weights of the BMA approach for different products: (a) additive (Add), multiplicative (Mul), CSR gridded SH (CSR), JPL gridded SH (JPL), GFZ gridded SH (GFZ), and mascons (MAS); (b) additive (Add), CSR gridded SH (CSR), JPL gridded SH (JPL), GFZ gridded SH (GFZ), and mascons (MAS).

evaluated by WGHM (scenario 1) is better than other GRACE TWS changes in terms of the highest medium (0.714) and mean (0.656) of NSE, and the second lowest mean of RMSD (33.0 mm) and bias (27.2 mm) for the 60 basins globally (Table 2 and Fig. 4(a)–(d)). Despite the lowest medians of NSE (0.617) and R^2 of (0.73), the additive correction approach resulted in the lowest medians of RMSD (28.3 mm) and absolute bias (24.2) mm among the six products compared. This means that though the timing of TWS changes from the additive correction may deviate from the other GRACE TWS changes and WGHM output, the overall amplitude of TWS changes from the additive correction approach is more consistent with the WGHM output. This may explain the reason of the highest allocated weight for the additive correction approach (0.232) in merging different TWS change products based on BMA. The reason of the high consistency of the amplitude of the TWS changes between the additive correction approach and WGHM output may be attributed to the fact that the additive correction uses a LSM (Noah in this study) for bias and leakage correction. Results here highlight that each product has its advantages and limitations. The BMA approach is an effective way to merge all the available products in terms of the consistency in both amplitude and timing relative to a reference model or observations to generate an output that includes their strengths but reduces uncertainties arising from the algorithms used for GRACE signal restoration.

The three gridded SH products perform similarly in terms of all statistical metrics, with the CSR gridded SH product showing a relatively higher medium of NSE (0.674) than JPL (0.660) and GFZ (0.657) (see Fig. 4(a1)). The mascons show an improved median of NSE (0.682) relative to all the GRACE TWS changes (Add: 0.617; CSR: 0.674; JPL: 0.660; GFZ: 0.657), and a lower median of RMSD (30.9 mm) and absolute bias (26.0 mm) than the three gridded SH

products. This reinforces that the mascon solution further improves TWS change estimates compared with the three gridded SH products by taking the WGHM output as the reference. In particular, the dependence of gridded SH products on the scaling factor from LSMs has been greatly reduced.

The BMA algorithm trained by PCR-GLOBWB (scenario 2) and Noah (scenario 3) and evaluated by WGHM output generated the second (0.691 or 0.650) and the third (0.686 or 0.647) highest medium or mean of NSE, respectively (Table 2), higher than other individual GRACE TWS change products. This means that the BMA approach can be effective in merging different GRACE TWS change products, given that the reference data for training are shown to be of lower uncertainties than the GRACE TWS changes to be used in the BMA approach. It is also apparent that depending on the reference data for training the BMA algorithm, the BMA-based TWS changes can be slightly different in this case. This implies that uncertainties for the BMA algorithm exist if the poor reference data for training are used. Use of the TCH approach illustrated in Section 3.1 is one of the effective ways to quantify relative uncertainties in different products and determine which product can be used as TWS_R.

We evaluated the BMA approach using PCR-GLOBWB (scenarios 4–6, trained by WGHM, PCR-GLOBWB, and Noah, respectively) and Noah (scenarios 7–9, trained by WGHM, PCR-GLOBWB, and Noah, respectively). Similar to scenarios 1–3, the BMA algorithm provided the highest medium of NSE (0.727, scenario 5), the second highest medium of NSE (0.716, scenario 4), and the fourth highest medium of NSE (0.714, scenario 6), for validation using PCR-GLOBWB output, and generally higher than other individual GRACE TWS changes. For validation with Noah TWS output, BMA provided the highest medium of NSE (0.734, scenario 9), the second highest medium of NSE (0.734, scenario 7),

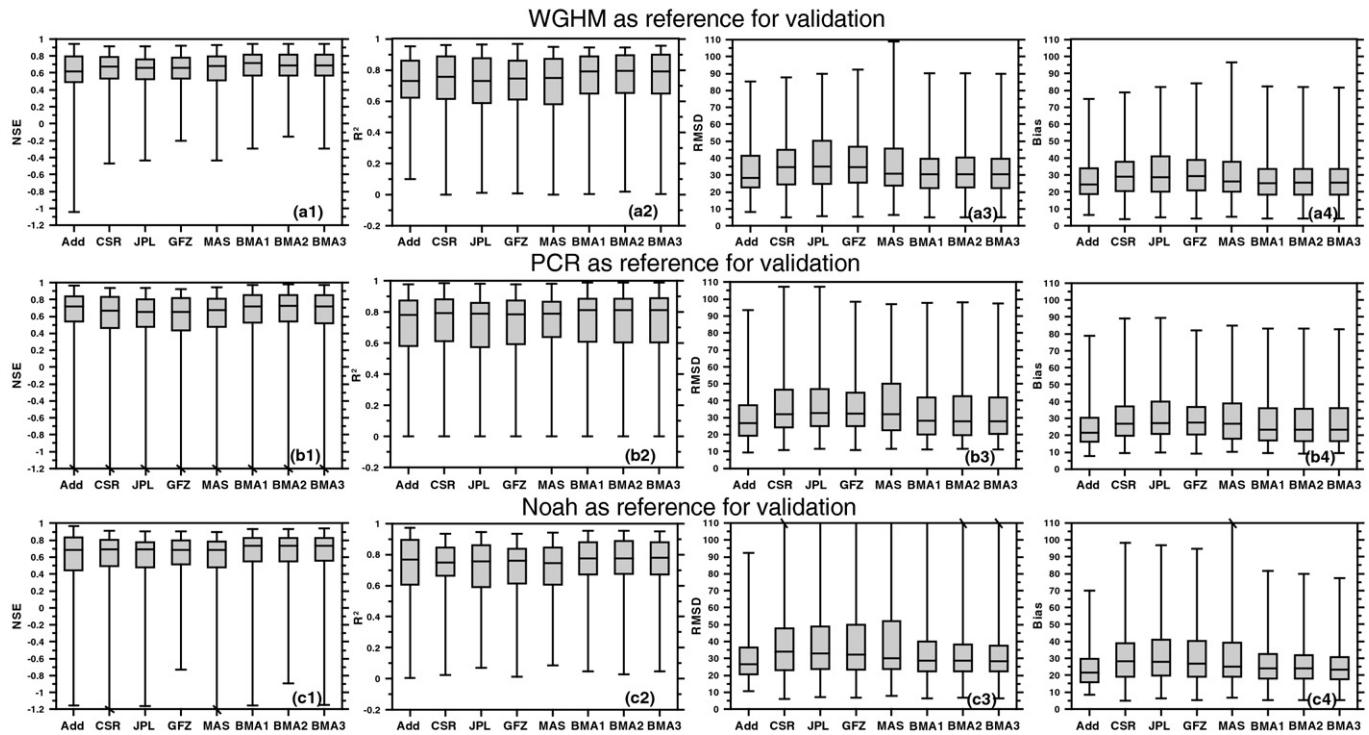


Fig. 4. Boxplots of Nash-Sutcliffe efficiency (NSE, a(1)–c(1)) coefficients, coefficient of determination (R^2 , a(2)–c(2)), root mean square difference (RMSD, a(3)–c(3)), and mean absolute bias (a(4)–c(4)) of GRACE-based TWS anomalies from the BMA approach (using five GRACE TWS anomalies, i.e., additive (Add), CSR gridded SH (CSR), JPL gridded SH (JPL), GFZ gridded SH (GFZ), mascons (MAS)) and each individual GRACE TWS anomalies with reference to WGHM (the first horizontal panel), PCR-GLOBWB (the second horizontal panel), and Noah TWS output (the third horizontal panel) for the validation period 2007–2009. BMA1, BMA2, and BMA3 denote training the BMA algorithm using WGHM, PCR-GLOBWB, and Noah TWS outputs for the period 2003–2006, respectively.

and the third highest medium of NSE (0.733, scenario 8), higher than other individual GRACE TWS changes. For scenarios 4–9, the additive correction approach provided TWS changes of the lowest RMSD and bias relative to the reference model output for validation, similar to scenarios 1–3. This again reinforces the necessity of using BMA to reduce uncertainties in different GRACE TWS changes.

3.2.2. Statistics at the basin scale

Given that WGHM TWS output is completely independent of GRACE-observed TWS changes, and the statistical evidence of the lowest uncertainty of WGHM TWS output using the TCH approach (Section 3.1), we used BMA-based TWS changes from scenario 1 (i.e., trained and validated by WGHM output for different periods) for the

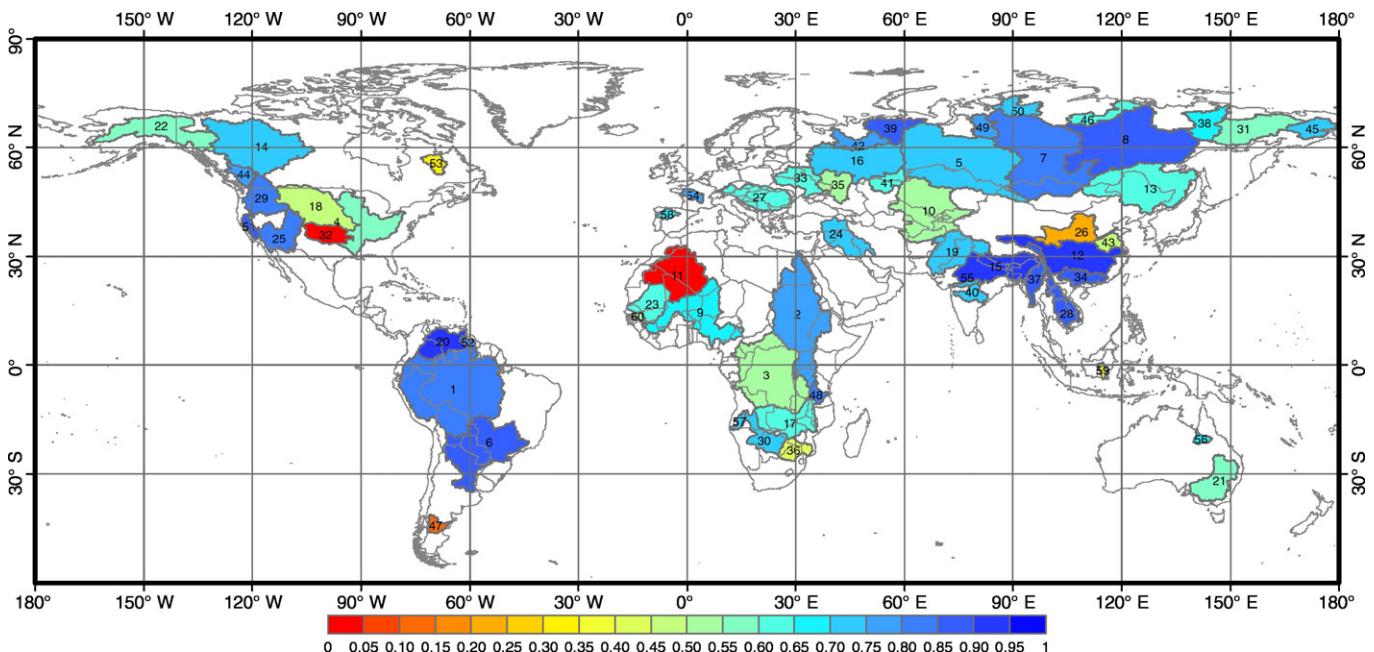


Fig. 5. Global distribution of NSE values of the merged TWS changes using BMA with reference to WGHM TWS change output for the validation period 2007–2009.

following discussion. The mascons-based TWS changes were also compared for the training, validation, and an extended period (Jan 2010–Mar 2015). Overall, 32 out of the 60 basins (53%) show NSE values higher than 0.7 (excellent), and 43 out of the 60 basins (72%) show NSE values higher than 0.6 (acceptable). In particular, the BMA-based TWS changes compare well with WGHM output during the period 2007–2009 for the Orinoco (NSE = 0.945), Changjiang (NSE = 0.938), Ganges (NSE = 0.913), Pechora (NSE = 0.90), Mekong (NSE = 0.894), Lena (NSE = 0.891), Zhuijiang (NSE = 0.88), Central Valley (NSE = 0.88), and Parana (NSE = 0.87) basins (see Figs. 5 and 6). Across these basins, the BMA-based TWS changes are well consistent with the WGHM output in terms of both amplitude and timing. In addition, the mascons are also similar to the BMA-based counterparts, though there are appreciable differences for the Central Valley where the mascons show a greater decreasing trend, particularly for the extended period (Fig. 6).

On the other hand, there are 28% basins (17) where BMA does not perform well compared with the WGHM output. In particular, Tamanraset (NSE = −0.298), Arkansas (NSE = 0.049), Chubut (NSE = 0.141), Huanghe (NSE = 0.226), Koksoak (NSE = 0.309), Barito (NSE = 0.371), Limpopo (NSE = 0.441), Missouri (NSE = 0.453), and Huai (NSE = 0.48) show relatively low consistency between the BMA-based TWS changes and WGHM TWS output (Figs. 5 and 7). This is primarily attributed to the fact that all TWS change products and signal restoration approaches deviate greatly from the model output over these basins, resulting in the poor performance of BMA for merging TWS changes. For instance, the overall mean of NSE for the additive (0.089), CSR gridded SH (−0.469), JPL gridded SH (−0.432), GFZ gridded SH (−0.18), and mascons (−0.253) for the Tamanraset basin is only −0.257, resulting in the low NSE of −0.298 for the BMA approach.

All of basins with lower NSE for BMA are extremely dry (e.g., Tamanraset), intensively irrigated basins (e.g., the upper reach of the Arkansas basin), or impacted by regions with marked cryospheric processes (e.g., the Chubut basin impacted by glacier melting of Patagonia). In these cases, the BMA approach may not work well due to the poor

reference (i.e., WGHM TWS output in this study) and/or the lower signal to noise ratio of GRACE observations. Over extremely dry areas, uncertainties in model forcing and model parameters associated with evapotranspiration and soil moisture may result in large uncertainties in TWS estimates from hydrological models (Long et al., 2014a; Long et al., 2013). In addition, GHMs may not perform well over cryospheric regions due to limitations in model physics. Therefore, the low consistency between model output and GRACE-derived TWS changes may not necessarily reflect the low reliability of GRACE-derived TWS changes. Over areas with intensive irrigation, more in situ water-level measurements are needed to evaluate the reliability of GRACE-derived TWS changes. Under the condition that BMA does not perform well due to the poor choice of reference, more in situ measurements or a better choice of TWS_R for a specific basin may be needed to better merge different TWS changes.

3.3. Trends in TWS changes for sixty basins globally from different approaches

3.3.1. Three gridded SH products

Trends in TWS changes derived from three gridded SH products for the 60 basins for the period Apr 2002–Mar 2015 are shown in Fig. 8, indicating high consistency among the three gridded SH products. Specifically, 45 basins show statistical significance ($p = 0.05$) and 2 basins show statistical insignificance indicated by all the three products. However, there are still 13 basins (4 were large river basins, i.e., the Nile (ID = 2), Congo (ID = 3), Ob (ID = 5), and Lena (ID = 8) and 5 were medium-size basins, i.e., Danube (ID = 27), Limpopo (ID = 36), Indigirk (ID = 38), Pechora (ID = 39), and N. Davina (ID = 42), referring to Fig. 1) where the three gridded SH products indicate inconsistent statistical significance. This results from generally low magnitudes in TWS trends due to the fact that (1) depletion and recovery in TWS were essentially compensated during the study period and/or (2) depletion and recovery in TWS were compensated by the spatial averaging effect. Therefore, the statistical test of trends in TWS over these basins is fairly sensitive to processing strategies of GRACE data.

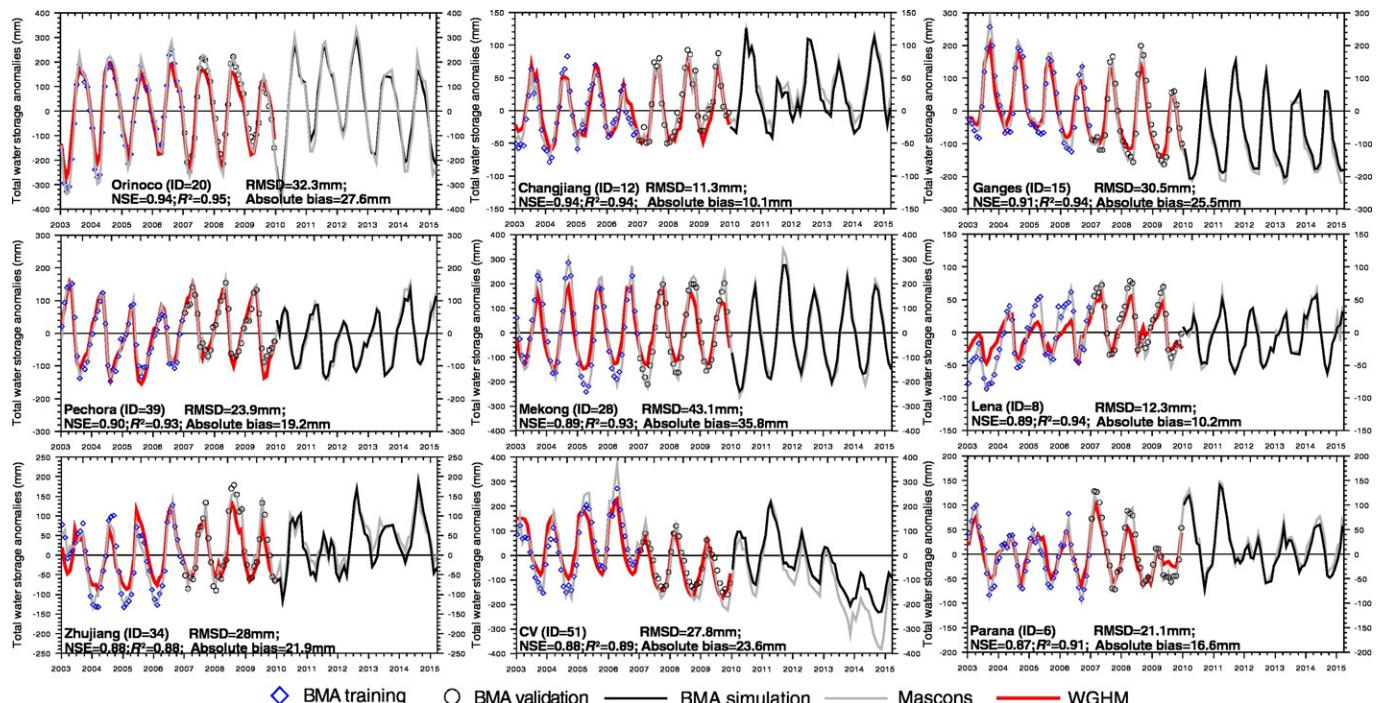


Fig. 6. Training (2003–2006), validation (2007–2009), and an extend period (2010–2015) of TWS changes from BMA for nine basins (i.e., Orinoco, Changjiang, Ganges, Pechora, Mekong, Lena, Zhuijiang, Central Valley, and Parana) with relatively better performance, showing statistical metrics including NSE, R², RMSD, and mean absolute bias, and corresponding WGHM TWS change output (2003–2009) as the reference and mascons for the entire study period.

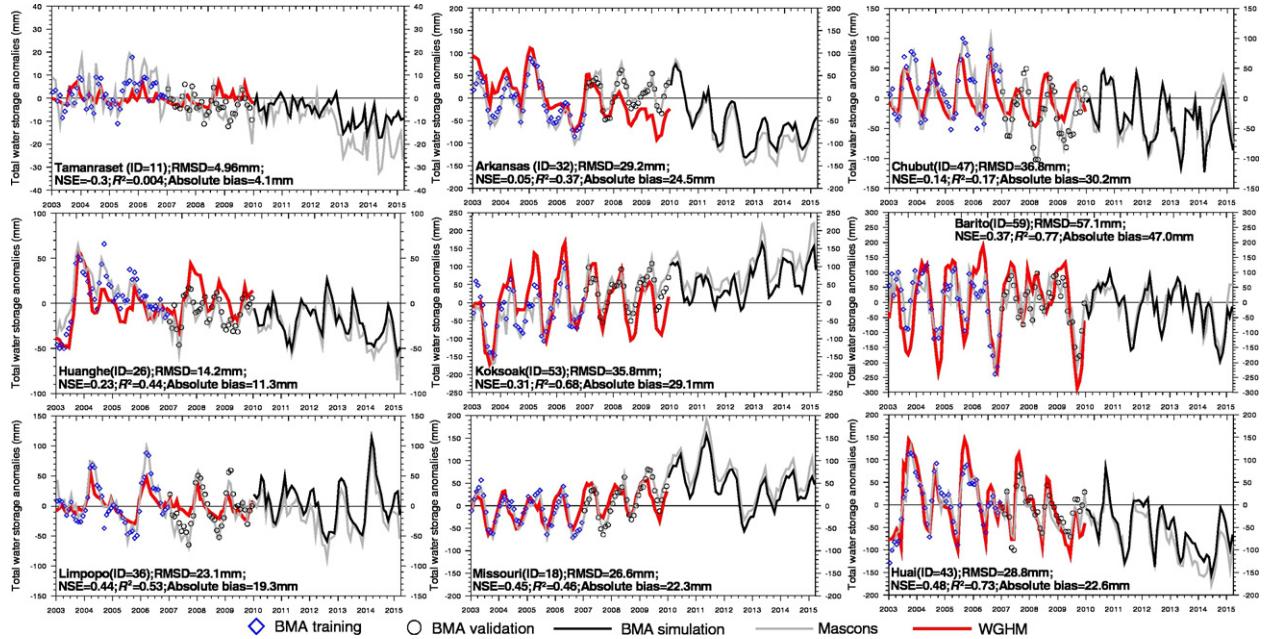


Fig. 7. Training (2003–2006), validation (2007–2009), and an extended periods (2010–2015) of TWS changes from BMA for nine basins (i.e., Tamanrasset, Arkansas, Chubut, Huanghe, Koksoak, Barito, Limpopo, Missouri, and Huai) with relatively poor performance, showing statistical metrics of NSE, R^2 , RMSD, and absolute bias, and corresponding WGHM TWS change output (2003–2009) as the reference and mascons for the entire study period.

The three gridded SH products still show consistent statistical tests of trends in large basins that are greatly impacted by climate change or extremes, e.g., the Amazon basin that showed a significantly increasing trend of ~ 4.3 mm/a due to the recovery of the drought from 2010 (Lewis et al., 2011). Overall, the gridded SH products are generally consistent. In the following discussion, we focus mainly on comparison of the CSR gridded SH product with TWS changes from the additive, mascons, and BMA approaches.

3.3.2. Comparison of BMA with other products

Trends in TWS derived from the CSR gridded SH, additive correction, mascons, and BMA approaches at the basin scale are shown in Fig. 9. The four approaches generate relatively more disparate trends compared with the three gridded SH products shown in Section 3.3.1. First, there are 42 basins for which the four approaches-derived TWS changes

show consistent statistical test results (statistical significance: 40 and statistical insignificance: 2). Second, standard deviations of the trends from the four approaches are generally higher in small basins, indicating a larger disparity in TWS trends over small basins. Standard deviations in small basins account for 50% (5) of the top 10 highest standard deviations. This is likely caused by the fact that the spatial heterogeneity in TWS changes may be amplified in small basins so that differences in GRACE processing strategy are more likely to be manifested (Long et al., 2015a). Third, basins with marked cryospheric processes (e.g., Yukon, Mackenzie, and Fraser), intensive irrigation (e.g., Central Valley and Tigris), and basins impacted by surrounding regions that are largely affected by cryospheric processes (e.g., the Irrawaddy basin affected by glacier melting over the eastern Himalayan region in the Tibetan Plateau) show relatively large differences in TWS trends. This is because for these basins, LSM output may only represent limited knowledge of

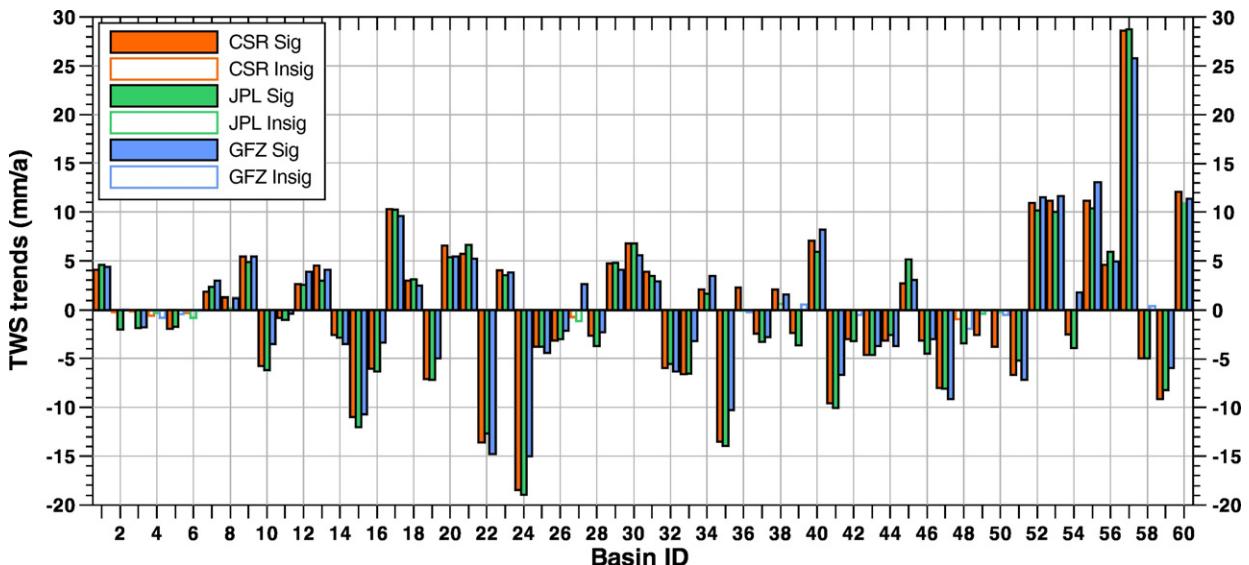


Fig. 8. TWS trends derived from three gridded SH products (CSR, JPL, and GFZ) for 60 basins examined in this study. 'Sig' in the legend means statistical significance ($p < 0.05$); 'Insig' means statistical insignificance ($p \geq 0.05$).

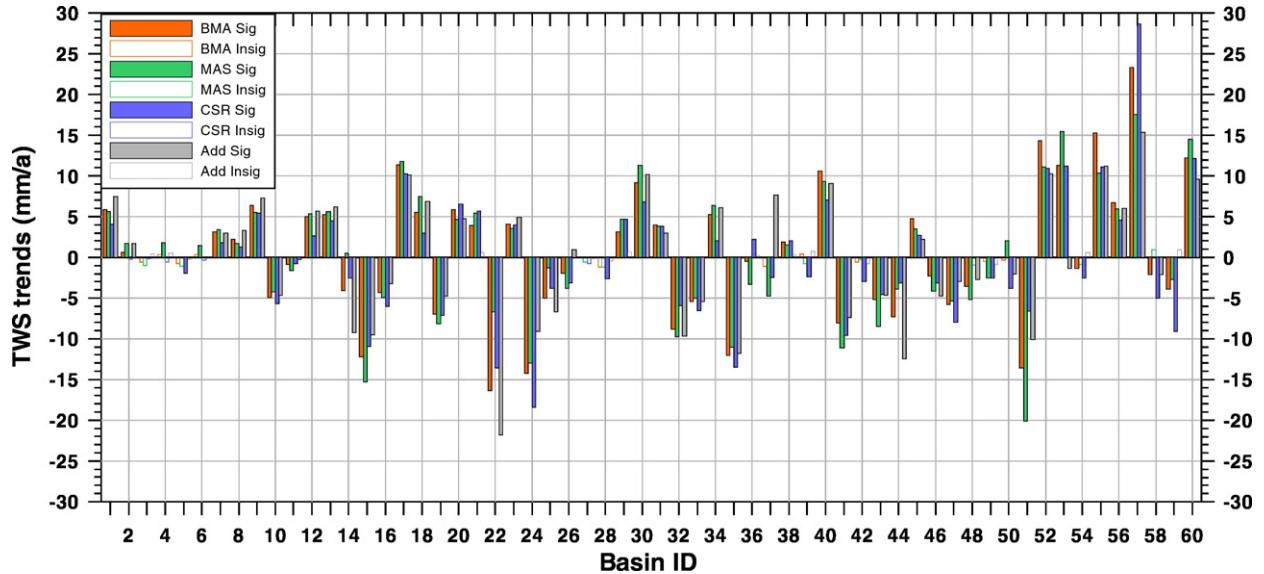


Fig. 9. TWS trends derived from the mascons, CSR gridded SH product, additive correction, and BMA for 60 basins examined in this study. 'Sig' in the legend means statistical significance ($p < 0.05$); 'Insig' means statistical insignificance ($p \geq 0.05$).

the realistic hydrological processes. TWS changes derived from the approaches that depend on a certain LSM, i.e., the additive approach depends on a LSM for computing bias and leakage, and the CSR gridded SH product requires scaling factors based on CLM4.0 output to rescale the filtered TWS changes from GRACE, may apparently deviate from the mascons that do not require a priori information provided by LSM output. Fourth, the BMA-based trends are generally within the range of the five products because of its nature for merging all products based on WGHM output as the reference.

3.4. Basins with mostly decreasing or increasing TWS trends derived from the mascons

Based on the mascons for the period Apr 2002–Mar 2015, the top 10 decreasing rates in TWS are found in the (1) Central Valley (-20.1 ± 1.5 mm/a), (2) Ganges (-15.2 ± 0.6 mm/a), (3) Tigris ($-13.0 \pm$

0.6 mm/a), (4) Ural (-11.2 ± 0.5 mm/a), (5) Don (-11.0 ± 0.8 mm/a), (6) Arkansas (-9.7 ± 0.7 mm/a), (7) Huai (-8.5 ± 0.8 mm/a), (8) Indus (-8.1 ± 0.3 mm/a), (9) Yukon (-6.7 ± 0.4 mm/a), and (10) Chubut (-5.4 ± 0.3 mm/a) (Figs. 10 and 11, and Table 3). Overall, the relatively higher decreasing rates are found in (1) basins that were likely impacted by significant climate variability and extremes, e.g., glacier melting in the Yukon and Chubut (Jacob et al., 2012) and the heat waves and drought in Russia and Kazakhstan (the Ural and Don basins) in 2010 when TWS reached lowest during the entire period. However, causes for the persistent drying trends from 2007 through 2010 for the Ural and Don basins need to be further investigated; (2) basins impacted by intensive human activities (e.g., irrigation) for the Central Valley (Famiglietti et al., 2011; Scanlon et al., 2012), Tigris (Longuevergne et al., 2013; Voss et al., 2013), Arkansas (i.e., the central High Plains Aquifer) (Longuevergne et al., 2010; Strassberg et al., 2009), and Indus (Rodell et al., 2009); and (3) basins jointly affected by climate

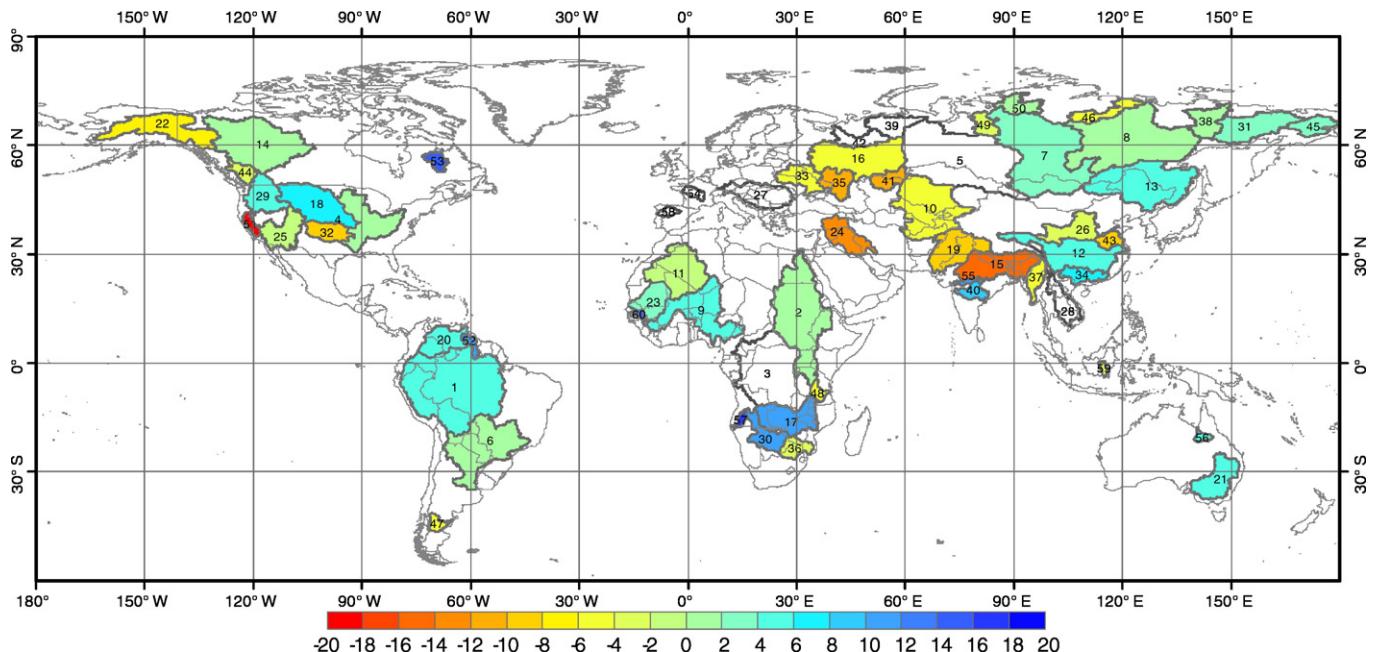


Fig. 10. Global distribution of basins with statistically significant ($p < 0.05$) trends (mm/a) in TWS derived from the mascons for the period Apr 2002–Mar 2015.

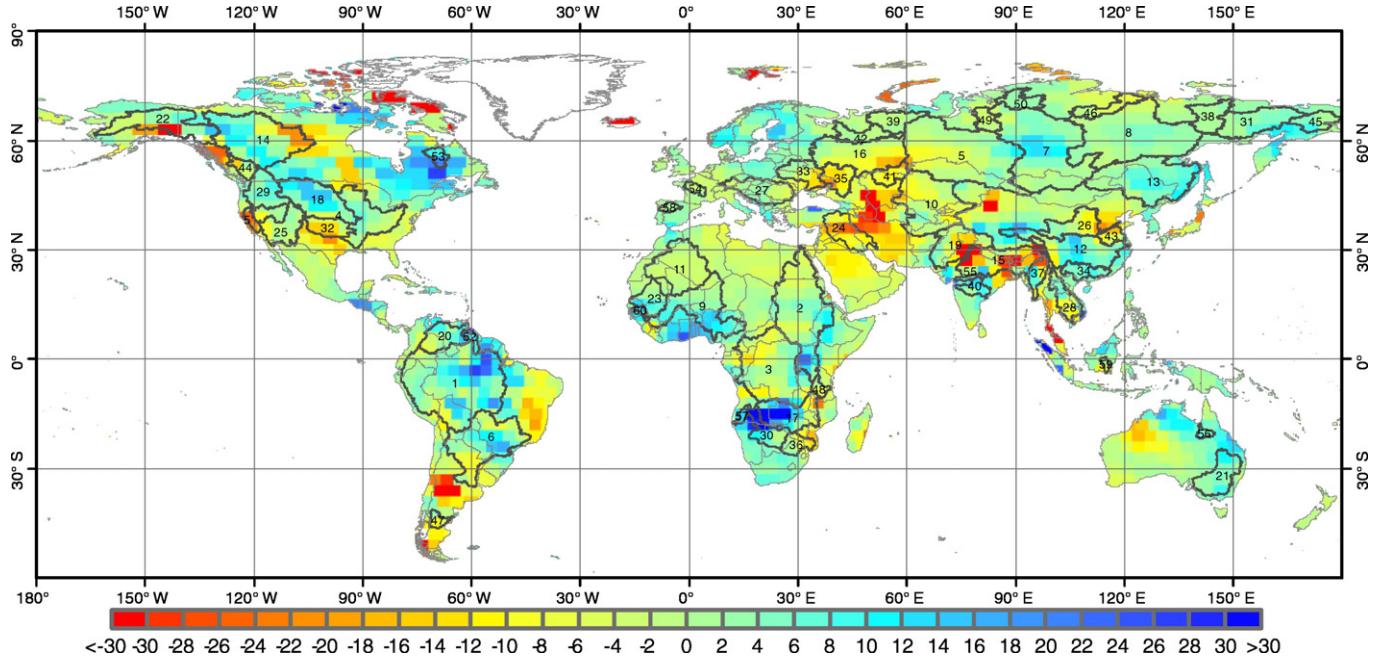


Fig. 11. Global distribution of trends (mm/a) in TWS derived from the mascons for the period Apr 2002–Mar 2015.

variability and human activities, e.g., the Ganges whose groundwater storage has been largely depleted (Long et al., 2016; Rodell et al., 2009) and showed glacier melting in the east Himalayan region (Jacob et al., 2012). Reasons for TWS depletion over the Huai River are not clear and need further investigation. Note that there are some other regions showing significant decreasing trends in TWS (e.g., south Parana and north Chubut, northeast Parana basin, the Caspian Sea located in the south Don and Ural basins and northeast Tigris basin, and the east of the Tianshan Mountains in northwest China) (Fig. 11).

Increases in TWS merit consideration, which can suppress global sea level rise (Reager et al., 2016). Basins showing significant increasing trends in TWS are distributed mainly in (1) the South Africa region (e.g., Cunene: 17.5 ± 1.0 mm/a, Zambezi: 11.8 ± 0.9 mm/a, and Okavango: 11.3 ± 0.4 mm/a), (2) the west Africa region (e.g., Gambia: 14.5 ± 1.0 mm/a and Niger: 5.5 ± 0.3 mm/a), (3) North America (e.g., Missouri: 7.4 ± 0.7 mm/a and Koksoak: 15.4 ± 0.4 mm/a), (4) the north of South America (e.g., Essequibo: 11.1 ± 1.5 mm/a) and Amazon (5.6 ± 0.8 mm/a), (5) central India (e.g., Narmada: 10.3 ± 0.8 mm/a and Godavari: 9.3 ± 0.8 mm/a), and (6) the north Tibetan Plateau (17.4 mm/a) and central Yangtze River basin (14 mm/a) in China. Most of the increased TWS is attributed to increased precipitation (Reager et al., 2016). Human-induced surface water storage changes are also responsible for part of the increases in TWS, e.g., increases in TWS in the central Yangtze were caused primarily by water impoundment of the Three Gorges Reservoir (Long et al., 2015b). Increases in TWS over the north Tibetan Plateau were primarily characterized by lake expansion resulting from both increased precipitation and glacier melting (Yang et al., 2014).

3.5. Basins with mostly decreasing or increasing TWS trends derived from the BMA output

Based on BMA, the top 10 basins with mostly decreasing TWS trends are found in the (1) Yukon (-16.3 ± 0.5 mm/a), (2) Tigris (-14.3 ± 0.6 mm/a), (3) Central Valley (-13.6 ± 1.3 mm/a), (4) Ganges (-12.2 ± 0.6 mm/a), (5) Don (-12.1 ± 0.8 mm/a), (6) Arkansas (-8.8 ± 0.6 mm/a), (7) Ural (-8.0 ± 0.5 mm/a), (8) Fraser (-7.3 ± 1.2 mm/a), (9) Indus (-6.9 ± 0.2 mm/a), and (10) Dnieper basins (-5.4 ± 0.5 mm/a) (Figs. 12 and 13, and Table 3). Similar to the mascons, increases in TWS are generally distributed in the south

and west Africa regions (e.g., Cunene: 23.4 ± 1.2 mm/a, Zambezi: 11.4 ± 0.9 mm/a, Okavango: 9.2 ± 0.4 mm/a, Gambia: 12.2 ± 0.9 mm/a, and Niger: 6.4 ± 0.3 mm/a), (2) central India (Narmada: 15.2 ± 1.0 mm/a and Godavari: 10.6 ± 0.8 mm/a), (3) South America (e.g., Essequibo: 14.4 ± 1.6 mm/a), and (4) North America (e.g., Koksoak: 11.3 ± 0.4 mm/a and Missouri: 5.5 ± 0.6 mm/a) (Figs. 12 and 14, and Table 3).

In general, basins with significant decreasing or increasing trends derived from both the mascons and BMA are found to be distributed over the same region. However, the exact rates for TWS changes are notably different. Overall, the mean absolute relative difference in TWS decreasing rates between the mascons and BMA is ~35% for 8 basins in the top 10 with mostly decreasing TWS trends from both approaches. The difference is up to ~140% for the Yukon basin (MAS: -6.7 ± 0.4 mm/a; BMA: -16.3 ± 0.5 mm/a), ~32% for the Central Valley (MAS: -20.1 ± 1.5 mm/a; BMA: -13.6 ± 1.3 mm/a), ~30% for the Indus basin (-8.1 ± 0.3 mm/a for the MAS and -5.7 ± 0.4 mm/a for the BMA), ~29% for the Ural basin (MAS: -11.2 ± 0.5 mm/a; BMA: -8.0 ± 0.5 mm/a), and ~20% for the Ganges (MAS: -15.2 ± 0.6 mm/a; BMA: -12.2 ± 0.6 mm/a). This is also the case for basins showing relatively large increasing trends. The mean absolute relative difference in TWS increasing rates between the mascons and BMA is ~24%, and is up to ~48% for the Narmada (MAS: 10.3 ± 0.8 mm/a; BMA: 15.3 ± 1.0 mm/a), ~34% for the Cunene (MAS: 17.5 ± 1.0 mm/a; BMA: 23.4 ± 1.2 mm/a), ~30% for the Essequibo (MAS: 11.1 ± 1.5 mm/a; BMA: 14.4 ± 1.6 mm/a), ~27% for the Koksoak (MAS: 15.4 ± 0.4 mm/a; BMA: 11.3 ± 0.4 mm/a), and ~19% for the Okavango (MAS: 11.3 ± 0.4 mm/a; BMA: 9.2 ± 0.4 mm/a). These differences indicate that different processing approaches for GRACE data may result in different magnitudes of increases or decreases in TWS trends. However, the general patterns of trends in TWS are almost the same at regional and global scales (Figs. 10–12). Because we incorporated a priori information on TWS changes from a GHM, information on human-induced variations can essentially be translated into the interpretation of BMA-based TWS changes.

3.6. EOF analysis of TWS changes from the mascons

Dominant seasonal and interannual variability and long-term trends are captured by the first four modes of the EOF analysis, which explain

Table 3

Characteristics of basins globally examined in this study and related statistics on trends in total water storage changes from the mascons (MAS) and BMA for the period Apr 2002–Mar 2015. 'Std.' means the standard deviation of linear regression for trends; 'Conf.' means the half value of the confidence interval; and 'R²' means the coefficient of determination. The gray lines show divisions of the large, medium-size, and small basins.

Basin ID	Name	Trends (mm/a)		Std. (mm/a)		Conf. (mm/a)		R ²		p value	
		MAS	BMA	MAS	BMA	MAS	BMA	MAS	BMA	MAS	BMA
1	Amazon	5.58	5.82	0.81	0.82	1.33	1.36	0.24	0.25	0.00	0.00
2	Nile	1.72	0.62	0.22	0.24	0.37	0.40	0.28	0.04	0.00	0.01
3	Congo	-1.00	-0.60	0.60	0.65	1.00	1.07	0.02	0.01	0.10	0.35
4	Mississippi	1.75	0.35	0.58	0.57	0.96	0.95	0.06	0.00	0.00	0.54
5	Ob	-1.10	-0.80	0.61	0.63	1.01	1.05	0.02	0.01	0.07	0.21
6	Parana	1.48	0.34	0.41	0.40	0.68	0.66	0.08	0.01	0.00	0.40
7	Yenisei	3.38	3.15	0.40	0.44	0.65	0.72	0.32	0.25	0.00	0.00
8	Lena	1.66	2.19	0.44	0.46	0.72	0.76	0.09	0.13	0.00	0.00
9	Niger	5.50	6.36	0.30	0.30	0.50	0.50	0.68	0.75	0.00	0.00
10	Aral	-4.26	-4.91	0.28	0.30	0.47	0.50	0.59	0.63	0.00	0.00
11	Tamanrasset	-1.61	-0.82	0.13	0.07	0.22	0.12	0.49	0.45	0.00	0.00
12	Yangtze	5.33	4.99	0.20	0.18	0.34	0.30	0.82	0.83	0.00	0.00
13	Amur	5.62	5.21	0.47	0.45	0.77	0.75	0.48	0.46	0.00	0.00
14	Mackenzie	0.53	-4.06	0.22	0.23	0.36	0.37	0.04	0.68	0.02	0.00
15	Ganges	-15.23	-12.18	0.55	0.56	0.91	0.93	0.83	0.75	0.00	0.00
16	Volga	-4.90	-4.34	0.59	0.54	0.98	0.90	0.31	0.29	0.00	0.00
17	Zambezi	11.82	11.36	0.94	0.93	1.55	1.54	0.51	0.49	0.00	0.00
18	Missouri	7.42	5.48	0.69	0.62	1.14	1.03	0.43	0.34	0.00	0.00
19	Indus	-8.14	-6.94	0.27	0.21	0.44	0.35	0.86	0.88	0.00	0.00
20	Orinoco	4.66	5.82	1.37	1.29	2.27	2.14	0.07	0.12	0.00	0.00
21	Murray	5.43	3.92	0.88	0.70	1.46	1.16	0.20	0.17	0.00	0.00
22	Yukon	-6.66	-16.34	0.44	0.46	0.73	0.75	0.60	0.89	0.00	0.00
23	Senegal	3.59	4.09	0.32	0.24	0.52	0.40	0.46	0.65	0.00	0.00
24	Tigris	-12.99	-14.29	0.64	0.57	1.06	0.94	0.73	0.81	0.00	0.00
25	Colorado	-1.26	-5.02	0.42	0.40	0.70	0.66	0.05	0.51	0.00	0.00
26	Huanghe	-3.83	-1.93	0.31	0.31	0.51	0.51	0.50	0.20	0.00	0.00
27	Danube	-0.57	0.02	0.76	0.67	1.25	1.12	0.00	0.00	0.46	0.97
28	Mekong	-1.17	-1.19	0.84	0.76	1.38	1.26	0.01	0.02	0.16	0.12
29	Columbia	4.68	3.10	0.67	0.57	1.10	0.95	0.24	0.16	0.00	0.00
30	Okavango	11.28	9.16	0.44	0.36	0.73	0.59	0.81	0.81	0.00	0.00
31	Kolyma	3.81	4.02	0.50	0.49	0.82	0.80	0.28	0.31	0.00	0.00
32	Arkansas	-9.73	-8.84	0.71	0.59	1.18	0.98	0.55	0.59	0.00	0.00
33	Dnieper	-5.01	-5.41	0.51	0.53	0.84	0.88	0.39	0.40	0.00	0.00
34	Zhujiang	6.34	5.25	0.54	0.52	0.90	0.86	0.47	0.40	0.00	0.00
35	Don	-11.03	-12.07	0.83	0.81	1.37	1.34	0.54	0.59	0.00	0.00
36	Limpopo	-3.34	-0.48	0.21	0.19	0.35	0.31	0.62	0.04	0.00	0.01
37	Irrawaddy	-4.74	-1.13	1.38	0.82	2.28	1.35	0.07	0.01	0.00	0.17
38	Indigirk	1.55	1.85	0.37	0.34	0.61	0.56	0.10	0.16	0.00	0.00
39	Pechora	-0.75	0.44	0.70	0.64	1.16	1.07	0.01	0.00	0.29	0.49
40	Godavari	9.29	10.58	0.82	0.81	1.36	1.34	0.45	0.52	0.00	0.00
41	Ural	-11.15	-8.02	0.50	0.54	0.82	0.89	0.77	0.59	0.00	0.00
42	N.Davina	-0.27	-0.64	0.58	0.54	0.97	0.89	0.00	0.01	0.64	0.23
43	Huai	-8.49	-5.15	0.81	0.82	1.34	1.36	0.41	0.20	0.00	0.00
44	Fraser	-3.94	-7.26	1.23	1.15	2.03	1.90	0.06	0.21	0.00	0.00
45	Anadyr	3.48	4.79	0.78	0.77	1.29	1.28	0.11	0.20	0.00	0.00
46	Olenek	-4.19	-2.28	0.41	0.43	0.68	0.71	0.41	0.16	0.00	0.00
47	Chubut	-5.36	-5.73	0.32	0.43	0.53	0.72	0.64	0.53	0.00	0.00
48	Rufiji	-5.19	-3.53	0.85	1.15	1.41	1.90	0.19	0.06	0.00	0.00
49	Taz	-2.55	-0.52	0.76	0.83	1.25	1.37	0.07	0.00	0.00	0.53
50	Pyasina	1.99	-0.37	0.79	0.69	1.31	1.14	0.04	0.00	0.01	0.59
51	CV	-20.08	-13.55	1.53	1.26	2.53	2.08	0.53	0.43	0.00	0.00
52	Essequibo	11.11	14.35	1.54	1.63	2.55	2.69	0.25	0.34	0.00	0.00
53	Koksoak	15.41	11.28	0.40	0.40	0.67	0.66	0.90	0.84	0.00	0.00
54	Loire	-0.85	-1.35	0.70	0.56	1.16	0.93	0.01	0.04	0.23	0.02
55	Narmada	10.32	15.22	0.80	0.98	1.32	1.62	0.52	0.61	0.00	0.00
56	Flinders	5.93	6.70	1.04	0.75	1.72	1.24	0.17	0.34	0.00	0.00
57	Cunene	17.53	23.36	1.01	1.21	1.66	2.00	0.66	0.71	0.00	0.00
58	Douro	0.90	-2.08	0.74	0.50	1.23	0.83	0.01	0.10	0.23	0.00
59	Barito	-2.74	-3.93	0.39	0.69	0.64	1.14	0.25	0.17	0.00	0.00
60	Gambia	14.48	12.21	1.02	0.86	1.70	1.42	0.56	0.57	0.00	0.00

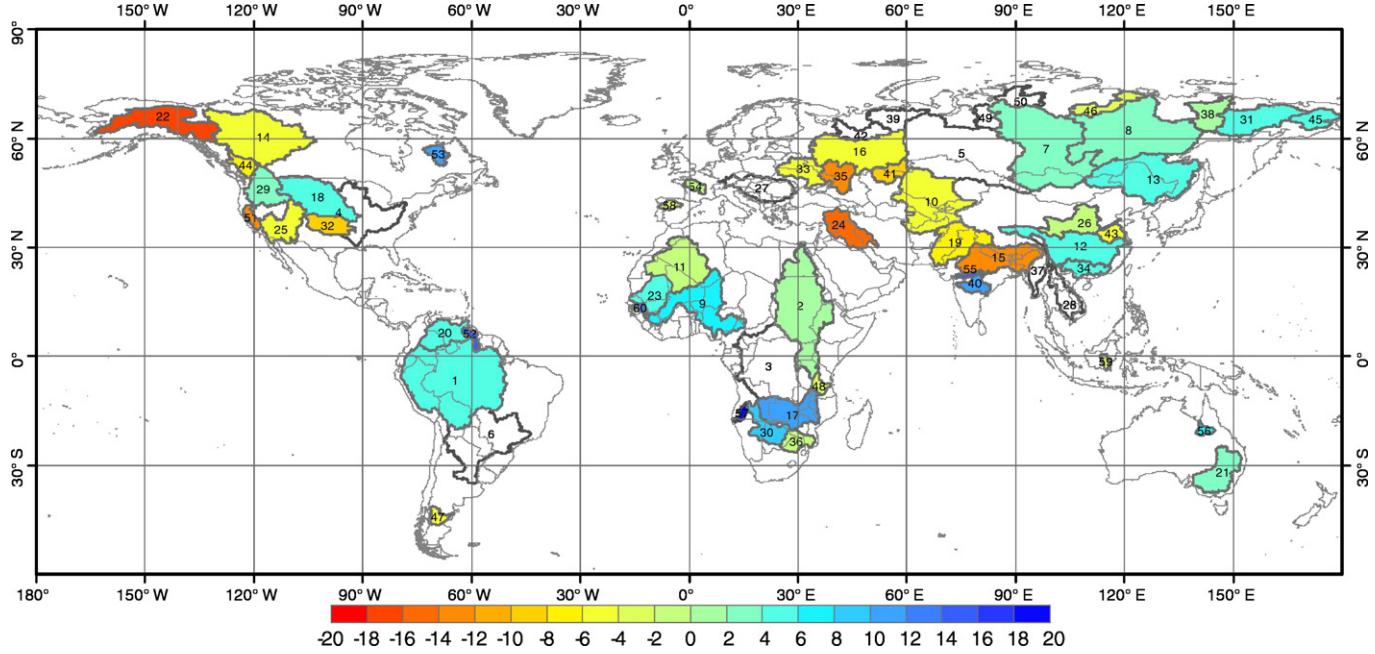


Fig. 12. Global distribution of basins with statistically significant ($p < 0.05$) trends (mm/a) in TWS derived from the BMA approach for the period Apr 2002–Mar 2015.

44.4%, 29.5%, 6.1% and 3.2% of the total variance, respectively. Fig. 15 shows spatial patterns (EOF modes 1–4) of TWS changes and associated amplitude time series (PCs 1–4) for the first four modes of the EOF analyses. Positive (negative) values in the modes (shown in blue (red) in the left panel of Fig. 15) multiplied by positive (negative) values of PCs (shown in blue (red) in the right panel of Fig. 15) yield positive deviations, whereas positive (negative) values of EOF modes 1–4 multiplied by negative (positive) PC values yield negative deviations.

PCs of mode 2 and mode 3 are associated with the dominant seasonal features. The amplitude time series of mode 2 (i.e., PC 2) is approximately a cosine wave with peaks in Apr–May (i.e., boreal spring) and

troughs in Oct–Nov (i.e., boreal autumn). Both mode 2 and PC 2 capture the 6-month phase shift between a spring maximum (positive anomalies) and autumn minimum (negative anomalies) across the western coast of North America (e.g., Fraser and Columbia), Amazon, South Africa (e.g., Zambezi), Central Asia (e.g., Aral), most regions of Russia and north Australia, versus the opposite pattern over the northern part of South America (e.g., Orinoco and Essequibo) and the southern part of South America (e.g., Chubut), Central Africa (e.g., Gambia), and South and Southeast Asia (e.g., Ganges, Narmada, Godavari, Irrawaddy and Mekong). These results are also consistent with the dominant seasonal signal derived by STL (Humphrey et al., 2016) that

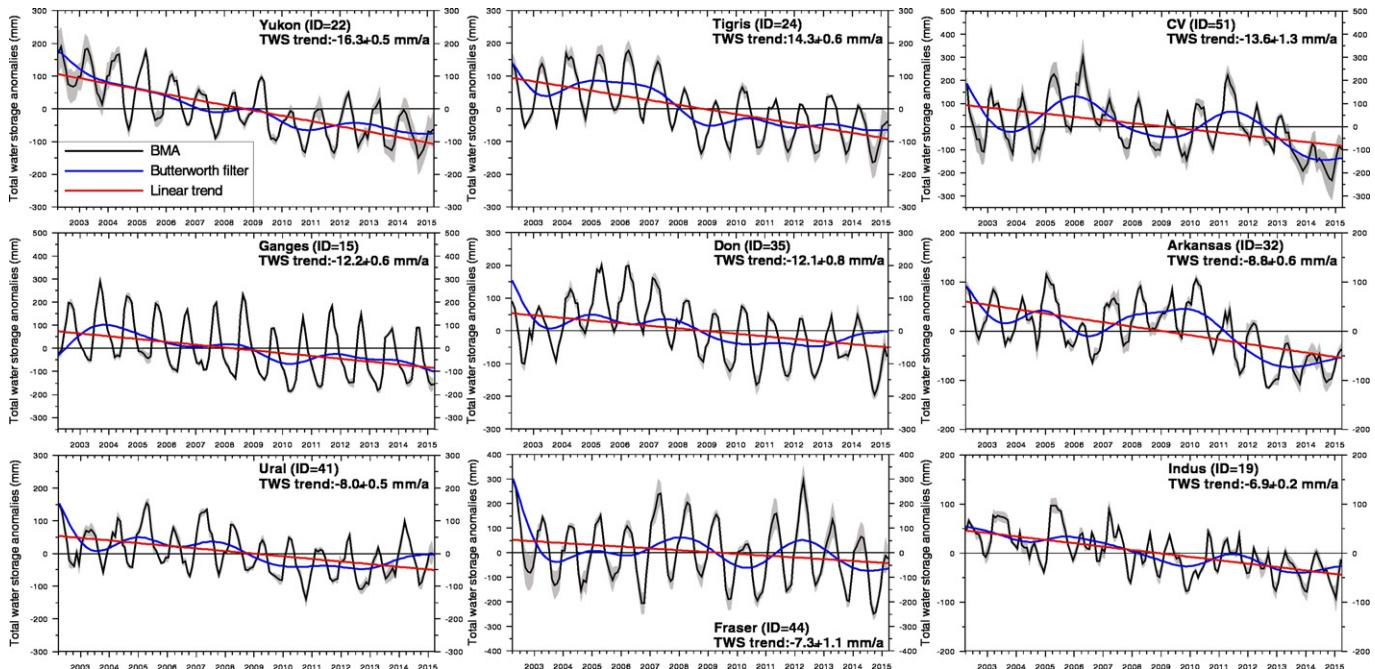


Fig. 13. TWS anomaly time series from BMA for a 13-year period (Apr 2002–Mar 2015, black line), and the TWS time series with seasonal and semiannual variations removed using the Butterworth filter (blue line) and its linear trend (red line) based on the filtered time series for nine mostly decreasing TWS basins. Gray areas show uncertainties in the BMA-based TWS anomaly time series. The uncertainties among different products were approximated by the standard deviation of TWS changes from the mascons, CSR gridded SH product, additive correction, and BMA.

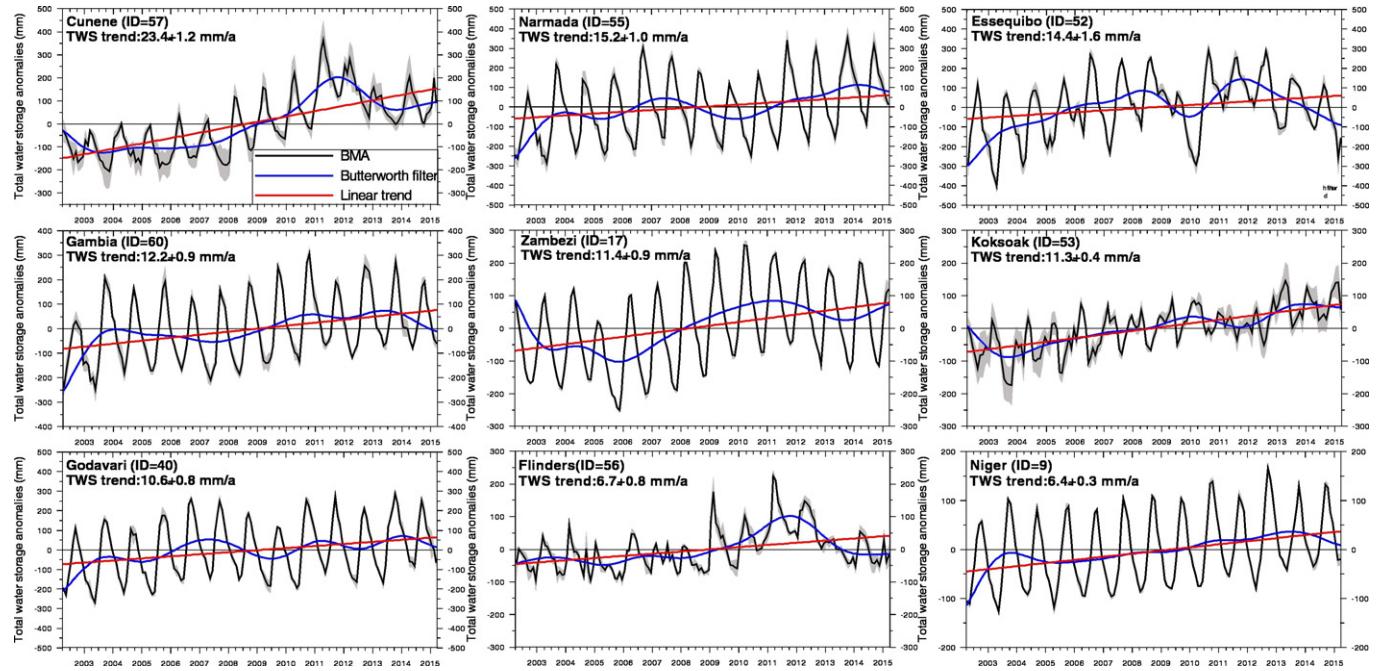


Fig. 14. TWS anomaly time series from BMA for a 13-year period (Apr 2002–Mar 2015, black line), and the TWS time series with seasonal and semiannual variations removed using the Butterworth filter (blue line) and its linear trend (red line) based on the filtered time series for nine mostly increasing TWS basins. Gray areas show uncertainties in the BMA-based TWS anomaly time series. The uncertainties among different products were approximated by the standard deviation of TWS changes from the mascons, CSR gridded product, additive correction, and BMA.

Amazon, central Arica, India, and northwest America showed marked seasonal variations in TWS. Humphrey et al. (2016) indicated that in the Northern Hemisphere, peaks in TWS generally occurred in spring for cold and temperate regions and in autumn for subtropical regions. The minimum TWS occurred in autumn for cold and temperate regions and in spring for sub-tropical regions. This is consistent with what mode 2 of the EOF analysis conveyed in our study.

PC 3 shows a seasonal cycle with peaks in Dec–Jan (boreal winter) and troughs in Jul–Aug (boreal summer). During boreal winter, the spatial pattern of mode 3 is strongly negative in the northern part of South America (e.g., Amazon, Orinoco, and Essequibo), South China (i.e., southeast Changjiang), southeast Ganges, and the Caspian Sea (which was not included in the study basins here). Negative values of mode 3 multiplied by positive values in the amplitude time series (PC 3) yield negative winter deviations in these regions. However, TWS experiences positive deviations during boreal winter in North America (e.g., Yukon and Koksoak), southern Amazon, Congo, south India (e.g., Narmada and Godavari), and Eurasia between 40°N and 70°N. The negative anomaly occurs during boreal summer. The combination of modes 2 and 3 describes a consecutive spatiotemporal variation in TWS.

PCs 1 and 4 indicate that both modes 1 and 4 are associated with long-term trends and interannual variability of TWS. Based on PC 1, the climatological shift occurred in 2008–2009. The spatial pattern of mode 1 shows positive TWS deviations in the west of North America (e.g., Yukon and Central Valley), Arkansas, Iceland, southeast Africa (e.g., Rufiji), Caspian Sea, Indus and west Ganges, North China Plain, and northwest Australia during the period 2002–2008, but negative TWS deviations during 2009–2015 in terms of the long-term trends (referring to Fig. 13). TWS in northeast North America (e.g., Koksoak), South Africa (e.g., Zambezi and Cunene), central India (e.g., Godavari) and Central-South China (e.g., middle Yangtze) experienced shifts from negative anomalies (i.e., positive values of mode 1 multiplied by negative amplitude time series) during 2002–2008 to positive anomalies (i.e., positive values multiplied by positive amplitude time series) during 2009–2015 (referring to Fig. 14). These results are generally consistent with the picture of interannual variability derived by STL (Humphrey et al., 2016) that performed analyses based on filtered

gridded SH without scaling. Joint use of the mascons and EOF in our study leads to a more detailed spatial distribution of interannual variability illustrated above.

Mode 4 exhibits the interannual variability with a correlation coefficient of -0.38 between the multivariate ENSO Index (MEI) and PC 4. The interannual shifts mainly occurred in 2004 and 2012 based on the amplitude time series. The spatial pattern of mode 4 is highlighted by negative anomalies in the southeast Yukon basin, central Amazon and south Parana basins in South America, upper Nile basin in Africa, west Ganges, Zhujiang River basin in China, south Mekong, and southeast Murray-Darling. These regions experienced positive TWS deviations from 2002 to mid-2004, negative deviations during 2005–2011, and positive deviations again during 2012–2015. However, regions with positive values of mode 4, i.e., central Canada, the Central Valley, northern and southeastern parts of South America (e.g., Orinoco and Essequibo), the Caspian Sea, south India (e.g., Godavari), the Huai basin in China, and north Australia, experienced negative deviations during 2002–2004, positive deviations during 2005–2011, and negative deviations during 2012–2015.

4. Conclusion

Uncertainties in various TWS change products from GRACE satellites hamper an elevated understanding of hydrological processes for a specific large basin and disaggregation of individual components from TWS changes. First, the TCH approach is applied to quantify relative uncertainties in TWS changes from GRACE observations, LSM and GHM simulations. Second, BMA and the WGHM model which accommodates impacts of human-induced water withdrawals on water storage, are jointly used to merge a range of TWS change products from GRACE. Trends in TWS from the merged product, the newly released mascons, and traditional processing (the additive correction and three gridded SH products) are examined in detail. Results show that the WGHM TWS output provides the lowest uncertainty among all products/outputs examined over 60 basins globally. The BMA approach trained by WGHM, PCR-GLOBWB, or Noah generally provides more consistent TWS changes than individual GRACE TWS change products with

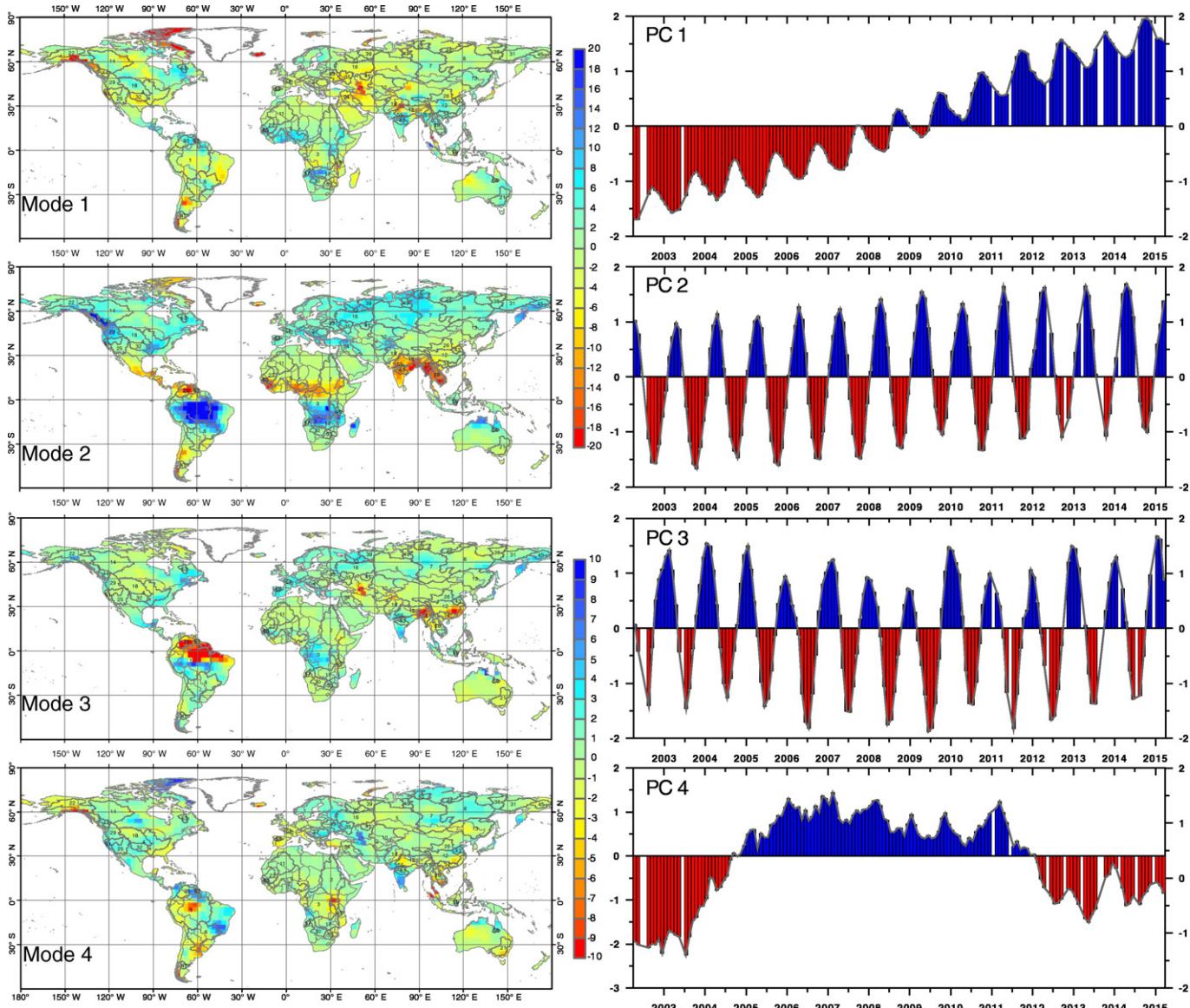


Fig. 15. Spatial patterns excluding Greenland and Antarctic for EOF modes 1–4 during the period Apr 2002–Mar 2015 (the left panel) and the normalized amplitude time series (principal components) for modes 1–4. Blank areas in PCs 1–4 denote months without TWS data.

reference to these training models during the validation period, which effectively reduces uncertainties by improving both amplitude and timing of TWS changes from GRACE. The mascons show the highest median of NSE (0.682) in all individual GRACE products with reference to the WGHM output for the 60 basins globally. Uncertainties in the BMA-based TWS changes exist if a poor reference model for training is selected. Therefore, it is necessary to ensure that the training data set has a relatively lower uncertainty than GRACE.

Trends in TWS globally for the period Apr 2002–Mar 2015 are examined. Increases in TWS mainly occur in the South and West Africa regions, South America, and North America. Significant depletion is found over major aquifers globally, including the Tigris, Central Valley, Ganges, upper Arkansas, and Indus basins attributed to intensive groundwater-fed irrigation. Basins subject to climate extremes, e.g., the Don and Ural basins also show large water depletion. In addition, glacier melting is the primary reason for TWS depletion over high-latitude (e.g., Yukon) and middle-latitude alpine (e.g., east Ganges and upper Brahmaputra in the Tibetan Plateau) basins. EOF analysis indicates marked seasonal and interannual variability in GRACE-observed TWS changes, with the interannual variability derived from EOF analysis

consistent with that derived at individual basins. This study underscores the need to incorporate a priori information from some hydrological model to reduce uncertainties in TWS changes from GRACE, and represents an advanced understanding of how TWS changes across a range of basins under varying climatic settings and levels of anthropogenic activities globally.

Acknowledgement

This study was supported mostly by the National Natural Science Foundation of China (Grant Nos. 91547210, 51579128 and 91437214), with some additional support from the key project of the Chinese Academy of Sciences (Grant No. KZZD-EW-13-04) and initial support from the foundation of the State Key Laboratory of Hydroscience and Engineering of Tsinghua University (Grant No. KZZD-EW-13-04). We appreciate Petra Döll (e-mail: p.doell@em.uni-frankfurt.de) with the Institute of Physical Geography, Goethe University Frankfurt, Germany, for her providing the WGHM output, and Yoshihide Wada (e-mail: wada@iiasa.ac.at) with the International Institute for Applied Systems Analysis, Austria, for his providing the PCR-GLOBWB output for this analysis. CSR

RL05 Level 2 data are available at the Center for Space Research at The University of Texas at Austin ([ftp://podaac.jpl.nasa.gov/allData/grace/L2/CSR/RL05/](http://podaac.jpl.nasa.gov/allData/grace/L2/CSR/RL05/)); gridded GRACE TWS anomaly data are available at the Jet Propulsion Laboratory, California Institute of Technology (<http://grace.jpl.nasa.gov/data/gracemonthlymassgridsland/>); mascons-based TWS anomaly data are available at the JPL website ([ftp://podaac.jpl.nasa.gov/allData/tellus/L3/mascon/RL05/JPL/](http://podaac.jpl.nasa.gov/allData/tellus/L3/mascon/RL05/JPL/)); GLDAS-1 output is available at the Goddard Earth Sciences Data and Information Services Center, NASA (<http://disc.sci.gsfc.nasa.gov/hydrology/data-holdings>). Comments provided by four anonymous reviewers, editors, and Dr. Xun Sun from East China Normal University are highly appreciated.

References

- Ahmed, M., Sultan, M., Wahr, J., Yan, E., 2014. The use of GRACE data to monitor natural and anthropogenic induced variations in water availability across Africa. *Earth Sci. Rev.* 136, 289–300.
- Alcamo, J., Doll, P., Henrichs, T., Kaspar, F., Lehner, B., Rosch, T., Siebert, S., 2003. Development and testing of the WaterGAP 2 global model of water use and availability. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques* 48, 317–337.
- Arendt, A., Luthcke, S., Gardner, A., O'Neil, S., Hill, D., Moholdt, G., Abdalati, W., 2013. Analysis of a GRACE global mascon solution for Gulf of Alaska glaciers. *J. Glaciol.* 59, 913–924.
- van Beek, L.P.H., Eikelboom, T., van Vliet, M.T.H., Bierkens, M.F.P., 2012. A physically based model of global freshwater surface temperature. *Water Resources Research* 48.
- Breña-Naranjo, J.A., Kendall, A.D., Hyndman, D.W., 2014. Improved methods for satellite-based groundwater storage estimates: a decade of monitoring the high plains aquifer from space and ground observations. *Geophys. Res. Lett.* 41, 6167–6173.
- Bulmer, M.G., 1979. *Principles of Statistics*. Dover, New York.
- Castle, S.L., Thomas, B.F., Reager, J.T., Rodell, M., Swenson, S.C., Famiglietti, J.S., 2014. Groundwater depletion during drought threatens future water security of the Colorado River Basin. *Geophysical Research Letters* 41, 2014GL061055.
- Chen, J.L., Tapley, B.D., Wilson, C.R., 2006. Alaskan mountain glacial melting observed by satellite gravimetry. *Earth Planet. Sci. Lett.* 248, 368–378.
- Chen, J.L., Wilson, C.R., Tapley, B.D., Blankenship, D.D., Ivins, E.R., 2007. Patagonia Icefield melting observed by Gravity Recovery and Climate Experiment (GRACE). *Geophysical Research Letters* 34 (n/a-n/a).
- 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.
- Chen, J.L., Li, J., Zhang, Z.Z., Ni, S.N., 2014. Long-term groundwater variations in Northwest India from satellite gravity measurements. *Glob. Planet. Chang.* 116, 130–138.
- Doll, P., Kaspar, F., Lehner, B., 2003. A global hydrological model for deriving water availability indicators: model tuning and validation. *J. Hydrol.* 270, 105–134.
- 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 Resour. Res.* 50, 5698–5720.
- Duan, Q.Y., Phillips, T.J., 2010. Bayesian estimation of local signal and noise in multimodel simulations of climate change. *Journal of Geophysical Research: Atmospheres* 115 (n/a-n/a).
- Famiglietti, J.S., Lo, M., Ho, S.L., Bethune, J., Anderson, K.J., Syed, T.H., Swenson, S.C., de Linage, C.R., Rodell, M., 2011. Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophys. Res. Lett.* 38, L03403.
- Farinotti, D., Longuevergne, L., Moholdt, G., Duethmann, D., Molg, T., Bolch, T., Vorogushyn, S., Guntner, A., 2015. Substantial glacier mass loss in the Tien Shan over the past 50 years. *Nature Geosci* 8, 716–722.
- Feng, W., Zhong, M., Lemoine, J.M., Biancale, R., Hsu, H.T., Xia, J., 2013. Evaluation of groundwater depletion in North China using the Gravity Recovery and Climate Experiment (GRACE) data and ground-based measurements. *Water Resour. Res.* 49, 2110–2118.
- Forootan, E., Rietbroek, R., Kusche, J., Sharifi, M.A., Awange, J.L., Schmidt, M., Omondi, P., Famiglietti, J., 2014. Separation of large scale water storage patterns over Iran using GRACE, altimetry and hydrological data. *Remote Sens. Environ.* 140, 580–595.
- Frappart, F., Papa, F., Silva, J., Ramillien, G., Prigent, C., Seyler, F., Calmant, S., 2012. Surface freshwater storage and dynamics in the Amazon basin during the 2005 exceptional drought. *Environ. Res. Lett.* 7, 044010.
- Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian Model Averaging: A Tutorial (With Comments by M. Clyde, David Draper and E. I. George, and a Rejoinder by the Authors). pp. 382–417.
- Horwath, M., Dietrich, R., 2009. Signal and error in mass change inferences from GRACE: the case of Antarctica. *Geophys. J. Int.* 177, 849–864.
- Huang, Z., Pan, Y., Gong, H., Yeh, P.J.F., Li, X., Zhou, D., Zhao, W., 2015. Subregional-scale groundwater depletion detected by GRACE for both shallow and deep aquifers in North China plain. *Geophys. Res. Lett.* 42, 1791–1799.
- Humphrey, V., Gudmundsson, L., Seneviratne, S.I., 2016. Assessing global water storage variability from GRACE: trends, seasonal cycle, subseasonal anomalies and extremes. *Surv. Geophys.* 37, 357–395.
- Immerzeel, W.W., van Beek, L.P.H., Bierkens, M.F.P., 2010. Climate change will affect the Asian water towers. *Science (New York, N.Y.)* 328, 1382–1385.
- Jacob, T., Wahr, J., Pfeffer, W.T., Swenson, S., 2012. Recent contributions of glaciers and ice caps to sea level rise. *Nature* 482, 514–518.
- Joodaki, G., Wahr, J., Swenson, S., 2014. Estimating the human contribution to groundwater depletion in the Middle East, from GRACE data, land surface models, and well observations. *Water Resour. Res.* 50, 2679–2692.
- Klees, R., Zapreeva, E.A., Winsemius, H.C., Savenije, H.H.G., 2007. The bias in GRACE estimates of continental water storage variations. *Hydrol. Earth Syst. Sci.* 11, 1227–1241.
- Landerer, F.W., Swenson, S.C., 2012. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour. Res.* 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.
- Lewis, S.L., Brando, P.M., Phillips, O.L., van der Heijden, G.M.F., Nepstad, D., 2011. The 2010 Amazon drought. *Science* 331, 554.
- Long, D., Scanlon, B.R., Longuevergne, L., Sun, A.-Y., Fernando, D.N., Save, H., 2013. GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas. *Geophys. Res. Lett.* 40, 3395–3401.
- Long, D., Longuevergne, L., Scanlon, B.R., 2014a. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resour. Res.* 50, 1131–1151.
- Long, D., Shen, Y.J., Sun, A.Y., Hong, Y., Longuevergne, L., Yang, Y.T., Li, B., Chen, L., 2014b. Drought and flood monitoring over a large karst plateau in Southwest China using extended GRACE data. *Remote Sensing of Environment* 145–160.
- Long, D., Longuevergne, L., Scanlon, B.R., 2015a. Global analysis of approaches for deriving total water storage changes from GRACE satellites. *Water Resour. Res.* 51, 2574–2594.
- Long, D., Yang, Y.T., Wada, Y., Hong, Y., Liang, W., Chen, Y.N., Yong, B., Hou, A.Z., Wei, J.F., Chen, L., 2015b. Deriving scaling factors using a global hydrological model to restore GRACE total water storage changes for China's Yangtze River basin. *Remote Sens. Environ.* 168, 177–193.
- Long, D., Chen, X., Scanlon, B.R., Wada, Y., Hong, Y., Singh, V.P., Chen, Y., Wang, C., Han, Z., Yang, W., 2016. Have GRACE satellites overestimated groundwater depletion in the Northwest India Aquifer? *Scientific Reports* 6, 24398.
- Longuevergne, L., Scanlon, B.R., Wilson, C.R., 2010. GRACE hydrological estimates for small basins: evaluating processing approaches on the High Plains Aquifer, USA. *Water Resources Research* 46.
- Longuevergne, L., Wilson, C.R., Scanlon, B.R., Crétaux, J.F., 2013. GRACE water storage estimates for the Middle East and other regions with significant reservoir and lake storage. *Hydrol. Earth Syst. Sci.* 17, 4817–4830.
- Morton, D.C., Nagol, J., Carabajal, C.C., Rosette, J., Palace, M., Cook, B.D., Vermote, E.F., Harding, D.J., North, P.R.J., 2014. Amazon forests maintain consistent canopy structure and greenness during the dry season. *Nature* 506, 221–224.
- Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F.T., Flörke, M., Döll, P., 2014. Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrol. Earth Syst. Sci.* 18, 3511–3538.
- Premoli, A., Tavella, P., 1993. A revisited three-cornered hat method for estimating frequency standard instability. *IEEE Trans. Instrum. Meas.* 42, 7–13.
- Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* 133, 1155–1174.
- Ramillien, G., Frappart, F., Seoane, L., 2014. Application of the regional water mass variations from GRACE satellite gravimetry to large-scale water management in Africa. *Remote Sens.* 6, 7379.
- Reager, J.T., Thomas, B.F., Famiglietti, J.S., 2014. River basin flood potential inferred using GRACE gravity observations at several months lead time. *Nature Geosci* 7, 588–592.
- Reager, J.T., Gardner, A.S., Famiglietti, J.S., Wiese, D.N., Eicker, A., Lo, M.-H., 2016. A decade of sea level rise slowed by climate-driven hydrology. *Science* 351, 699–703.
- Richey, A.S., Thomas, B.F., Lo, M.-H., Famiglietti, J.S., Swenson, S., Rodell, M., 2015a. Uncertainty in global groundwater storage estimates in a Total Groundwater Stress framework. *Water Resour. Res.* 51, 5198–5216.
- Richey, A.S., Thomas, B.F., Lo, M.-H., Reager, J.T., Famiglietti, J.S., Voss, K., Swenson, S., Rodell, M., 2015b. Quantifying renewable groundwater stress with GRACE. *Water Resour. Res.* 51, 5217–5238.
- 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.
- Rodell, M., Velicogna, I., Famiglietti, J.S., 2009. Satellite-based estimates of groundwater depletion in India. *Nature* 460, 999–1002.
- Scanlon, B.R., Longuevergne, L., Long, D., 2012. Ground referencing GRACE satellite estimates of groundwater storage changes in the California Central Valley, USA. *Water Resources Research* 48.
- Stoffelen, A., 1998. Toward the true near-surface wind speed: error modeling and calibration using triple collocation. *Journal of Geophysical Research: Oceans* 103, 7755–7766.
- Strassberg, G., Scanlon, B.R., Chambers, D., 2009. Evaluation of groundwater storage monitoring with the GRACE satellite: case study of the High Plains aquifer, central United States. *Water Resources Research* 45.
- Swenson, S., Wahr, J., 2007. Multi-sensor analysis of water storage variations of the Caspian Sea. *Geophys. Res. Lett.* 34.
- Tapley, B.D., Bettadpur, S., Ries, J.C., Thompson, P.F., Watkins, M.M., 2004. GRACE measurements of mass variability in the Earth system. *Science* 305, 503–505.
- Tavella, P., Premoli, A., 1991. Characterization of frequency standard instability by estimation of their covariance matrix. *Proceedings of 23rd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, pp. 265–276.

- Tavella, P., Premoli, A., 1994. Estimating the instabilities of N-clocks by measuring differences of their readings. *Metrologia* 30, 479–486.
- 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, 2014GL059323.
- Trabucco, A., Zomer, R.J., 2009. In: C.G.f.I.A. Research (Ed.), Global Aridity Index (Global-aridity) and Global Potential Evapotranspiration (Global-PET) Geospatial Database Published online, available from the CGIAR-CSIRO GeoPortal at: <http://www.cgiar.org/>.
- Velicogna, I., Wahr, J., 2006a. Acceleration of Greenland ice mass loss in spring 2004. *Nature* 443, 329–331.
- Velicogna, I., Wahr, J., 2006b. Measurements of time-variable gravity show mass loss in Antarctica. *Science* 311, 1754–1756.
- Venegas, S.A., Mysak, L.A., Straub, D.N., 1997. Atmosphere-ocean coupled variability in the South Atlantic. *J. Clim.* 10, 2904–2920.
- Voss, K.A., Famiglietti, J.S., Lo, M., de Linage, C., Rodell, M., Swenson, S.C., 2013. Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region. *Water Resour. Res.* 49, 904–914.
- Wada, Y., van Beek, L.P.H., Bierkens, M.F.P., 2011. Modelling global water stress of the recent past: on the relative importance of trends in water demand and climate variability. *Hydrol. Earth Syst. Sci.* 15, 3785–3808.
- Wada, Y., Wisser, D., Bierkens, M.F.P., 2014. Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. *Earth System Dynamics* 5, 15–40.
- Wang, S., Pan, M., Mu, Q., Shi, X., Mao, J., Brümmer, C., Jassal, R.S., Krishnan, P., Li, J., Black, T.A., 2015. Comparing evapotranspiration from eddy covariance measurements, water budgets, remote sensing, and land surface models over Canada. *J. Hydrometeorol.* 16, 1540–1560.
- Watkins, M.M., Wiese, D.N., Yuan, D.-N., Boening, C., Landerer, F.W., 2015. Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *Journal of Geophysical Research: Solid Earth* 120, 2648–2671.
- Yang, K., Wu, H., Qin, J., Lin, C., Tang, W., Chen, Y., 2014. Recent climate changes over the Tibetan Plateau and their impacts on energy and water cycle: a review. *Glob. Planet. Chang.* 112, 79–91.