Assessment of Three Common Methods for Estimating Terrestrial Water Storage Change with Three Reanalysis Datasets

SHANSHAN DENG, SUXIA LIU, AND XINGGUO MO

Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, and University of Chinese Academy of Sciences, College of Resources and Environment, Sino-Danish College, Beijing, China

(Manuscript received 7 October 2018, in final form 27 August 2019)

ABSTRACT

Terrestrial water storage change (TWSC) plays a crucial role in the hydrological cycle and climate system. To date, methods including 1) the terrestrial water balance method (PER), 2) the combined atmospheric and terrestrial water balance method (AT), and 3) the summation method (SS) have been developed to estimate TWSC, but the accuracy of these methods has not been systematically compared. This paper compares the spatial and temporal differences of the TWSC estimates by the three methods comprehensively with the GRACE data during the 2002–13 period. To avoid the impact of different inputs in the comparison, three advanced reanalysis datasets are used, namely 1) the National Centers for Environmental Prediction (NCEP)–Department of Energy (DOE) Reanalysis II (NCEP R2), 2) the ECMWF interim reanalysis (ERA-Interim), and 3) the Japanese 55-Year Reanalysis (JRA-55). The results show that all estimates with PER and AT considerably overestimate the long-term mean on a regional scale because the data assimilation in the reanalysis opens the water budget. The difficulty of atmospheric observation and simulation in arid and polar tundra regions is the documented reason for the failure of the AT method to represent the TWSC phase over 30% of the region found in this study. Although the SS result exhibited the best overall agreement with GRACE, the amplitude of TWSC based on SS differed substantially from that of GRACE and the similarity coefficient of the global distribution between the SS-derived estimate and GRACE is still not high. More detailed considerations of groundwater and human activities, for example, irrigation and reservoir impoundments, can help SS to achieve a higher accuracy.

1. Introduction

Terrestrial water storage (TWS) includes all forms of water stored on or below land surfaces such as snow/ice, soil moisture, groundwater, surface water, lakes/reservoirs, canopy interception, and biomass water (Hirschi et al. 2007). It plays a key role in the terrestrial and global water cycle and its components are of great importance to the climate system (Seneviratne et al. 2004; Hirschi et al. 2006a; Hirschi et al. 2007; Syed et al. 2008). Terrestrial water storage change (TWSC), which is defined as the spatiotemporal evolution of TWS, is an important water balance component. Improved quantitative estimates of TWSC are important for the management of water resources, agriculture, and ecosystems (Yeh and Famiglietti 2008). Moreover, TWSC estimates support our understanding of global climate change, for instance, the impact of TWSC on sea levels (Pokhrel et al. 2012) and on Earth rotation parameters such as length of day and polar

motion (Chao and O'Connor 1988; Adhikari and Ivins 2016; Liu et al. 2018). Previous studies have shown that TWSC is instrumental for assessing extreme climate events, such as floods, droughts, and heatwaves (Andersen et al. 2005; Reager and Famiglietti 2009; Reager et al. 2015; Yi and Wen 2016; Sinha et al. 2017).

However, observation data of TWSC are extremely scarce. Prior to the Gravity Recovery and Climate Experiment (GRACE) satellite mission, which was launched in 2002, no systematic, spatiotemporally resolved, and regional- to continental-scale observations of TWSC were available (Swenson et al. 2006). TWSC estimates from GRACE have been compared with in situ hydrological observations in several areas, showing promising performance (Swenson et al. 2006; Strassberg et al. 2007). Because the uncertainty of GRACE-derived TWSC is smaller than the uncertainty of TWSC simulation approaches (Alkama et al. 2010), GRACE data have been widely used from basin to regional and continental scales (Feng et al. 2012; Li et al. 2016; Wang and Li 2016; Yang et al. 2017).

Corresponding author: Suxia Liu, liusx@igsrr.ac.cn

DOI: 10.1175/JCLI-D-18-0637.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

Given the limited temporal coverage of GRACE data (less than two decades), it is necessary to develop methods to estimate longer-term TWSC data. Generally, there are two ways to estimate TWSC. The water balance method estimates TWSC by determining the difference between precipitation and the sum of evapotranspiration and runoff (Tang et al. 2010) or by replacing the residual between precipitation and evapotranspiration by transforming the atmospheric water balance equation (Seneviratne et al. 2004; Hirschi et al. 2006a; Hirschi et al. 2006b; Mueller et al. 2011; Hirschi and Seneviratne 2017). Non-water balance methods assess the major components of TWSC directly. For example, Chen and Wilson (2005) used soil moisture and snow data from the National Centers for Environmental Prediction (NCEP) reanalysis and Hirschi et al. (2006a) used in situ soil moisture and snow cover depth observations to represent TWSC. Seneviratne et al. (2004) used soil moisture, snow cover, and groundwater observations to estimate TWSC. Alkama et al. (2010) represented the TWSC as the total change in model-simulated soil moisture, snow, vegetation interception, and stream water content. Reanalysis data are often used to estimate TWSC because of the length of the available record, global coverage, and the consistency in both hydrological and atmospheric data (Hirschi et al. 2006a; Hirschi et al. 2006b; Yeh and Famiglietti 2008; Mueller et al. 2011; Hirschi and Seneviratne 2017).

Although numerous past TWSC studies have used water balance and non-water balance methods with reanalysis data, few studies have systematically compared both approaches. Previous studies chose one method over another method based only on theoretical deduction without a comparison (e.g., Chen and Wilson 2005). Thus, in this study, we focus on and analyze the differences between commonly used methods, including the terrestrial water balance method (hereafter called PER), the atmospheric and terrestrial water balance method (hereafter called AT), and the summation method (hereafter called SS), to estimate TWSC. To analyze the impact of using different reanalysis data, this study employs three advanced reanalysis datasets, the NCEP–Department of Energy (DOE) Reanalysis II (NCEP R2), the ECMWF interim reanalysis (ERA-Interim, hereinafter ERA-I), and the Japanese 55-Year Reanalysis (JRA-55) datasets, to calculate the TWSC using the three methods.

This study follows the path of many previous studies (Alkama et al. 2010; Zhang et al. 2016; Hirschi and Seneviratne 2017; Zhang et al. 2017) and uses GRACE-derived TWSC as the validation data to evaluate the estimates. First, the similarity of the global TWSC distribution derived from each estimate and GRACE is analyzed. Second, the global spatial mean of the estimates at monthly to yearly time scales are compared with GRACE.

Third, the consistency of the regional TWSC between each estimate and GRACE is discussed by evaluating the globally distributed differences in the phase, long-term average, and amplitude. The possible reasons for the differences of the applied methods are discussed, and guidance is provided for choosing potential approaches to increase the accuracy of TWSC estimates.

2. Methods and data

a. Data

This study employs three reanalysis datasets, namely the NCEP R2 from NCEP, the ERA-I from the European Centre for Medium-Range Weather Forecasts, and the JRA-55 from the Japan Meteorological Agency (JMA). Precipitation, evapotranspiration, runoff, precipitable water, divergence of moisture flux, volumetric soil water layer, and snow water equivalent from these datasets are used to estimate TWSC. The GRACE Tellus LAND Grids dataset provided by NASA is used for comparison with the estimates. The Köppen–Geiger climate classification dataset is used for assessing the accuracy of the estimates.

All data in this study, except for the climate classification data, were interpolated to the NCEP T62 Gaussian grid (192×94 , $\sim 222 \text{ km} \times 222 \text{ km}$, the lowest spatial resolution of all data) at monthly temporal scale. Limited by the time span of GRACE (2002–16) and JRA-55 (1979–2014), the study period ranges from May 2002 to December 2014.

1) NCEP R2 DATASET

The NCEP R2 (<https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html>) is an improved version of the NCEP R1, which is a state-of-the-art analysis/forecast system including assimilation of atmospheric, ocean, and land surface observations (see Table 1). Compared with NCEP R1, the NCEP R2 has corrected human errors, has higher resolution, and provides a more accurate picture of soil wetness, near-surface temperature, hydrological budget, winter precipitation of polar regions, hydrological and energy budget in polar regions, and snow cover (Kanamitsu et al. 2002).

Since air temperature data are not provided by NCEP R2, these data are derived from NCEP R1 and the sublimation data are from the NOAA-CIRES Twentieth Century Reanalysis version 2c, which is available at https://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2c.html. The wind levels are set to be the same as the specific humidity to calculate the divergence of the atmospheric moisture flux.

2) ERA-I DATASET

The ERA-I dataset is one of the latest global atmospheric reanalysis datasets from the ECMWF and was

TABLE 1. NCEP R2 data used in the study.

Variable	Temporal coverage	Spatial resolution	Levels
Water runoff	Jan 1979–Dec 2016	Global T62 Gaussian grid (192×94)	Surface
Precipitation rate	Jan 1979–Dec 2016	Global T62 Gaussian grid (192×94)	Surface
Latent heat net flux	Jan 1979–Dec 2016	Global T62 Gaussian grid (192×94)	Surface
Precipitable water	Jan 1979–Dec 2016	$2.5^\circ \times 2.5^\circ$ global grid (144×73)	Surface
u wind/ v wind	Jan 1979–Dec 2016	$2.5^\circ \times 2.5^\circ$ global grid (144×73)	1000–30 hPa
Specific humidity	Jan 1948–May 2017	$2.5^\circ \times 2.5^\circ$ global grid (144×73)	1000–300 hPa
Pressure	Jan 1979–Dec 2016	$2.5^\circ \times 2.5^\circ$ global grid (144×73)	Surface
Soil moisture	Jan 1979–Dec 2016	Global T62 Gaussian grid (192×94)	0–200 cm
Water equivalent of snow depth	Jan 1979–Dec 2016	Global T62 Gaussian grid (192×94)	Surface

launched in 2011. The primary goal of ERA-I is to address several problems encountered in its former version (ERA-40), which includes the low accuracy in representing the hydrological cycle and the inconsistency in time of the reanalyzed fields. By correcting the incorrect data assimilation processes, the ERA-I improves the moist physical processes and provides a better hydrological cycle presentation (Dee et al. 2011).

This study uses data from ERA-I, which includes seven variables downloaded from <http://apps.ecmwf.int/datasets/> (as shown in Table 2) to calculate TWSC.

3) JRA-55 DATASET

JRA-55, which was launched in 2013, is the second Japanese global atmospheric reanalysis project conducted by the JMA. It is one of the latest and most advanced reanalysis datasets and features many improvements over its predecessor, the Japanese 25-Year Reanalysis (JRA-25). JRA-55 aims to provide a high-quality time-consistent dataset for studying climate change and multidecadal variability (Kobayashi et al. 2015). Compared with JRA-25, JRA-55 has shown significant improvements in temporal coverage and temperature bias in the lower stratosphere and the dry land surface problem in the Amazon basin (Ebata et al. 2011). (The data listed in Table 3 were downloaded from <https://rda.ucar.edu> and are used for estimating TWSC.)

4) GRACE TELLUS LAND GRIDS DATASET

The GRACE Tellus LAND Grids dataset has been updated monthly and is available at [https://grace.jpl.](https://grace.jpl.nasa.gov/data/get-data/monthly-mass-grids-land/)

[nasa.gov/data/get-data/monthly-mass-grids-land/](https://grace.jpl.nasa.gov/data/get-data/monthly-mass-grids-land/). The updates of the dataset include the removal of atmospheric pressure and mass contributions, as well as postprocessing (Swenson and Wahr 2006), such as destriping, and Gaussian filters to reduce the errors in GRACE observation. In addition, much energy has also been removed at small spatial scales. A provided land-grid scaling was applied to these data to restore the energy removed by sampling and postprocessing (Landerer and Swenson 2012).

This study averages the solutions from three different processing centers, which are the Jet Propulsion Laboratory (JPL), Geoforschungs Zentrum Potsdam (GFZ), and the Center for Space Research (CSR) at the University of Texas, Austin. Each dataset represents the TWS mass deviation relative to the baseline average from January 2004 to December 2009. To compare the results with other estimates, the time average from May 2002 to December 2014 of each dataset is also subtracted in this study. Given that the baseline average from January 2004 to December 2009 has been subtracted in each dataset, the time average from May 2002 to December 2014 is equal to the time average of the other periods (May 2002–December 2003 and January 2010–December 2014). In addition, gaps in the GRACE time series were filled using gap-filling algorithms as presented in section 2b(1)(iv).

5) KÖPPEN-GEIGER CLIMATE CLASSIFICATION DATASET

To assess the accuracy of the estimates of TWSC, the global Köppen-Geiger climate classification dataset

TABLE 2. ERA-I data used in the study.

Variable	Temporal coverage	Spatial resolution	Levels
Runoff	Jan 1979–Mar 2017	$2.5^\circ \times 2.5^\circ$ global grid (144×73)	Surface
Evaporation	Jan 1979–Mar 2017	$2.5^\circ \times 2.5^\circ$ global grid (144×73)	Surface
Total precipitation	Jan 1979–Mar 2017	$2.5^\circ \times 2.5^\circ$ global grid (144×73)	Surface
Total column water vapor	Jan 1979–Mar 2017	$2.5^\circ \times 2.5^\circ$ global grid (144×73)	From the surface up to 0.1 hPa
Vertical integral of divergence of moisture flux	Jan 1979–Mar 2017	$2.5^\circ \times 2.5^\circ$ global grid (144×73)	From the surface up to 0.1 hPa
Snow depth	Jan 1979–Mar 2017	$2.5^\circ \times 2.5^\circ$ global grid (144×73)	Surface
Volumetric soil water layer	Jan 1979–Mar 2017	$2.5^\circ \times 2.5^\circ$ global grid (144×73)	0–289 cm

TABLE 3. JRA-55 data used in the study.

Variable	Temporal coverage	Spatial resolution	Levels
Water runoff	Jan 1958–Dec 2016	$1.25^\circ \times 1.25^\circ$ global grid (288×145)	Surface
Evapotranspiration	Jan 1958–Dec 2016	$1.25^\circ \times 1.25^\circ$ global grid (288×145)	Surface
Total precipitation	Jan 1958–Dec 2016	$1.25^\circ \times 1.25^\circ$ global grid (288×145)	Surface
Precipitable water	Jan 1958–Dec 2016	$1.25^\circ \times 1.25^\circ$ global grid (288×145)	1000–1 hPa
Zonal/meridional water vapor flux	Jan 1958–Dec 2016	$1.25^\circ \times 1.25^\circ$ global grid (288×145)	1000–1 hPa
Snow depth	Jan 1958–Dec 2016	$1.25^\circ \times 1.25^\circ$ global grid (288×145)	Surface
Mass concentration of condensed water in soil	Jan 1958–Dec 2016	$1.25^\circ \times 1.25^\circ$ global grid (288×145)	Underground layers

(Kottek et al. 2006) is used in this study. It is based on datasets from the Climatic Research Unit (CRU) of the University of East Anglia and the Global Precipitation Climatology Centre (GPCC) at the German Weather Service. The new digital Köppen–Geiger global climate classification dataset for the period of 1950–2000 can be downloaded from <http://koeppen-geiger.vu-wien.ac.at/present.htm>.

b. Methods

This study uses three methods (PER, AT, SS) and three reanalysis datasets (NCEP R2, ERA-I, JRA-55) to calculate TWSC. Thus there are nine sets of estimates based on the combinations of methods and inputs: NCEP/PER, NCEP/AT, NCEP/SS, ERA/PER, ERA/AT, ERA/SS, JRA/PER, JRA/AT, and JRA/SS. In addition, for each method, the average of three results is calculated to show the general performance of each method. To assess the accuracy of these estimates, six assessment indices in space and time are used to compare the results with the estimated TWSC, including the similarity coefficient of spatial distribution, the Nash–Sutcliffe efficiency, correlation coefficient, lag time, difference in mean value, and fractional difference of amplitude.

1) TWSC CALCULATION METHODS

(i) *Terrestrial water balance method (PER)*. In any given region, the TWSC can be estimated by the water balance equation:

$$\frac{\partial S(t)}{\partial t} = P(t) - E(t) - R(t), \quad (1)$$

where $S(t)$ represents the TWS in time t , $P(t)$ is the total precipitation, $E(t)$ is the evapotranspiration, and $R(t)$ is the runoff in this region. The units of all variables are millimeters per month (mm month^{-1}). The evapotranspiration data are not provided in the ERA-I and NCEP R2 reanalysis datasets. Therefore, for ERA-I, we use the evaporation data as a substitute for evapotranspiration. For NCEP R2, the evapotranspiration is estimated by Eq. (2) (e.g., Su and Feng 2015):

$$E = \frac{Q_E}{L_e} + S_{ub}, \quad (2)$$

$$L_e \approx 2.501 \times 10^6 - 2370 T, \quad (3)$$

where L_e (J kg^{-1}) is the latent heat of evaporation, which can be estimated from Eq. (3) (Jacobson 2005, 38–40), T ($^\circ\text{C}$) is the surface air temperature, Q_E (W m^{-2}) is the latent heat flux, and S_{ub} (mm) is the sublimation.

(ii) *Combined atmospheric and terrestrial water balance method (AT)*. The combined terrestrial and atmospheric water balance equation is defined in the following equation (e.g., Hirschi et al. 2006a):

$$\frac{\partial S(t)}{\partial t} = -\frac{\partial W(t)}{\partial t} - \nabla Q(t) - R(t), \quad (4)$$

where $\partial W/\partial t$ (mm) is the change in the column storage of water vapor and is equal to the changes in precipitable water in a region and time t , ∇Q (mm) is the water vapor flux convergence at time t , and the meaning of $R(t)$ is the same as in Eq. (1). In NCEP R2, the water vapor flux convergence needs to be calculated with specific humidity, wind, and pressure as shown in Eqs. (5) and (6):

$$Q = \frac{1}{g} \int_{Ap_s}^{Ap_t} Vq dAp, \quad (5)$$

$$\nabla Q = \frac{\partial Q_v}{\partial y} + \frac{\partial Q_u}{\partial x} - \frac{Q_v}{R_e} \times \tan(\phi), \quad (6)$$

where Q ($\text{kg m}^{-1} \text{s}^{-1}$) is the water vapor flux, V (m s^{-1}) is the zonal or meridional wind, q (kg kg^{-1}) is the specific humidity, Ap (hPa) is the atmospheric pressure, Ap_t is the top level of the atmospheric pressure, and Ap_s is the pressure at the ground surface. According to Eq. (5), the zonal (meridional) water vapor flux Q_u (Q_v) is known. Then Eq. (6) can use Q_u and Q_v to estimate the water vapor flux convergence with a central difference algorithm, which is similar to the wind divergence (Bluestein 1992, 113–114). In Eq. (6), ∂x (∂y) represents the zonal (meridional) distance per unit area, R_e (m) is Earth's radius, and ϕ is the latitude.

(iii) *The summation method (SS)*. The TWSC also can be expressed as the sum of all changes in water storage from the ground surface to a certain depth defined by the reanalysis model; this is approximately equal to changes in soil and snow water and ignores the groundwater change:

$$\frac{\partial S(t)}{\partial t} = \frac{\partial [S_n(t) + S_o(t)]}{\partial t}, \quad (7)$$

where $S_o(t)$ (mm) is the storage of soil water and $S_n(t)$ (mm) is the snow water equivalent in a region and time t . For the snow water equivalent, the snow depth, whose units are meters of water equivalent, is used in the three reanalysis datasets. For the storage of soil water, the (volumetric) soil moisture (fraction) from NCEP R2, the volumetric soil water (fraction) from ERA-I, and the mass concentration of condensed water in the soil (kg m^{-3}) from JRA-55 are used. These data are multiplied by the thickness of each soil layer defined in the reanalysis to obtain the soil water storage per unit of depth. Note that the data from JRA-55 have to be divided first by the density of water to convert the data to soil volume water content. The total depth of the soil defined by each reanalysis dataset is shown in Tables 1–3 (the level of soil water data).

(iv) *Converting mass anomalies to change rates*. The mass anomaly data of GRACE are converted into mass change rates. In a previous study (Forootan et al. 2017), a central difference operator was used to transform the mass anomalies to the mass change rate. However, this may cause possible phase shifts produced by the central differentiation operator. Therefore, in this study, to avoid phase shifts, we do not use the central differentiation operator but compute the mass change rate of a particular month by determining the difference between the mass anomalies on the last day and the first day of each month. Considering that the mass anomaly data of GRACE are not equally spaced in time, this study uses the Lagrange interpolation method to compute the data on the last day and the first day:

$$\begin{aligned} Y(t) = & Y(t_{i-1}) \frac{(t-t_i)(t-t_{i+1})}{(t_{i-1}-t_i)(t_{i-1}-t_{i+1})} \\ & + Y(t_i) \frac{(t-t_{i-1})(t-t_{i+1})}{(t_i-t_{i-1})(t_i-t_{i+1})} \\ & + Y(t_{i+1}) \frac{(t-t_{i-1})(t-t_i)}{(t_{i+1}-t_{i-1})(t_{i+1}-t_i)}, \end{aligned} \quad (8)$$

where $Y(t)$ is the interpolation result in time t , and $Y(t_{i-1})$, $Y(t_i)$, and $Y(t_{i+1})$ are the three datasets of the original series that are closest to $Y(t)$ in time. The results

FIG. 1. The monthly global mean of TWSC from GRACE (solid line with circles) and its Lagrange interpolation result (dashed line with crosses) from May 2002 to December 2014.

of the Lagrange interpolation are shown in Fig. 1. It is evident that the curve of the interpolated data and the curve of the original GRACE data are in high agreement.

2) ASSESSMENT INDICES

(i) *Similarity in spatial distribution*. The similarity coefficient is used to quantify the level of similarity between the spatial distribution of the estimates and the GRACE data; it is calculated as described by Lin (2006):

$$\begin{aligned} SC_t = & \frac{\sum_{y=1}^n \sum_{x=1}^m q_e(x, y, t) q_G(x, y, t)}{\sqrt{\sum_{y=1}^n \sum_{x=1}^m q_e(x, y, t)^2} \sqrt{\sum_{y=1}^n \sum_{x=1}^m q_G(x, y, t)^2}}, \\ t = & 1, 2, 3, \dots, l, \end{aligned} \quad (9)$$

where SC_t is the similarity coefficient at time t , l is the length of the time period, m is the number of zonal grid cells, and n is the number of meridional grid cells. The spatial distributions of the two fields tend to be similar when the value is close to 1, tend to be opposite when the value is close to -1 , and tend to be unrelated when the value is close to 0. Note that it is required that the data size, spatial extent, numbering method, and spatial resolution of the two fields are the same.

(ii) *Consistency in time*. To compare the global means of spatial time series data, the Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe 1970) is used to describe the accuracy of the estimates. It is defined as

$$\text{NSE} = 1 - \frac{\sum_{t=1}^l [q_e(t) - q_G(t)]^2}{\sum_{t=1}^l [q_G(t) - \mu_G]^2}, \quad (10)$$

where $q_e(t)$ and $q_G(t)$ are the time series of the estimates and GRACE, respectively; μ_G is the mean of the GRACE values and l is the length of the time period. The NSE ranges from $-\infty$ to 1. The closer the NSE is to 1, the more accurate the simulation. When the NSE is far less than zero, the estimated value is not accurate.

However, the NSE of the global mean is an overall measure of accuracy that is partly affected by the global differences in the fluctuation pattern, phase, mean value, and amplitude between the estimates and the GRACE data. To determine the underlying reasons for the differences in accuracy, the differences in the fluctuation pattern, phase, amplitude, and mean value are computed in each grid.

First, the zero-lag correlation coefficient is calculated to measure the similarity in fluctuation patterns between the estimates and the GRACE data. The smaller the areas with insignificant zero-lag correlation coefficients, the greater the agreement is between the estimates of TWSC and the GRACE data. Lag correlations from -6 months to 6 months are computed for areas with insignificant zero-lag correlation to analyze the phase change between the estimates and the GRACE data. Pearson's correlation coefficient is used to measure the linear correlation between $q_e(t)$ and $q_G(t)$ based on Eq. (11). Equation (12) is used to test the statistical significance level of the correlation between the two variables (e.g., Liu et al. 2009):

$$R = \frac{\text{cov}[q_G(t), q_e(t)]}{\sigma_G \sigma_e}, \quad (11)$$

$$\frac{-1 - Z\sqrt{l-2}}{l-1} \leq R \leq \frac{-1 + Z\sqrt{l-2}}{l-1}, \quad (12)$$

where R is the correlation coefficient, σ indicates the standard deviations, and cov indicates the covariance. When R is outside of the interval of Eq. (12), the correlation is statistically significant; $Z = 1.645, 1.96$, and 2.575 respectively correspond to the significance levels of $0.1, 0.05$, and 0.01 .

Second, the mean values of the estimates and the GRACE data are used to show the level of imbalance between the input and output of the water budget. Since the mean value of the TWSC monthly series in each grid is the mass change rate on a monthly scale, it reflects the regional water balance. The imbalance in the GRACE data can be regarded as the real changes in the water cycle

caused by climate change or anthropogenic influences because of its relatively high accuracy. Thus the difference between the imbalance of the estimates and the GRACE data should be the bias in the simulation. The similarity in the global distribution of the mean value between each estimate and GRACE is computed to determine if there is any bias in the simulation.

Third, the fractional difference in amplitude between the estimates and the GRACE data is computed to assess the bias in the amplitude of the estimates. The amplitude here is defined as the differences between the maximum and minimum value in the multiyear averaged annual cycle of TWSC. To remove the effect of the uneven spatial distribution of the TWSC amplitude orders, the fractional difference in the amplitude is calculated as the difference in the amplitude between the estimates and the GRACE data divided by the amplitude of the GRACE data.

(iii) *Uncertainty.* In this study, there are two approaches to calculate the uncertainties of the results. For each estimation method, the overall performance is represented as the performance of estimate averaged over three reanalysis datasets (hereafter called three-input average) and the uncertainty is represented as the differences between estimates of three different reanalysis datasets. For time-varying (space-varying) assessment index, the overall performance is represented as a time average (space average) and the uncertainty is represented as ± 1 standard deviation.

3. Results

a. The differences in the evolution of global mean on monthly, mean annual cycle, and yearly scales

Figures 2a–c show the averaged global mean time series over three reanalysis datasets estimated by the three methods and GRACE data on monthly, multiyear mean annual cycle, and yearly scales. It is found that there are large uncertainties in the results of PER and AT in 2014, which appear to be outliers (Figs. 2a,c). A closer check of the document and data of the reanalysis datasets indicates that this result is related to the change in the data processing approach of JRA-55 starting in 2014. The entire JRA-55 production was completed in 2013 and thereafter was continued on a near-real-time basis (Kobayashi et al. 2015). We thus treat the global mean estimates of JRA/PER and JRA/AT in 2014 as outliers. To remove the effect of the inconsistent input from JRA-55, all estimates in 2014 have been removed before the calculation of all indices used later in this study. It should be noted that this inconsistency in time of the estimates derived from JRA/PER and JRA/AT

FIG. 2. (a) Monthly time series, (b) multiyear mean annual cycle, and (c) yearly changes of global mean TWSC from three different methods and GRACE during the period of 2002–14, and (d) the Nash–Sutcliffe efficiency of the global mean TWSC for the estimates from three different methods and different inputs on different time scales during the period of 2002–13 (the outliers of 2014 are removed). Note that (b) shows the multiyear monthly averages of (a), and (c) is the yearly average of (a). The error bars in (a)–(c) show the uncertainty of each method, i.e., the change in the results for the three different reanalysis input. The solid lines in (a)–(c) are the three-input average. In *x*-axis labels of (d), M means monthly changes, A means the annual cycle, and Y means yearly changes.

only affects the others indices slightly compared to the effect on the yearly global mean changes.

As shown in Fig. 2a, there is a strong fluctuation in the monthly global mean TWSC and, generally, there is a high consistency between the estimates and the GRACE data. The NSE of the estimates shows that all the estimated monthly series match the GRACE data well; the NSE is 0.77 for AT, 0.81 for SS, and 0.62 for PER. This similar fluctuation pattern reflects the multiyear average annual cycle. As shown in Fig. 2b, the annual cycles are similar for the estimates and the GRACE data; the lowest values occur around June and the highest values around January. For the multiyear mean annual cycle, the NSE is 0.83 for

AT, 0.89 for SS, and 0.62 for PER. On a yearly time scale, the NSE is 0.41 for AT, 0.81 for SS, and 0.73 for PER. All results of three-input average show higher accuracy of the SS in the global mean time series over PER and AT.

However, we notice that the accuracy of the global mean TWSC estimates of each method varies for different inputs (Fig. 2d). For the PER method, JRA/PER results in a global mean TWSC with high accuracy on both the yearly and monthly time scales whereas NCEP/PER fails to provide good performance on both time scales and ERA/PER misestimates on a yearly time scale. For the AT method, although NCEP/AT and JRA/AT have relatively low accuracy for the global mean yearly

FIG. 3. (a) Monthly time series, (b) multiyear mean annual cycles, and (c) yearly changes in the similarity coefficients of global distribution between the estimates from the three methods and GRACE during the period of 2002–13, and (d) the comparison of three different methods with different inputs. Note that (b) shows the multiyear monthly averages of (a), and (c) is the yearly average of (a). The error bars in (a)–(c) show the uncertainty of each method, i.e., the change in of the results for the three different reanalysis input. The solid lines in (a)–(c) are the result between GRACE and the three-input averaged estimate of each method. The error bars in (d) show the uncertainty of the time-averaged similarity coefficient, represented as ± 1 standard deviation. In the x -axis labels of (d), M means monthly changes, A means the annual cycle, and Y means yearly changes.

change, ERA/AT has high accuracy on both time scales. For the SS method, the estimates are generally highly consistent with the GRACE data, except for ERA/SS on a yearly time scale.

b. The evolution of differences in spatial similarity on monthly, mean annual cycle, and yearly scales

Averaged over three reanalysis datasets, the similarity coefficient of the global distribution between the estimates and GRACE of each month is fluctuating over time (Fig. 3a). The global distributions of TWSC derived from SS and PER are quite similar to the GRACE data with mean similarity coefficients of 0.47 ± 0.07 and 0.42 ± 0.11 , followed by AT (0.22 ± 0.24). The fluctuation in the monthly series is also observed in the multiyear mean

monthly changes (as shown in Fig. 3b) with a cycle of about 5–7 months. The yearly variations are much more stable, as shown in Fig. 3c. Overall, the three-input average result of SS is slightly better than that of PER and much better than that of AT.

Figure 3d shows the similarity coefficients for all method/reanalysis dataset combinations during the study period. It is found that the accuracy of the three methods depends on the different inputs. For NCEP R2 and ERA-I, the ranking of the accuracy of the three methods is similar to that of the three-input average. NCEP/AT and ERA/AT have low accuracy for the global distribution. However, JRA-55 provides relatively high accuracy in the results of all three methods after removing the outliers present in 2014.

FIG. 4. (a)–(l) Global lag time between GRACE and estimates in the regions with insignificant zero-lag correlation between GRACE and estimates at the 0.05 level, (m) the global distribution of the regions with arid climates and polar tundra climate according to the Köppen–Geiger climate classification during the period of 1951–2000, and (n) the comparison of results in (a)–(l). Note that (a)–(c) are the results of the three-input average, and (d)–(l) are results of three different methods with three different reanalysis datasets.

c. The differences in the lag correlation between the monthly estimates and GRACE in each grid

The zero-lag correlation between the monthly estimates and the GRACE data in each grid cell is calculated to measure the similarity of temporal evolution on a

regional scale. The averaged correlation over the three reanalysis data is illustrated in Figs. 4a–c. About 93% of the areas is in significant zero-lag correlation with GRACE at the 0.05 significance level for PER and SS and 68% for AT. The averaged zero-lag correlation coefficients on the global scale for AT, PER, and SS are 0.32

± 0.30 , 0.54 ± 0.26 , and 0.56 ± 0.27 , respectively (all at the 0.05 significance level). It is found that AT performs worst for the temporal evolution at the regional scale.

The zero-lag correlation results of the three methods with three different reanalysis datasets are shown in Figs. 4d–l. For NCEP R2, the percentage of the significant zero-lag correlated area is 37% for AT, and 73% and 74% for PER and SS, respectively. For ERA-I, the significant areas for AT account for 65% of all areas whereas those for PER and SS are both 85%. The JRA-55 results are slightly better than those of ERA-I and the percentages are 70% for AT, 84% for PER, and 91% for SS. No matter which reanalysis datasets are used, the AT method performs worst for the temporal evolution on the regional scale. PER and SS have a similar better performance.

The nonzero lag correlation is further calculated for the areas with insignificant zero-lag correlations between the estimates and GRACE data. The lag times with the highest lag correlations between the estimates and GRACE are defined as the phase differences between the estimates and GRACE for these regions (Figs. 4a–l). It is shown that the AT results have the least agreement with the GRACE data in the phase with the largest lag time (3.38 ± 1.67 months for the average over three reanalysis datasets, 3.89 ± 1.59 months for NCEP R2, 3.36 ± 1.64 months for ERA-I, and 3.62 ± 1.52 months for JRA 55). Also, the SS performs best with the smallest lag time (2.78 ± 1.54 months for the average over three reanalysis datasets, 2.96 ± 1.71 months for NCEP R2, 2.86 ± 1.69 months for ERA-I, and 2.78 ± 1.58 months for JRA 55).

d. The differences in global distribution of the mean estimate of TWSC over the study period

Over a long period, based on the principle of water balance, the average of TWSC over the study period (2002–13) should equal to zero. The mean value averaged over the study period of the estimates and the GRACE data in each grid cell are shown in Figs. 5a–l; the results clearly show the issue of the water imbalanced over the globe. The differences of the global distributions of water-imbalanced issue are noticeably different among the three methods. The SS-derived estimates have the highest agreement with the GRACE data and the similarity coefficients of the global distribution are 0.37 for the average over three reanalysis datasets, 0.28 for NCEP R2, 0.21 for ERA-I, and 0.29 for JRA 55. The AT-derived estimates have the least agreement with GRACE and the similarity coefficients of the global distribution are 0.02 for the average over three reanalysis datasets, -0.01 for NCEP, -0.02 for ERA, and 0.09 for JRA. For the PER method, the result of the average over three reanalysis datasets is 0.12 but

the result of NCEP is only 0.03 whereas those of ERA and JRA are 0.19 and 0.17, respectively.

Over the 12 years from 2002 to 2013, similar to the GRACE data, the result of SS shows a slight decrease trend of the mean value of TWSC (water deficit) in most areas of the globe and a slight increase trend (water sufficient) in tropical regions of the Northern Hemisphere and some subtropical regions of the Southern Hemisphere such as southern Africa. However, the results of PER and AT show a different global distribution pattern of the decreasing/increasing trend of TWSC from the GRACE data. AT even shows an opposite decreasing or increasing pattern. As shown in Fig. 5n, the large number of outliers in the results of AT and PER may be the reasons why AT and PER do not produce a similar pattern to that of the GRACE data.

e. The differences in global distribution of amplitude differences

As shown in Fig. 6, the global distribution of the fractional differences in the amplitude varies between the GRACE data and the estimates from different reanalysis datasets. Almost every estimate has considerable differences in regional TWSC amplitude with GRACE.

The amplitudes of the areas with high amplitudes around 200 mm (the black areas in Fig. 6m), that is, central Africa, northern Asia, and the Amazon basin, are more likely to be underestimated in estimates (the red areas in Figs. 6a–l). In contrast, the amplitudes of the areas with low amplitudes near 0 mm (the white areas in Fig. 6m), that is, eastern Australia, the Sahara Desert, southern Africa, the Great Plains, the central Andes Mountains, the Paraná basin, and most areas of eastern Asia, are more likely to be overestimated in estimates (the blue areas in Figs. 6a–l), especially the estimates derived from AT method (the dark blue areas Figs. 6b, 6e, 6h, and 6k).

According to the boxplots of all results, the estimates derived from AT methods perform worst as the median of fractional difference being 0.29 for estimate averaged over three reanalysis datasets, 0.28 for NCEP R2, 0.81 for ERA-I, and 0.87 for JRA 55. The estimates averaged over three reanalysis datasets based on the PER and SS methods, NCEP/SS and ERA/SS, are better than other estimates with the mean fractional differences being 0.21, 0.16, 0.03, and -0.16 , respectively.

4. Discussion

a. Reason for the unstable performances of the water balance methods

In this study, we found that the time-averaged regional TWSC estimates based on PER and AT differed

FIG. 5. Globally distributed mean value (mm) of TWSC from (a)–(l) estimates and (m) GRACE, and (n) the boxplots of results in (a)–(m). Note that (a)–(c) are the results of the three-input average, and (d)–(l) are the results of three different methods and three different reanalysis datasets. The central red line in the boxes indicates the median, and the blue bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The outliers, which fall below the 25th percentile minus 1.5 times the interquartile range (IQR, being equal to the difference between 75th and 25th percentiles) or above the 75th percentile plus 1.5 times the IQR, are plotted individually using red plus signs (+).

substantially from those of GRACE (Fig. 5). Eicker et al. (2016) examined the long-term consistency between TWSC estimates based on PER method and different reanalysis datasets and GRACE-derived TWSC and also found a lack of agreement. Eicker et al. (2016)

concluded that the lack of agreement was attributed to the time-varying biases in the reanalysis data. Since the TWSC in this study is the mass change rate, the time-averaged TWSC in each grid also reflects the regional water balance conditions.

FIG. 6. (a)–(l) Global fractional difference between the amplitude of estimates and GRACE, (m) the global amplitude (mm) of GRACE-derived TWSC, and (n) the boxplots of the results in (a)–(l). Note that (a)–(c) are the results of the three-input average, and (d)–(l) are the results of three different methods and three different reanalysis datasets. The central line in the boxes indicates the median and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The outliers, which fall below 25th percentile minus 1.5 times the IQR or above the 75th percentile plus 1.5 times the IQR, are plotted individually using plus signs (+).

Numerous previous studies have pointed out the imbalance of the water budgets in reanalysis (Seneviratne et al. 2004; Hirschi et al. 2006a; Yeh and Famiglietti 2008; Zeng et al. 2008; Tang et al. 2010; Mueller et al. 2011; Hirschi and Seneviratne 2017). The artificial water sources or sinks caused by the data assimilation processes are the main reason for the imbalance in the water budget.

Lorenz and Kunstmann (2012) also demonstrated the limited performance of reanalysis in closing the combined terrestrial and atmospheric water budget due to the forcing of atmospheric observations. Generally, the variables of the water cycle are constrained indirectly by many observations and thus are affected by any relevant data assimilation processes (Dee et al. 2011). Therefore,

there is no doubt that estimates based on the water balance theory and reanalysis datasets will likely fail to describe the real water balance conditions at the regional scale.

Although imbalances in the water budget occur, the objective of data assimilation during reanalysis is to improve the accuracy of certain variables and different effects may occur due to different methods and goals. For example, the NCEP R2 corrects the infiltration processes by subtracting (adding) the difference between the observed precipitation and modeled precipitation from (to) soil moisture (Kanamitsu et al. 2002). This process improves the soil moisture data, thus only NCEP/SS produces highly accurate yearly global mean TWSC whereas NCEP/PER is less accurate (Fig. 2d). In addition, ERA-I exhibits an improvement in closing the global terrestrial and atmospheric water balance (Lorenz and Kunstmann 2012) and results in high accuracy for the yearly global mean TWSC for ERA/AT (Fig. 2d) but the regional water balance condition of ERA/AT has low accuracy (Fig. 5h).

In short, the water balance methods are affected by different data assimilation approaches and are very sensitive to the water balance conditions of the datasets. Therefore, caution should be used when water balance methods are employed to estimate TWSC.

b. Reason for low accuracy of the AT method for the global distribution

As shown in the results of Fig. 4, there is a considerable bias of the AT estimates in some regions, which reduces the spatial accuracy of AT. The study of Mueller et al. (2011) also demonstrated that there was a lack of agreement between the AT-derived TWSC estimates and GRACE-derived TWSC in some basins. The results of most of their regions matched the results of this study. Mueller et al. (2011) concluded that the reason that AT showed a lack of agreement with the GRACE data in some regions was the relatively high uncertainty in moisture divergence. Draper and Mills (2008) investigated the accuracy of the moisture divergence of the reanalysis data over the Murray–Darling basin and found that the moisture flux of this basin was too small to simulate and was easily overshadowed by simulation errors. In this study, the AT-derived estimate over the Murray–Darling basin did not agree with the GRACE data whereas the result of PER was in good agreement (Fig. 4). As shown in Fig. 4, this study also found that regions with a large bias in the phase of the AT-derived estimates were mostly arid climate and polar tundra climate regions. It can be assumed that the moisture flux bias has a more immediate impact on AT because the inactive moisture fluxes in those areas increase the difficulty in observing and estimating the relevant atmospheric data. The PER

and SS, which are indirectly influenced, are less sensitive to the moisture flux bias than AT, as shown in Fig. 4.

c. Methods to improve the SS method

Although the SS result exhibited the best overall agreement with GRACE, all three methods, including SS, have some disadvantages when using reanalysis to estimate TWSC. For example, the amplitude of TWSC based on SS differed substantially from that of GRACE as shown in Fig. 6 and the similarity coefficient of the global distribution between the SS-derived estimate and GRACE was only 0.47.

One reason for the difference between SS and GRACE is that the groundwater has not been taken into account in the reanalysis datasets we used. Previous studies have pointed out the importance of groundwater in TWS (Alkama et al. 2010; Frappart et al. 2013; Sun et al. 2015). Zhang et al. (2017) investigated the consistency between the estimated TWSC from models with or without groundwater and the TWSC from GRACE. They found a better agreement with GRACE when the groundwater was taken into consideration. Thus the lack of groundwater information could have contributed to the lack of agreement between the estimates and GRACE. Since soil depth and terrain are unevenly distributed globally, the proportion of groundwater in TWSC varies spatially. This study showed underestimations of the amplitude in regions with abundant groundwater resources as shown in the study of Gleeson et al. (2016) (e.g., the Amazon basin and the northern part of Eurasia), as shown in Fig. 6. The uneven distribution of the groundwater contributed partly to the uneven global distribution in the differences in the amplitude and the global distribution between the estimates and GRACE.

Another reason is the human influence on the hydrological cycle through irrigation, reservoir impoundment, and other human activities, which are not fully considered in the reanalysis dataset. This problem will alter the natural curves of hydrological variations and cause discrepancies between the estimates and GRACE. According to Tang et al. (2010), although using measured data in the assimilation processes of reanalysis is equivalent to taking the human influence into account, the changes in the land surface schemes caused by human activities are not updated at the same time. This inconsistency between the input data may have also affected the estimates of TWSC.

Although all methods used in this study have some disadvantages in using reanalysis datasets to estimate TWSC, the SS-derived estimates exhibit a good agreement with the GRACE data. As long as the bias correction in the amplitude and global distribution is applied, it is expected that the use of SS results in high accuracy of the long-term TWSC series.

5. Conclusions

To find a better approach to estimate long-term and globally distributed TWSC with reanalysis datasets, in this study, we used three methods, namely PER, AT, and SS, and three reanalysis datasets, namely NCEP R2, ERA-I, and JRA-55, to estimate TWSC. The results show that the global mean changes of the averaged estimates over three reanalysis datasets by SS, with Nash–Sutcliffe efficiency coefficient (NSE) being 0.89 for multiyear mean annual cycle and 0.81 on a yearly time scale, are more consistent with the GRACE data than PER (AT), with NSE being 0.62 (0.83) and 0.73 (0.41). However, PER and AT can obtain the best performance based on single dataset of JRA-55 and ERA-I respectively. The temporal evolution of the global spatial distribution of the estimates by SS approaches, with the similarity coefficient being 0.47 ± 0.07 , is better than PER (0.42 ± 0.11 , 0.54 ± 0.26) and much better than AT (0.22 ± 0.24 , 0.32 ± 0.30) for the average over the three reanalysis datasets. So do the results based on single dataset of NCEP R2 and ERA-I respectively while JRA-55 provides relatively high accuracy in the spatial distribution for all three methods. It is found that the considerable overestimation of long-term mean on regional scale for AT and PER may be one of the reasons why AT and PER are incapable of matching a pattern similar to that of the GRACE data. Different performances of different reanalysis datasets can be explained by the different focus of the updating used by the reanalysis datasets through data assimilation. Besides, the difficulty of atmospheric observation and simulation in arid and polar tundra regions is the documented reason for the failure of the AT method to represent the TWSC phase over 30% of the region found in this study. We found that the estimate based on SS was, in general, more consistent with the GRACE data than the water balance category methods (PER and AT); relatively high accuracy was obtained in space and time, especially for the three-input averaged SS estimate.

Therefore, among the three methods, SS has the most potential to estimate the long-term TWSC using reanalysis data. Although the SS method provided the best agreement with GRACE, it was found that there was some bias in the amplitude and global distribution of the estimate; this was partly attributed to the lack of groundwater information and human influence in the reanalysis datasets. It is suggested that a bias correction method should be used prior to the analysis.

Acknowledgments. This paper is financially supported by National Key Research and Development Program of China projects (Grants 2017YFA0603702, 2018YFE0106500), the National Program on Key Basic Research Project of

China (Grant 2012CB957802), and the Danida Fellowship Centre EOForChina project (12-31359-K). Many thanks to Prof. Dr. Peter Bauer-Gottwein, Department of Environmental Engineering, Technical University of Denmark, for English editing of the manuscript. The first author also thanks the support of China Scholarship Council.

REFERENCES

- Adhikari, S., and E. R. Ivins, 2016: Climate-driven polar motion: 2003–2015. *Sci. Adv.*, **2**, e1501693, <https://doi.org/10.1126/sciadv.1501693>.
- Alkama, R., and Coauthors, 2010: Global evaluation of the ISBA-TRIP continental hydrological system. Part I: Comparison to GRACE terrestrial water storage estimates and in situ river discharges. *J. Hydrometeorol.*, **11**, 583–600, <https://doi.org/10.1175/2010JHM1211.1>.
- Andersen, O. B., S. I. Seneviratne, J. Hinderer, and P. Viterbo, 2005: GRACE-derived terrestrial water storage depletion associated with the 2003 European heat wave. *Geophys. Res. Lett.*, **32**, L18405, <https://doi.org/10.1029/2005GL023574>.
- Bluestein, H. B., 1992: *Synoptic-Dynamic Meteorology in Mid-latitudes*, Vol. I, Oxford University Press, 594 pp.
- Chao, B. F., and W. P. O'Connor, 1988: Global surface-water-induced seasonal variations in the Earth's rotation and gravitational field. *Geophys. J. Int.*, **94**, 263–270, <https://doi.org/10.1111/j.1365-246X.1988.tb05900.x>.
- Chen, J. L., and C. R. Wilson, 2005: Hydrological excitations of polar motion, 1993–2002. *Geophys. J. Int.*, **160**, 833–839, <https://doi.org/10.1111/j.1365-246X.2005.02522.x>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Draper, C., and G. Mills, 2008: The atmospheric water balance over the semiarid Murray-Darling River basin. *J. Hydrometeorol.*, **9**, 521–534, <https://doi.org/10.1175/2007JHM889.1>.
- Ebita, A., and Coauthors, 2011: The Japanese 55-Year Reanalysis “JRA-55”: An interim report. *SOLA*, **7**, 149–152, <https://doi.org/10.2151/SOLA.2011-038>.
- Eicker, A., E. Forootan, A. Springer, L. Longuevergne, and J. Kusche, 2016: Does GRACE see the terrestrial water cycle “intensifying”? *J. Geophys. Res.*, **121**, 733–745, <https://doi.org/10.1002/2015JD023808>.
- Feng, W., J. M. Lemoine, M. Zhong, and T. T. Hsu, 2012: Terrestrial water storage changes in the Amazon basin measured by GRACE during 2002–2010. *Chin. J. Geophys.*, **55**, 814–821.
- Forootan, E., A. Safari, A. Mostafaie, M. Schumacher, M. Delavar, and J. L. Awange, 2017: Large-scale total water storage and water flux changes over the arid and semiarid parts of the Middle East from GRACE and reanalysis products. *Surv. Geophys.*, **38**, 591–615, <https://doi.org/10.1007/s10712-016-9403-1>.
- Frappart, F., G. Ramillien, and J. Ronchail, 2013: Changes in terrestrial water storage versus rainfall and discharges in the Amazon basin. *Int. J. Climatol.*, **33**, 3029–3046, <https://doi.org/10.1002/joc.3647>.
- Gleeson, T., K. M. Bepko, S. Jasechko, E. Luijendijk, and M. B. Cardenas, 2016: The global volume and distribution of modern groundwater. *Nat. Geosci.*, **9**, 161–167, <https://doi.org/10.1038/ngeo2590>.
- Hirschi, M., and S. I. Seneviratne, 2017: Basin-scale water-balance dataset (BSWB): An update. *Earth Syst. Sci. Data*, **9**, 251–258, <https://doi.org/10.5194/essd-9-251-2017>.
- , —, and C. Schar, 2006a: Seasonal variations in terrestrial water storage for major midlatitude river basins. *J. Hydrometeorol.*, **7**, 39–60, <https://doi.org/10.1175/JHM480.1>.

- , P. Viterbo, and S. I. Seneviratne, 2006b: Basin-scale water-balance estimates of terrestrial water storage variations from ECMWF operational forecast analysis. *Geophys. Res. Lett.*, **33**, L21401, <https://doi.org/10.1029/2006GL027659>.
- , S. I. Seneviratne, S. Hagemann, and C. Schar, 2007: Analysis of seasonal terrestrial water storage variations in regional climate simulations over Europe. *J. Geophys. Res.*, **112**, D22109, <https://doi.org/10.1029/2006JD008338>.
- Jacobson, M. Z., 2005: *Fundamentals of Atmospheric Modeling*. 2nd ed. Cambridge University Press, 813 pp.
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S. K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643, <https://doi.org/10.1175/BAMS-83-11-1631>.
- Kobayashi, S., and Coauthors, 2015: The JRA-55 Reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, <https://doi.org/10.2151/jmsj.2015-001>.
- Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel, 2006: World map of the Köppen-Geiger climate classification updated. *Meteor. Z.*, **15**, 259–263, <https://doi.org/10.1127/0941-2948/2006/0130>.
- Landerer, F. W., and S. C. Swenson, 2012: Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour. Res.*, **48**, W04531, <https://doi.org/10.1029/2011WR011453>.
- Li, J. H., S. S. Wang, and F. Q. Zhou, 2016: Time series analysis of long-term terrestrial water storage over Canada from GRACE satellites using principal component analysis. *Can. J. Remote Sens.*, **42**, 161–170, <https://doi.org/10.1080/07038992.2016.1166042>.
- Lin, S., 2006: Method of forecasting regional sandstorm process in spring on seasonal scale (in Chinese). *J. Desert Res.*, **26**, 478–483.
- Liu, S., X. Mo, W. Zhao, V. Naeimi, D. Dai, C. Shu, and L. Mao, 2009: Temporal variation of soil moisture over the Wuding River basin assessed with an eco-hydrological model, in-situ observations and remote sensing. *Hydrol. Earth Syst. Sci.*, **13**, 1375–1398, <https://doi.org/10.5194/hess-13-1375-2009>.
- , S. Deng, X. Mo, and H. Yan, 2018: Indexing the relationship between polar motion and water mass change in a giant river basin. *Sci. China Earth Sci.*, **61**, 1065–1077, <https://doi.org/10.1007/s11430-016-9211-2>.
- Lorenz, C., and H. Kunstmann, 2012: The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. *J. Hydrometeorol.*, **13**, 1397–1420, <https://doi.org/10.1175/JHM-D-11-088.1>.
- Mueller, B., M. Hirschi, and S. I. Seneviratne, 2011: New diagnostic estimates of variations in terrestrial water storage based on ERA-Interim data. *Hydrol. Processes*, **25**, 996–1008, <https://doi.org/10.1002/hyp.7652>.
- Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models. Part 1: A discussion of principles. *J. Hydrol.*, **10**, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).
- Pokhrel, Y. N., N. Hanasaki, P. J. F. Yeh, T. J. Yamada, S. Kanae, and T. Oki, 2012: Model estimates of sea-level change due to anthropogenic impacts on terrestrial water storage. *Nat. Geosci.*, **5**, 389–392, <https://doi.org/10.1038/ngeo1476>.
- Reager, J. T., and J. S. Famiglietti, 2009: Global terrestrial water storage capacity and flood potential using GRACE. *Geophys. Res. Lett.*, **36**, L23402, <https://doi.org/10.1029/2009GL040826>.
- , A. C. Thomas, E. A. Sproles, M. Rodell, H. K. Beaudoin, B. Li, and J. S. Famiglietti, 2015: Assimilation of GRACE terrestrial water storage observations into a land surface model for the assessment of regional flood potential. *Remote Sens.*, **7**, 14 663–14 679, <https://doi.org/10.3390/rs71114663>.
- Seneviratne, S. I., P. Viterbo, D. Lüthi, and C. Schär, 2004: Inferring changes in terrestrial water storage using ERA-40 reanalysis data: The Mississippi River basin. *J. Climate*, **17**, 2039–2057, [https://doi.org/10.1175/1520-0442\(2004\)017<2039:ICITWS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2039:ICITWS>2.0.CO;2).
- Sinha, D., T. H. Syed, J. S. Famiglietti, J. T. Reager, and R. C. Thomas, 2017: Characterizing drought in India using GRACE observations of terrestrial water storage deficit. *J. Hydrometeorol.*, **18**, 381–396, <https://doi.org/10.1175/JHM-D-16-0047.1>.
- Strassberg, G., B. R. Scanlon, and M. Rodell, 2007: Comparison of seasonal terrestrial water storage variations from GRACE with groundwater-level measurements from the High Plains Aquifer (USA). *Geophys. Res. Lett.*, **34**, L14402, <https://doi.org/10.1029/2007GL030139>.
- Su, T., and G. L. Feng, 2015: Spatial-temporal variation characteristics of global evaporation revealed by eight reanalyses. *Sci. China Earth Sci.*, **58**, 255–269, <https://doi.org/10.1007/s11430-014-4947-8>.
- Sun, A. Y., J. Chen, and J. Donges, 2015: Global terrestrial water storage connectivity revealed using complex climate network analyses. *Nonlinear Processes Geophys.*, **22**, 433–446, <https://doi.org/10.5194/npg-22-433-2015>.
- Swenson, S., and J. Wahr, 2006: Post-processing removal of correlated errors in GRACE data. *Geophys. Res. Lett.*, **33**, L08402, <https://doi.org/10.1029/2005GL025285>.
- , P. J. F. Yeh, J. Wahr, and J. Famiglietti, 2006: A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois. *Geophys. Res. Lett.*, **33**, L16401, <https://doi.org/10.1029/2006GL026962>.
- Syed, T. H., J. S. Famiglietti, M. Rodell, J. Chen, and C. R. Wilson, 2008: Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resour. Res.*, **44**, W02433, <https://doi.org/10.1029/2006WR005779>.
- Tang, Q. H., H. L. Gao, P. Yeh, T. Oki, F. G. Su, and D. P. Lettenmaier, 2010: Dynamics of terrestrial water storage change from satellite and surface observations and modeling. *J. Hydrometeorol.*, **11**, 156–170, <https://doi.org/10.1175/2009JHM1152.1>.
- Wang, S. S., and J. H. Li, 2016: Terrestrial water storage climatology for Canada from GRACE satellite observations in 2002–2014. *Can. J. Remote Sens.*, **42**, 190–202, <https://doi.org/10.1080/07038992.2016.1171132>.
- Yang, P., J. Xia, C. S. Zhan, Y. F. Qiao, and Y. L. Wang, 2017: Monitoring the spatio-temporal changes of terrestrial water storage using GRACE data in the Tarim River basin between 2002 and 2015. *Sci. Total Environ.*, **595**, 218–228, <https://doi.org/10.1016/j.scitotenv.2017.03.268>.
- Yeh, P. J. F., and J. S. Famiglietti, 2008: Regional terrestrial water storage change and evapotranspiration from terrestrial and atmospheric water balance computations. *J. Geophys. Res.*, **113**, D09108, <https://doi.org/10.1029/2007JD009045>.
- Yi, H., and L. X. Wen, 2016: Satellite gravity measurement monitoring terrestrial water storage change and drought in the continental United States. *Sci. Rep.*, **6**, 19909, <https://doi.org/10.1038/srep19909>.
- Zeng, N., J. H. Yoon, A. Mariotti, and S. Swenson, 2008: Variability of basin-scale terrestrial water storage from a PER water budget method: The Amazon and the Mississippi. *J. Climate*, **21**, 248–265, <https://doi.org/10.1175/2007JCLI1639.1>.
- Zhang, L. J., H. Dobslaw, C. Dahle, I. Sasgen, and M. Thomas, 2016: Validation of MPI-ESM decadal hindcast experiments with terrestrial water storage variations as observed by the GRACE satellite mission. *Meteor. Z.*, **25**, 685–694, <https://doi.org/10.1127/metz/2015/0596>.
- , —, T. Stacke, A. Guntner, R. Dill, and M. Thomas, 2017: Validation of terrestrial water storage variations as simulated by different global numerical models with GRACE satellite observations. *Hydrol. Earth Syst. Sci.*, **21**, 821–837, <https://doi.org/10.5194/hess-21-821-2017>.