



**HAL**  
open science

## Global analysis of approaches for deriving total water storage changes from GRACE satellites

Di Long, Laurent Longuevergne, Bridget R. Scanlon

► **To cite this version:**

Di Long, Laurent Longuevergne, Bridget R. Scanlon. Global analysis of approaches for deriving total water storage changes from GRACE satellites. *Water Resources Research*, 2015, 51 (4), pp.2574-2594. 10.1002/2014WR016853 . insu-01137851

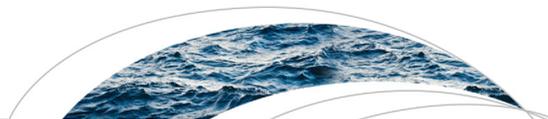
**HAL Id: insu-01137851**

**<https://insu.hal.science/insu-01137851>**

Submitted on 20 Apr 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



### RESEARCH ARTICLE

10.1002/2014WR016853

# Global analysis of approaches for deriving total water storage changes from GRACE satellites

Di Long<sup>1,2</sup>, Laurent Longuevergne<sup>3</sup>, and Bridget R. Scanlon<sup>1</sup>

#### Key Points:

- Model dependence of scaling factors was examined using six land surface models
- Temporal variability of scaling factors was examined using GLDAS-1 Noah
- Total water storage changes from three approaches were compared with WGHM output

#### Supporting Information:

- Supporting Information S1
- Table S1
- Table S2
- Table S3

#### Correspondence to:

D. Long,  
dlong@tsinghua.edu.cn

#### Citation:

Long, D., L. Longuevergne, and B. R. Scanlon (2015), Global analysis of approaches for deriving total water storage changes from GRACE satellites, *Water Resour. Res.*, 51, doi:10.1002/2014WR016853.

Received 31 DEC 2014

Accepted 16 MAR 2015

Accepted article online 23 MAR 2015

<sup>1</sup>Bureau of Economic Geology, Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA, <sup>2</sup>State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing, China, <sup>3</sup>Geosciences Rennes, UMR CNRS 6118, Université de Rennes 1, Rennes, France

**Abstract** Increasing interest in use of GRACE satellites and a variety of new products to monitor changes in total water storage (TWS) underscores the need to assess the reliability of output from different products. The objective of this study was to assess skills and uncertainties of different approaches for processing GRACE data to restore signal losses caused by spatial filtering based on analysis of  $1^\circ \times 1^\circ$  grid-scale data and in 60 river basins globally. Results indicate that scaling factors from six LSMs, including GLDAS-1 four models (Noah2.7, Mosaic, VIC, and CLM 2.0), CLM 4.0, and WGHM, are similar over most of humid, subhumid, and high-latitude regions but can differ by up to 100% over arid and semiarid basins and areas with intensive irrigation. Temporal variability in scaling factors is generally minor at the basin scale except in arid and semiarid regions, but can be appreciable at the  $1^\circ \times 1^\circ$  grid scale. Large differences in TWS anomalies from three processing approaches (scaling factor, additive, and multiplicative corrections) were found in arid and semiarid regions, areas with intensive irrigation, and relatively small basins (e.g.,  $\leq 200,000$  km<sup>2</sup>). Furthermore, TWS anomaly products from gridded data with CLM4.0 scaling factors and the additive correction approach more closely agree with WGHM output than the multiplicative correction approach. This comprehensive evaluation of GRACE processing approaches should provide valuable guidance on applicability of different processing approaches with different climate settings and varying levels of irrigation.

## 1. Introduction

Total water storage (TWS) change is defined as changes in water stored on the surface (e.g., lakes and reservoirs, rivers, and snow water equivalent), over the entire soil profile, and in groundwater. Traditional observations of TWS components (e.g., surface water storage (SWS) change, soil moisture storage (SMS) change, and groundwater storage (GWS) change) are made at point scales. Land surface models (LSMs) or hydrological models can simulate components of TWS changes; however, these models often lack processes to simulate some important components, e.g., deep SMS and GWS changes [Rodell *et al.*, 2004]. In addition, large variability in simulated water balance components among different models suggests a high level of uncertainty in some components, e.g., soil moisture during extremely dry conditions [Long *et al.*, 2013] and evapotranspiration during extremely wet conditions [Long *et al.*, 2014]. Furthermore, it is difficult to evaluate the reliability of TWS change and its components from LSMs due to the lack of observations at appropriate scales.

Since its launch in March 2002, Gravity Recovery and Climate Experiment (GRACE) satellite-observed TWS changes have been widely used to evaluate groundwater depletion in major aquifers [e.g., Feng *et al.*, 2013; Rodell *et al.*, 2009; Scanlon *et al.*, 2012b; Strassberg *et al.*, 2007], monitor droughts [e.g., Chen *et al.*, 2010b; Leblanc *et al.*, 2009; Long *et al.*, 2013; Tourian *et al.*, 2015], and floods [e.g., Chen *et al.*, 2010a; Reager and Famiglietti, 2009], and in data assimilation for improving LSMs [e.g., Eicker *et al.*, 2014; Houborg *et al.*, 2012; Lo *et al.*, 2010]. Because GRACE data are noisy, processing is required (truncating high degree and order of the spherical harmonic (SH) solutions and further filtering) to reduce noise; however, truncation and filtering also attenuate some of the signal [Eicker *et al.*, 2014]. Signal restoration for basin-scale studies generally includes bias and leakage corrections before applying GRACE data to resolve TWS changes, in most cases, in combination with synthetic data, e.g., global output of LSMs or based on hypotheses [Swenson and Wahr, 2002]. GRACE processing may not be a critical issue over large basins (e.g., area  $\geq 10^6$  km<sup>2</sup>) because bias and leakage offsets each other, but may be problematic for applying GRACE data to smaller basins less than

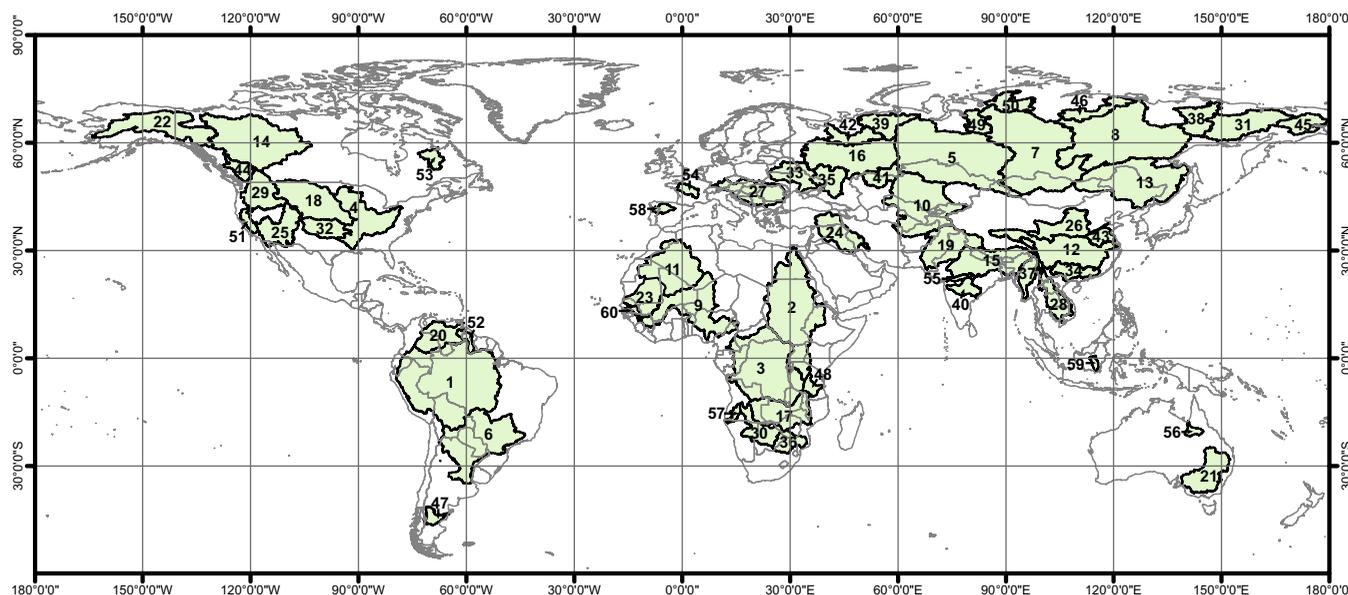
the GRACE footprint. The GRACE footprint is estimated to be  $\sim 200,000 \text{ km}^2$  based on the inherent resolution as a result of the satellite elevation ( $\sim 450 \times 450 \text{ km} = \sim 200,000 \text{ km}^2$ ). In addition, it is important to consider the impacts of different aspects of processing on estimated TWS changes, e.g., varying a priori LSMs and time periods considered, and climate and basin size to be applied.

There are generally three approaches to restore GRACE TWS change signals. (1) The first approach (the scaling factor approach) calculates a scaling factor for a grid cell or river basin by least squares fit between the filtered and unfiltered modeled TWS changes from LSMs. The GRACE TWS change is consequently restored by multiplying the filtered GRACE TWS change with the scaling factor. Traditional GRACE processing requires expertise in processing the standard GRACE SH geopotential coefficients. The gridded GRACE TWS anomaly products at a spatial resolution of  $1^\circ \times 1^\circ$  ( $111 \times 111 \text{ km} \sim 12,300 \text{ km}^2$  at the equator) developed using scaling factors derived from CLM4.0 are extremely valuable data sets because they do not require additional processing and have greatly popularized application of GRACE data by the hydrological community [Landerer and Swenson, 2012; Swenson and Wahr, 2006]. (2) The second approach (the additive correction approach) uses output from a LSM to add the signal loss between the unfiltered basin function used to extract TWS signal for a specific region and the filtered basin function using the same filters as applied to the GRACE data (bias correction), and removes the signal from the area surrounding the basin (leakage correction) [Klees et al., 2007; Longuevergne et al., 2010]. (3) The third approach (the multiplicative correction approach) requires calculation of a multiplicative factor that scales the amplitude of the filtered basin function to that of the unfiltered basin function and applies the multiplicative factor to the difference between filtered GRACE signal and the leakage [Longuevergne et al., 2010; Swenson and Wahr, 2007; Velicogna and Wahr, 2006]. Alternate methods for leakage corrections encompass the mascon-type approach [Tiwari et al., 2009], refocusing methods [Guo et al., 2010], or statistical modeling [Frootan et al., 2014].

The scaling factor approach is the most widely used and has been applied to individual basins or regions globally, and also to the more recent gridded GRACE TWS anomaly products [e.g., Cai et al., 2014; Pokhrel et al., 2013; Proulx et al., 2013; Sun, 2013]. However, the scaling factor may be sensitive to the LSM used because we assume that the LSM represents the actual mass changes but may neglect critical processes, such as irrigation and lateral water redistribution or storage compartments, such as surface water and groundwater [Longuevergne et al., 2013; Scanlon et al., 2012b]. Most applications use a single scaling factor over time; however, scaling factors may vary temporarily because of changes in spatial variability of TWS inside and outside a basin (typically different for seasonal variations and secular trends [Rodell et al., 2009]). In addition, performance of the scaling factor approach may vary in basins subject to different climatic forcing or anthropogenic conditions, particularly irrigation. The additive and multiplicative correction approaches require different assumptions, with the additive correction approach taking advantage of a priori information on spatial variability from LSMs to compute bias and leakage, and the multiplicative correction approach assuming uniform spatial distribution of TWS changes within a basin [Longuevergne et al., 2010]. However, differences in TWS anomalies from the three different approaches have not been systematically investigated. With the relatively recent development of the gridded product, it is important to compare TWS anomalies from this approach with output from more traditional processing approaches applied at the basin scale and also to consider varying climatic, anthropogenic, and basin size conditions.

There are other approaches of correcting bias and leakage in filtered GRACE TWS anomalies through an iterative process without using output from LSMs [e.g., Baur et al., 2009; Chen et al., 2013; Jacob et al., 2012; Wouters et al., 2008] or a regularization approach [Save et al., 2012]. These forward modeling approaches have been mostly used in cryosphere studies, and may also be suited for land hydrology studies. Global analyses of these approaches are still ongoing and will be discussed in the future.

The objectives of this study were therefore to: (1) examine variability in scaling factors with use of different LSMs (i.e., GLDAS-1 four LSMs, NCAR' CLM4, and the WaterGAP Global Hydrological Model (WGHM) at the gridded  $1^\circ \times 1^\circ$  scale and river basin scale); (2) quantify temporal variability in scaling factors at monthly scale using the GLDAS-1 Noah LSM at the  $1^\circ \times 1^\circ$  scale and river basin scale; and (3) compare TWS anomaly time series from the scaling factor, additive correction, and multiplicative correction approaches at the basin scale using the WGHM model TWS output as reference. This study builds on the work of Landerer and Swenson [2012] by expanding the number of LSMs considered from 46 in [Landerer and Swenson, 2012] to the 60 LSMs in this analysis. The impact of model dependence and temporal variability in scaling factors has not been thoroughly evaluated previously. Although different processing approaches have been compared in



**Figure 1.** Sixty river basins globally from the Simulated Topological Networks (STN-30p) examined in this study, containing the 46 basins selected in *Landerer and Swenson [2012]*. Basins ID and corresponding names are provided in Table 1.

some studies, most comparisons have been restricted to individual basins and none has been done at the global scale. Global analysis of these algorithms and related products should provide a holistic view of performance of different approaches, uncertainties in GRACE-derived TWS changes, and guidance on appropriate use of different processing approaches for hydrologic studies. As GRACE analysis continually evolves with the GRACE Follow-On Mission expected to launch in 2017, we need to understand impacts of processing and associated uncertainties to assess reliability of output.

## 2. Materials and Methods

### 2.1. GRACE Data and Processing

GRACE SH coefficients from the Center for Space Research (CSR, University of Texas at Austin) release 5 (RL05) were destriped [Swenson and Wahr, 2006], truncated at the maximum degree and order of 60, and filtered using a 300 km Gaussian filter and were subsequently converted into TWS anomalies for 60 river basins globally (Figure 1 and Table 1) for January 2003 through July 2013. Gridded GRACE TWS anomaly products at the spatial resolution of  $1^\circ \times 1^\circ$  from SH coefficients provided by CSR, the Jet Propulsion Laboratory (JPL), and German Research Center for Geoscience (GFZ) centers, and scaling factors derived from NCAR's CLM4.0 [Landerer and Swenson, 2012] were also obtained for comparison with TWS anomalies from the additive and multiplicative correction approaches. All these GRACE TWS anomalies were evaluated with TWS anomalies from the WGHM model (Version 2.2, STANDARD) (see below).

### 2.2. Land Surface Models

Here we adopted six LSMs to investigate the model dependence, including Noah2.7, Mosaic, VIC, and CLM2.0 in Global Land Data Assimilation System-1 (GLDAS-1) [Rodell et al., 2004], NCAR's CLM4.0 [Gent et al., 2011], and the WaterGAP WGHM2.2 model [Alcamo et al., 2003; Doll et al., 2003; Müller Schmied et al., 2014]. Temporal variability in scaling factors for 2003 through 2009 was investigated using the GLDAS-1 Noah model as an example.

There is no GWS component for LSMs in GLDAS-1. Both CLM4.0 and WGHM have a groundwater compartment accommodating natural recharge and discharge. In addition, WGHM2.2 simulates GWS changes resulting from human activities, e.g., irrigation. In this study, WGHM TWS changes are considered as a reference for comparison with the three approaches to restore GRACE signals because WGHM2.2 comprehensively considers water storage changes in surface water, soil layers, and aquifers, and accommodates

**Table 1.** Characteristics of Global River Basins Studied<sup>a</sup>

Basin ID	Name	Area (km <sup>2</sup> )	AI	Climate	TWS <sub>a</sub> (mm)	TWS <sub>sa</sub> (mm)	Total Error (mm)	STNR
1	Amazon	5,854,000	1.27	H	211.4	13.5	12.5	79.1
2	Nile	3,826,000	0.29	SA	41.9	6.0	8.5	17.3
3	Congo	3,699,000	0.89	H	50.8	16.0	13.0	26.6
4	Mississippi	3,203,000	0.68	H	62.6	9.9	8.0	36.2
5	Ob	3,026,000	0.74	H	64.0	9.5	6.5	27.0
6	Parana	2,664,000	0.71	H	75.6	6.6	16.6	22.7
7	Yenisei	2,579,000	0.82	H	52.7	13.8	7.1	20.5
8	Lena	2,442,000	0.78	H	41.4	13.5	9.5	27.0
9	Niger	2,240,000	0.32	SA	88.4	18.3	12.6	17.1
10	Aral	2,148,000	0.30	SA	46.3	3.4	5.2	28.2
11	Tamanraset	1,818,000	0.02	A	2.2	1.2	5.9	2.3
12	Changjiang	1,794,000	0.98	H	58.0	12.1	15.5	22.8
13	Amur	1,755,000	0.77	H	10.0	5.8	6.9	22.9
14	Mackenzie	1,695,000	0.76	H	57.2	7.3	11.9	19.1
15	Ganges	1,628,000	0.90	H	139.9	46.6	21.6	31.7
16	Volga	1,476,000	0.94	H	80.7	6.4	14.6	22.0
17	Zambezi	1,364,000	0.54	SH	147.0	33.9	20.0	22.1
18	Missouri	1,352,000	0.49	SA	48.2	8.4	8.4	25.5
19	Indus	1,143,000	0.37	SA	22.3	25.2	18.2	11.5
20	Orinoco	1,039,000	1.33	H	195.0	44.8	25.2	41.6
21	Murray	1,032,000	0.35	SA	19.9	4.6	13.5	12.5
22	Yukon	856,000	0.68	H	78.3	9.4	18.4	16.2
23	Senegal	847,000	0.16	A	43.0	11.7	11.3	9.7
24	Tigris	823,000	0.28	SA	91.6	4.7	12.3	21.8
25	Colorado	808,000	0.25	SA	37.3	3.0	9.4	9.6
26	Huanghe	795,000	0.50	SH	23.5	7.2	10.8	11.9
27	Danube	788,000	0.92	H	86.6	4.8	17.9	18.9
28	Mekong	759,000	1.02	H	214.4	19.9	28.5	30.5
29	Columbia	732,000	0.68	H	103.2	12.4	16.1	20.3
30	Okavango	701,000	0.29	SA	41.2	10.4	20.9	4.6
31	Kolyma	666,000	0.75	H	48.2	12.1	15.1	16.8
32	Arkansas	544,000	0.54	SH	46.8	3.6	15.1	12.5
33	Dnieper	500,000	0.80	H	75.0	2.7	19.7	11.9
34	Zhujiang	450,000	1.18	H	100.2	16.1	31.2	15.6
35	Don	423,000	0.62	SH	89.2	3.9	19.9	14.8
36	Limpopo	420,000	0.33	SA	33.0	11.2	16.5	10.0
37	Irrawaddy	406,000	1.24	H	107.6	15.8	41.6	5.9
38	Indigirk	334,000	0.73	H	34.2	9.1	17.1	10.0
39	Pechora	314,000	1.24	H	100.1	26.3	27.1	11.0
40	Godavari	312,000	0.64	SH	176.2	54.1	34.5	14.1
41	Ural	296,000	0.37	SA	69.4	2.8	16.4	9.8
42	N. Davina	288,000	1.11	H	88.1	12.4	31.4	7.0
43	Huai	244,000	0.73	H	45.3	28.8	28.6	10.1
44	Fraser	238,000	0.97	H	163.0	17.0	32.4	15.5
45	Anadyr	226,000	1.00	H	90.4	17.0	17.2	13.6
46	Olenek	223,000	0.75	H	44.4	15.4	21.7	5.7
47	Chubut	197,000	0.23	SA	66.1	9.7	29.1	8.4
48	Rufiji	187,000	0.62	SH	124.2	15.2	32.4	13.1
49	Taz	171,000	1.34	H	101.8	39.5	34.4	6.6
50	Pyasina	164,000	1.35	H	93.8	33.3	27.6	7.7
51	Central Valley	156,000	0.55	SH	128.1	7.5	25.2	14.2
52	Essequibo	151,000	1.17	H	153.2	45.2	62.6	8.0
53	Koksoak	150,000	1.69	H	67.5	18.5	33.7	5.3
54	Loire	118,000	0.88	H	82.7	12.5	34.8	7.8
55	Narmada	114,000	0.64	SH	190.5	72.7	51.7	8.6
56	Flinders	110,000	0.27	SA	44.7	18.8	26.2	4.7
57	Cunene	98,000	0.40	SA	151.2	27.5	45.6	4.2
58	Douro	97,000	0.64	SH	77.5	16.2	35.1	5.1
59	Barito	74,000	2.02	H	70.9	28.4	48.3	4.8
60	Gambia	72,000	0.46	SA	152.8	39.2	36.6	7.9

<sup>a</sup>P denotes mean annual precipitation. The aridity index (AI, mean annual precipitation/mean annual potential evapotranspiration) is available at the Consultative Group for International Agricultural Research Consortium for Spatial Information (CGIAR CSI) (<http://www.cgiar-csi.org/data/globalaridity-and-pet-database>). Climate of the river basins is categorized as humid (AI > 0.65), subhumid (AI ≤ 0.65 and > 0.5), semiarid (AI ≤ 0.5 and > 0.2), arid (AI ≤ 0.2 and > 0.05), and hyperarid (AI ≤ 0.05). In this study, arid and hyperarid are grouped into arid. TWS<sub>a</sub> and TWS<sub>sa</sub> are annual and semiannual variability of TWS changes using the CSR gridded product and least squares fit using sine and cosine functions. Total errors are the standard deviation of time variable measurement error and leakage error. STNR represents signal-to-noise ratio of TWS<sub>a</sub> to total error.

human impacts on SWS and GWS changes. Detailed descriptions about these LSMs are provided in the supporting information.

### 2.3. Studied River Basins

Sixty river basins globally representing a broad range of climate, land cover, and extent of irrigation were selected in this study. These 60 basins include all 46 basins used in the analysis of gridded and basin-scale GRACE products in *Landerer and Swenson* [2012]. More basins in our study were selected especially those in relatively dry areas to provide a more comprehensive evaluation of difference approaches for deriving GRACE TWS changes. 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, see a global map of AI in Figure S1) [*Trabucco and Zomer*, 2009], including humid (AI > 0.65), subhumid (AI ≤ 0.65 and > 0.5), semiarid (AI ≤ 0.5 and > 0.2), arid (AI ≤ 0.2 and > 0.05), and hyperarid (AI ≤ 0.05). In this study, arid and hyperarid are grouped into arid. Precipitation from the TRMM Multisatellite Precipitation Analysis (TMPA 3B43) Product (lat: 50°S–50°N; long: 180°W–180°E) [*Huffman et al.*, 2007] or GLDAS-1 forcing (lat: 60°S–90°N; long: 180°W–180°E) [*Rodell et al.*, 2004] were extracted to calculate monthly precipitation anomalies for these basins.

Based on these criteria, 57% of the river basins (34) are categorized as humid climate, 15% (9) as subhumid climate, 25% (15) as semiarid, and 3% (2) as arid (including the hyperarid). Some of the basins have been subjected to large-scale groundwater depletion from irrigation, including the upper reach of the Arkansas basin (the Central High Plains), Sacramento-San Joaquin River basin in California, the Tigris basin in the Middle East, and the lower reach of the Indus River [*Rodell et al.*, 2009; *Scanlon et al.*, 2012a, 2012b; *Voss et al.*, 2013].

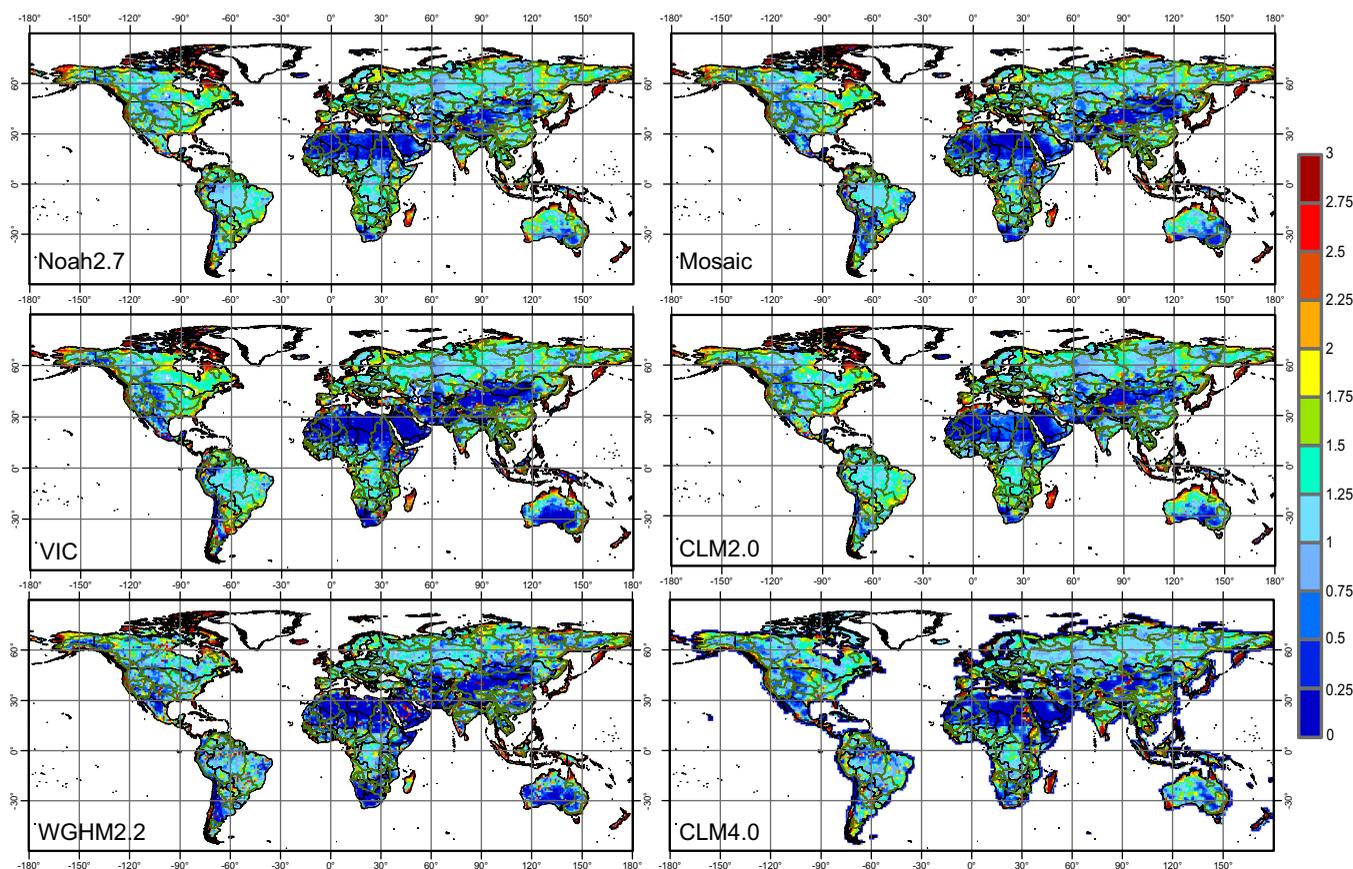
### 2.4. Scaling Factor Approach

For the scaling factor approach [*Landerer and Swenson*, 2012], GRACE TWS anomalies are derived by two steps. First, GRACE SH coefficients are destriped, truncated at the maximum degree and order of 60, and filtered using a 300 km Gaussian filter. After performing a series of low-pass filtering in the spectral domain, the filtered SH coefficients were converted to the spatial domain at a resolution of 1° × 1°. Second, the scaling factors (*k*), termed gain factors in [*Landerer and Swenson*, 2012], were computed by applying the same filters as applied to GRACE data to a numerical LSM (NCAR's CLM4). The filtered GRACE TWS anomalies at a grid cell were corrected by multiplying with the scaling factor for the same grid cell. Spatially averaging TWS anomalies over an arbitrary region with an area ≥ 200,000 km<sup>2</sup> is still required when using the gridded products. Applying filtering to a synthetic mass distribution is sometimes referred to as “forward modeling” and generates a mass distribution similar to what GRACE sees.

In short, the scaling factor is the multiplicative factor that minimizes the difference between the filtered and unfiltered modeled monthly TWS anomalies for a specific region of interest at both 1° × 1° and river basin scales (equation (1)).

$$M = \sum_{i=1}^T (\bar{S}_{m,i} - k \hat{S}_{m,i})^2 \quad (1)$$

where *M* is the objective function that we want to minimize;  $\bar{S}_{m,i}$  and  $\hat{S}_{m,i}$  are the unfiltered and filtered modeled TWS anomalies, respectively, for month *i*; *k* is the scaling factor derived through a least squares regression; *T* is the number of months used (January 2003 through December 2009), i.e., 84 months consistent with the time period in *Landerer and Swenson* [2012]. The scaling factors are provided separately from the filtered gridded GRACE data on the JPL website (<http://grace.jpl.nasa.gov/data/gracemonthlymassgrids-land/>). A schematic of GRACE data processing using the scaling factor approach is shown in Figure S2 (supporting information). The scaling factor ranges from negative values up to −20, with 90% ranging from 0.3 to 3.0 based on GLDAS-1 Noah (see Figure 2 in section 3.1.1). Spatially, variability in scaling factors is generally smooth, with some pixels showing up. Estimation of scaling values relies on a priori LSM quality to model both pixel and large-scale behavior. Computed values need to be compared to references for interpretation (Table 2). At grid cell scale, the scaling factor translates the modeled pixel behavior with respect to a large-scale behavior. For example, if the pixel behavior is exactly similar to the large scale, then *k* = 1.



**Figure 2.** Spatial distributions of scaling factors (spatial resolution 1° × 1°) across the global land surface derived from GLDAS-1 four LSMs (Noah2.7, Mosaic, VIC, and CLM2.0), WGHM2.2, and the scaling factors generated by CLM4.0 provided by the JPL website. Scaling factors greater than 3 were set to 3 whereas negative scaling factors were set to 0.

Values beyond the 90% range, i.e., negative values, below 0.3, or larger than 3.0 should be carefully interpreted.

Errors in the restored TWS anomalies include GRACE Level 1 (L1) measurement error in monthly gravity field solutions [Wahr et al., 2006] and leakage error (leakage error here includes both bias and leakage errors defined in our study, see section 2.5). The leakage error results from the deviation from the filtered TWS anomalies applied with the scaling factors and original TWS anomalies from a LSM [Landerer and Swenson, 2012]. Both measurement error and leakage error for each grid cell are available at the JPL website (<http://grace.jpl.nasa.gov/data/gracemonthlymassgridsland/>). Measurement and leakage errors are assumed to be constant over time for each grid cell in the gridded products and are summed in quadrature to estimate the total error for each grid.

**Table 2.** Interpretation of Magnitudes of Scaling Factors in Terms of Pixel Behavior With Respect to Large-Scale Mean

Scaling Factor Value	Interpretation
$k < 0$	Out of phase
$k = 0$	Uncorrelated behavior
$0 < k < 0.3$	Leakage prominent (e.g., desert)
$0.3 < k < 1$	Regular, lower amplitude at pixel, leakage important
1	Exact same behavior
$1 < k < 3.0$	Regular, larger amplitude of pixel
$k > 3.0$	Region close to ocean (average with 0), or large local behavior (e.g., river)

At the basin scale, measurement or leakage errors of a study basin are spatially correlated and should be lower than those calculated by simply averaging errors of grid cells within a basin. Therefore, to obtain a more realistic error estimate for a study basin based on measurement error or leakage error at grid cells for the basin, an approach accommodating spatial correlation of errors was used following [Landerer and Swenson, 2012]:

$$\sigma_{mb(lb)} = \sqrt{\text{var}/N} \tag{2}$$

$$\text{Var} = \sum_{i=1}^N \sum_{j=1}^N w_i w_j \text{cov}(w_i, w_j) \quad (3)$$

$$\text{cov}(x_i, x_j) = \sigma_i \sigma_j \exp\left(\frac{-d_{ij}^2}{2d_0^2}\right) \quad (4)$$

$$d_{i,j} = a \frac{\pi}{180} \sqrt{[(\text{long}(i) - \text{lon}(j)) \cos(\text{lat}(i))]^2 + [\text{lat}(i) - \text{lat}(j)]^2} \quad (5)$$

where  $\sigma_{mb}$  is the measurement error of a basin;  $\sigma_{lb}$  is the leakage error of a basin;  $N$  is the number of grid cells in a basin; subscripts  $i$  and  $j$  represent two different grid cells in a basin; var is the measurement or leakage error variance of mean TWS anomalies of a basin;  $w$  is the area weight at each grid cell in the basin and simplified to  $1/N$  under the assumption of equal contribution from each grid cell to the basin average TWS anomalies; cov is the covariance between two grid cells;  $\sigma$  is the standard deviation of measurement or leakage error of a grid cell;  $d_{ij}$  is the distance between two grid cells;  $d_0$  is a decorrelation-length scale, i.e., 300 km for measurement error and 100 km for leakage error;  $a$  is Earth's radius (6371 km); long and lat denote longitude and latitude of a grid cell.

To investigate the dependence of scaling factors on LSMs, TWS anomalies for January 2003 to December 2009, consistent with the time period in *Landerer and Swenson* [2012] from Noah2.7, Mosaic, VIC, and CLM2.0 in GLDAS-1 and WGHM2.2 were used to create scaling factors at both  $1^\circ \times 1^\circ$  and river basin scales globally. In combination with the scaling factors from NCAR's CLM4.0, there are six sets of scaling factors examined in this study. To investigate temporal variability in scaling factors, monthly scaling factors at  $1^\circ \times 1^\circ$  and basin scales for January 2003 to December 2009 were derived from the Noah model as an example. The coefficient of variation (CV) was used to quantify variability in scaling factors in the model and over time.

### 2.5. Additive Correction Approach

To extract TWS anomalies for a specific region of interest  $R$  with an area of  $R_0$ , a basin function  $h(X)$  (referring to the red rectangle in Figure S3a) is defined as 1 inside the basin ( $R$ ) and 0 outside the basin:

$$h(X) = \begin{cases} 1 & \text{if } X \in R \\ 0 & \text{if } X \in \Omega - R \end{cases} \quad (6)$$

where  $X$  is position on the Earth's surface  $X = (\theta, \lambda)$  and  $\Omega$  represents the entire Earth surface. If the maximum degree and order of SH coefficients ( $L_{\max}$ ) approaches infinity, the basin function approaches its ideal shape. However, for finite  $L_{\max}$  (degree 60 in this study), the basin function only approximates the exact  $h(X)$ . The effective basin function (EBF)  $\hat{h}$  is defined as the spatial function of the sphere and its corresponding SH coefficients have the same  $L_{\max}$  and filtered with the same filter as applied to GRACE data (referring to the black curve in Figure S3a). The EBF also describes the ability of GRACE to determine an average water storage change within the basin of interest [Longuevergne et al., 2010]. Attributes of the EBF are twofold: (1) mean value of the EBF over the basin of interest is no longer unity, resulting in a biased estimate; and (2) the EBF is not zero outside the basin, making the estimates sensitive to water storage changes outside the basin (leakage).

The filtered GRACE TWS anomaly for a specific basin of interest (Figure S3b) can therefore be written as [Wahr et al., 1998]:

$$\hat{S}_0 = \frac{1}{R_0} \int_{\Omega} \hat{S} \cdot h d\Omega \quad (7)$$

where  $\hat{S}$  is the filtered GRACE global TWS anomaly.

The additive correction approach uses output from an LSM, GLDAS Noah in this study, as a priori knowledge to estimate bias and leakage that can be numerically computed according to [Klees et al., 2007]:

$$S_B = \frac{1}{R_0} \int_R S_0 (h - \hat{h}) d\Omega \quad (8)$$

$$S_L = \frac{1}{R_0} \int_{\Omega-R} S_{\text{leak}} \cdot \hat{h} d\Omega \quad (9)$$

where  $S_B$  and  $S_L$  are bias and leakage, respectively;  $S_0$  and  $S_{\text{leak}}$  are assumed to reflect true water storage changes within and outside a study basin approximated by output from a LSM. Based on the definition of the additive correction approach, the final estimate of GRACE TWS anomaly is the filtered GRACE TWS anomaly, plus the bias (signal loss in a basin, referring to the gray area in Figure S3a), and minus the leakage (signal gain outside a study basin, referring to the blue area in Figure S3a), given as follows [Klees *et al.*, 2007]:

$$\bar{S}_0 = \hat{S}_0 + S_B - S_L \quad (10)$$

A schematic on the additive approach is provided in Figure S4. Errors in GRACE-derived TWS anomaly from the additive approach include: (1) GRACE measurement error and (2) propagation of bias/leakage correction [Longuevergne *et al.*, 2010] due to uncertainties in a priori SMS changes and/or missing components (e.g., deep SMS or GWS changes) in LSMs. Uncertainties in SMS changes were caused by missing components of LSMs (e.g., deep SMS (>2 m), redistribution of soil moisture due to irrigation, and GWS changes) and were quantified from the standard deviation of SMS changes among four LSMs (i.e., Noah, Mosaic, VIC, and CLM) in GLDAS-1. Uncertainty quantification for the additive correction approach needs to be further refined in the future to include the impact of missing components on the leakage correction. However, it would not impact most basins examined in this study because intensive irrigation is relatively localized (see the last part of section 2.3).

### 2.6. Multiplicative Correction Approach

The multiplicative correction approach assumes a uniform distribution of TWS changes within a basin [Chen *et al.*, 2005; Fenoglio-Marc *et al.*, 2006; Swenson and Wahr, 2007; Velicogna and Wahr, 2006], but considers the leakage error outside the basin. The multiplicative factor  $k_m$  is calculated as follows:

$$k_m = \left( \frac{1}{R_0} \int_R \hat{h} d\Omega \right)^{-1} \quad (11)$$

The corrected GRACE TWS anomaly based on the concept of the multiplicative correction approach can be computed as:

$$\bar{S}_0 = (\hat{S}_0 - S_L) \cdot k_m \quad (12)$$

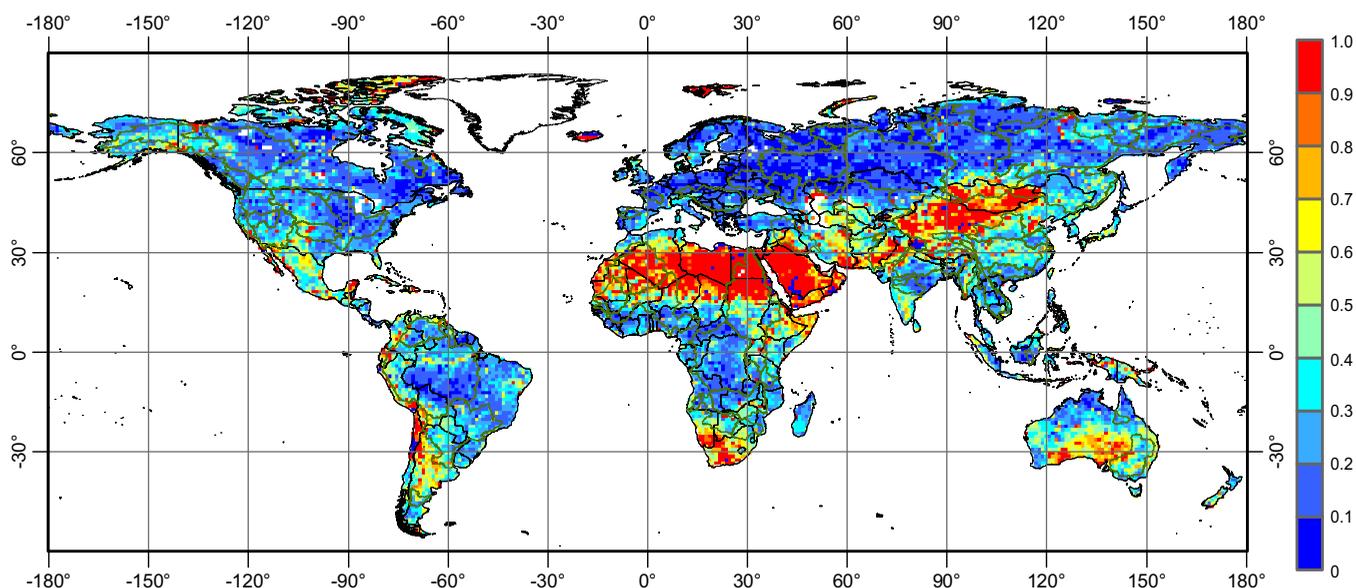
A schematic showing the processing of the multiplicative approach is given in Figure S5. It is worthwhile to emphasize the assumptions of the three approaches examined. The scaling factor approach assumes that the scaling factors from a LSM can correct for the bias and leakage effects integrally. The additive correction approach makes use of output from a LSM to correct for bias and leakage separately. The multiplicative correction approach partially uses output from LSMs for leakage correction and assumes uniform distribution of TWS changes within a basin. It is also emphasized that GLDAS-1 four LSMs and WGHM are used for generating scaling factors for examining the model dependence of the scaling factor approach (referring to section 3.1 in the following); the GLDAS-1 Noah model was adopted to examine the temporal variability model because it is one of the most widely used LSMs for GRACE signal restoration and groundwater disaggregation (referring to section 3.2); and GLDAS-1 Noah output was also taken as the a priori information for bias and leakage correction for the additive correction approach, and leakage correction for the multiplicative correction approach, both of which were compared with the scaling factor approach represented by CLM4.0 scaling factors and WGHM (referring to section 3.3). Again, the WGHM model output was considered as a reference for the three approaches for GRACE signal restoration, though uncertainties exist in the hydrological model, either [Haddeland *et al.*, 2014; Schewe *et al.*, 2014].

## 3. Results and Discussion

### 3.1. Model Dependence of Scaling Factor

#### 3.1.1. Global Scale at 1° × 1° Spatial Resolution

Magnitudes and spatial patterns of six sets of scaling factors were compared (Figure 2). In general, scaling factors from all the LSMs examined are similar over most grid cells, with 53% of the land surface showing



**Figure 3.** Coefficient of variation (CV) of scaling factors ( $1^\circ \times 1^\circ$ ) derived from GLDAS-1 four LSMs (Noah2.7, Mosaic, VIC, and CLM2.0), WGHM2.2, and CLM4.0 across the global land surface. CVs greater than 1 were set to 1.

CV (defined as the ratio between the standard deviation and the mean of scaling factors among models, and interpretable as a relative uncertainty)  $\leq 0.3$  (i.e., 30% uncertainty) and 75% showing  $CV \leq 0.5$  (Figure 3). However, 12% of the land surface shows  $CV > 0.8$  and even 8% show  $CV > 1$ . Relatively low dispersion in scaling factors was found at high latitudes, e.g., Russia and Europe, and humid areas, e.g., Central Amazon, eastern North America, and Central Africa (Figure 3). In contrast, large disparities in scaling factors (e.g.,  $CV \geq 0.6$ ) were found in relatively dry areas, e.g., the southern U.S. and Mexico, southwest and southern South America, North Africa, Middle East, Northwest China, Mongolia, and Central-South Australia where annual precipitation is generally  $\leq 200$  mm. For areas with intensive irrigation, scaling factors derived from different LSMs also show large differences, including the central U.S. (the central High Plains Aquifer), the Tigris and Euphrates region, North India region, and North China Plain. Detailed analyses of scaling factors at the basin scale and causes of variability are described in the following section.

The six LSMs generated similar global mean scaling factors ranging from 1.23 for CLM4.0 to 1.38 for Noah (Figure 4), which confirms that the processes of destriping (decorrelation), truncation, and filtering result in signal loss over most of the land surface; the filtered GRACE signal therefore needs to be compensated by multiplying by a scaling factor that is generally  $> 1.0$ . However, there are areas with scaling factors  $< 1.0$ , mainly in relatively dry areas with signal gain from the surrounding regions, e.g., the central and west U.S., southwestern South America, North Africa, Middle East, Central Asia, Northwest China, Mongolia, and Southeast/South Australia. The low-pass filtering process over these regions tends to amplify the signal; therefore, signal restoration is to remove signal gain (error from leakage into the basin) by multiplying by a scaling factor  $< 1.0$ .

Furthermore, the WGHM model generated the most heterogeneous scaling factors with the highest standard deviation of 1.61 of all LSMs. High spatial variability in TWS changes from WGHM is attributed to variability caused by human impacts on water storage changes related to agricultural, industrial and municipal water use which are not accounted for in GLDAS-1 LSMs. Standard deviations in scaling factors for the other five LSMs range from 0.92 for CLM4.0 to 1.43 for VIC. In terms of asymmetry in probability distributions (skewness), all LSMs are positively skewed, i.e., the mass of the probability distribution is concentrated on the left, with a longer tail to the right. Large scaling factors appear when local-scale behaviors differ from regional-scale behaviors, e.g., river storage changes and pumping in a small groundwater system. Skewness of scaling factors ranges from  $\sim 3.5$  for CLM4.0 to  $\sim 8.0$  for VIC. Relatively higher values of scaling factors ( $k > 3$ ) are localized and linked to a large mass concentration, with very high variations in TWS being filtered out during low-pass filtering, including areas along surface water bodies (e.g., the Nile and lower Yangtze) shown in CLM4.0 and areas with intensive irrigation (e.g., the upper Arkansas basin, the Middle East, North

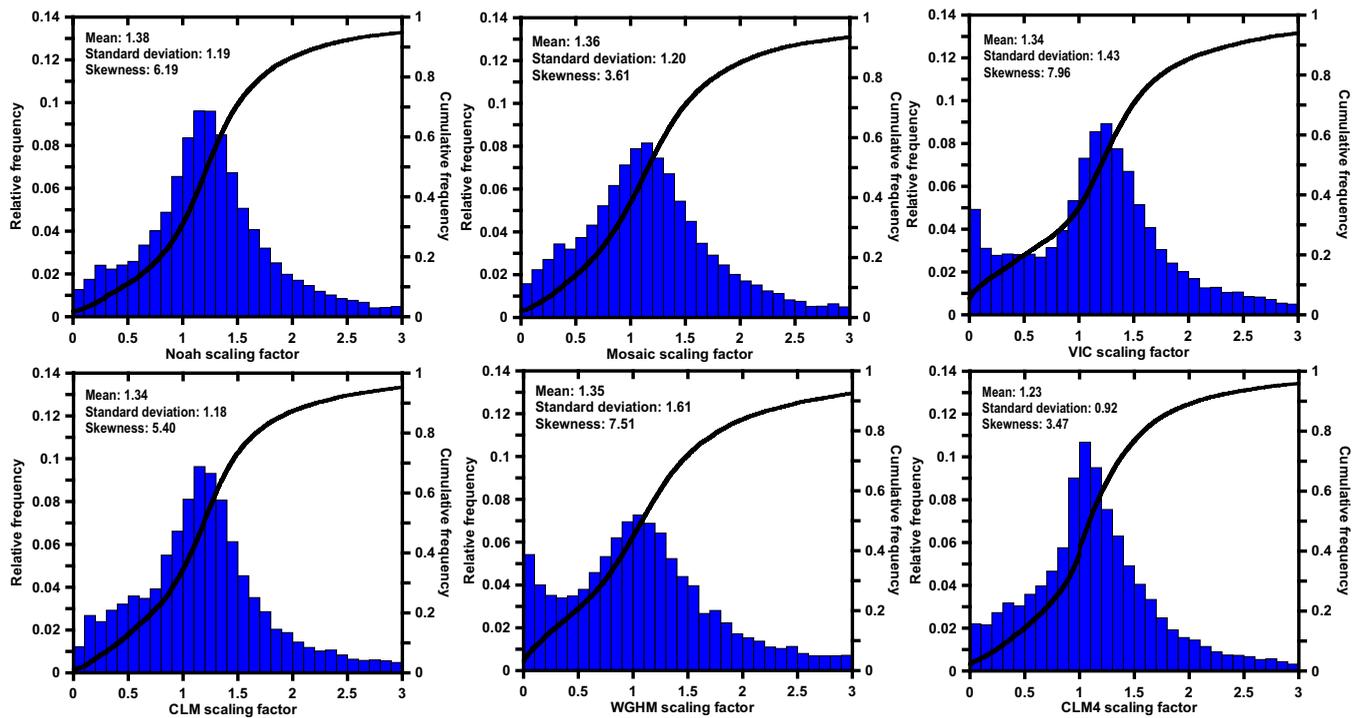


Figure 4. Frequency and cumulative distributions of scaling factors across the global land surface derived from the six LSMs examined, with showing mean, standard deviation, and skewness of these frequency distributions.

India, and North China Plain) shown in WGHM. Low values ( $k < 0.3$ ) are located in dry areas adjacent to regions with higher signal. Negative values mean local out of phase behavior with respect to the surrounding areas. In all these cases, the quality of the model (processes and parameterization) to represent small spatial-scale variability on  $1^\circ \times 1^\circ$  grids is critical and large uncertainties are expected in such regions.

### 3.1.2. Basin Scale

At the basin scale, CVs of scaling factors for the 60 basins evaluated are generally  $\leq 0.3$ , 53 basins (88%) with CVs  $\leq 0.2$ , and 47 basins (78%) with CVs  $\leq 0.1$  (Figure 5). Lower CVs of scaling factors are generally distributed in large (e.g., basin size  $\geq 1,000,000 \text{ km}^2$ ) except Nile, Aral, Tamanraset, Changjiang, Indus, and

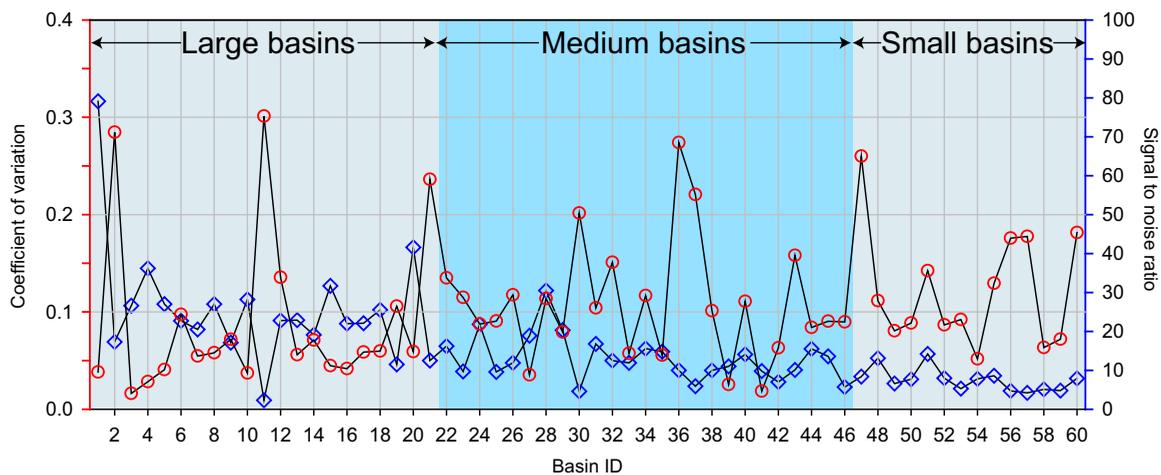


Figure 5. CVs of scaling factors derived from the six LSMs and signal-to-noise ratio (STNR) of GRACE signal for the 60 basins. Basin IDs and the corresponding basin names are presented in Table 1. Large basins have areas  $> 1,000,000 \text{ km}^2$ , medium basins with areas ranging from  $200,000$  to  $1,000,000 \text{ km}^2$ , and small basins with areas  $< 200,000 \text{ km}^2$  defined in this study.

Murray) and/or humid basins, e.g., Amazon (0.04), Zaire (0.02), Mississippi (0.03), Ob (0.04), Parana (0.1), Yenisei (0.05), and Lena (0.06). Relatively small differences in scaling factors reflect similarity in LSM TWS output [Landerer and Swenson, 2012]. Consistency in scaling factors over large and/or humid basins ( $CVs \leq 0.1$ ) derived from varying LSMs provides confidence in using GRACE data and its signal restoration with the scaling factor approach. Similarity in scaling factors in these large basins may be ascribed to bias and leakage of GRACE signals generally offsetting each other. Spatially averaging LSM output over a large region tends to filter out differences among various LSMs, resulting in similar scaling factors. However, this is not the case for some of the large basins, such as the Changjiang (Yangtze) River basin in China where the scaling factors range from 1.0 to 1.42 (Table 3). These differences may be related to the interactions between surface water and groundwater due to irrigation that are simplified or treated differently by various LSMs.

The scaling factors are generally much more variable across dry basins, though basin sizes are often enough for the footprint of GRACE, e.g., Tamanraset (0.3), Nile (0.2), Murray-Darling (0.24), Limpopo (0.27), and Chubut (0.26) (Figure 6 and Figure S6). For instance, the Nile basin shows the highest scaling factor of 1.21 from CLM4.0 but the lowest scaling factor of 0.55 from VIC. Therefore, CLM4.0 indicates signal loss but VIC indicates signal gain during the low-pass filtering process, which results in differences in amplitudes and surely long-term trends in TWS changes. This is also the case for the Murray-Darling basin where scaling factors vary from 0.63 (Mosaic) to 1.36 (CLM4.0) resulting in large variations in TWS recovery after the Millennium drought in 2010 (referring to precipitation anomalies in Figure S6). Differences in scaling factors over these arid regions are largely attributed to differences in partitioning water between infiltration and evapotranspiration in the LSMs; parameters associated with surface conductance for the land and atmosphere heat flux exchanges may differ markedly, resulting in large differences in output of TWS changes from LSMs [Long *et al.*, 2014, 2013].

Variability in scaling factors over humid to arid regions could be explained in part by the signal-to-noise-ratio (STNR, i.e., seasonality of TWS anomalies fitted using sine and cosine functions over the standard deviation of TWS anomalies). Scaling factor CVs seem to be negatively correlated with the STNR at the basin scale (Figure 5). This is reasonable because the lower STNR suggests a drier climate with amplified differences (errors) in partitioning energy and water fluxes among LSMs. Over arid and semiarid basins, the STNR values for both model output and GRACE are greatly reduced, highlighting the challenge in accurately simulating hydrological state and flux variables over these regions. The opposite may be true over humid/sub-humid regions, with generally higher STNRs favoring more consistent patterns among LSMs output.

However, some humid regions also exhibited relatively large variability in scaling factors, e.g., Irrawaddy (0.22) (Figure S7). For Irrawaddy, WGHM resulted in signal loss (scaling factor 1.35) and CLM4.0 showed a signal gain (scaling factor 0.67). A potential reason for this may be linked to proximity of basins with high amplitudes of storage variations, e.g., Irrawaddy (seasonal amplitude of TWS anomalies,  $TWS_a$ : 107.6 mm) adjacent to Ganges-Brahmaputra ( $TWS_a$ : 139.9 mm). To that end, LSM output in some of the humid may not be consistent and the associated scaling factors should be used with great caution. Deriving scaling factors from different LSMs provides an additional uncertainty estimate for GRACE signal.

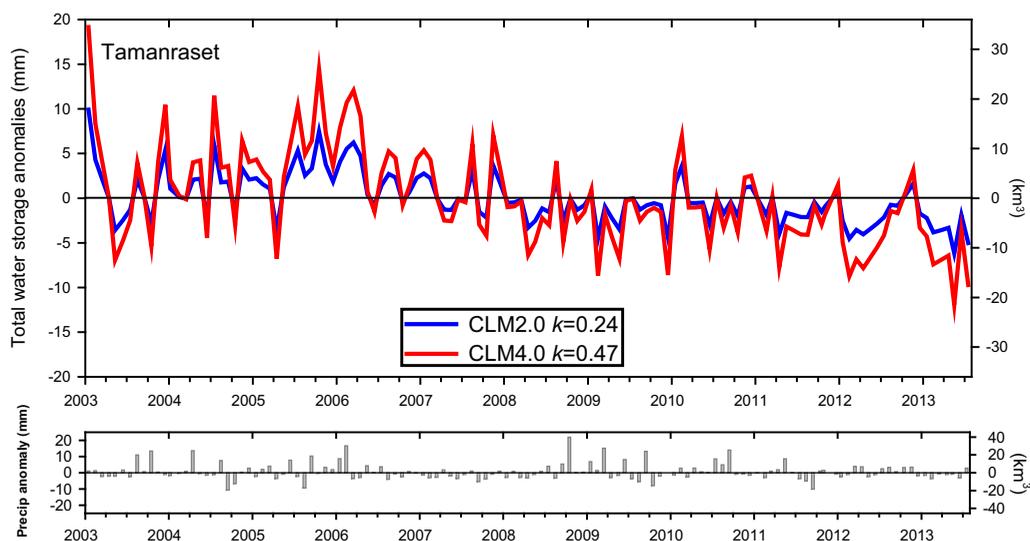
As was expected, basins with intensive irrigation show relatively large differences in scaling factors, attributed to whether natural and/or human-induced GWS changes are simulated. The Arkansas basin with intensive irrigation in its upper reaches (the Central High Plains) shows a scaling factor CV of 0.15 and the Sacramento-San Joaquin basin shows a CV of 0.14 (Figure 5). CLM4.0 provides a scaling factor of 0.79 but WGHM provides a scaling factor of 1.11 for the Arkansas basin (Figure 7). In this case, CLM4.0 indicates that the leakage is greater than the bias that is primarily due to a higher leakage-in signal in the eastern U.S. than the central and western U.S. The WGHM model accommodates human-induced groundwater storage changes over the Arkansas basin and the low-pass filtering process resulted in signal loss. Therefore, a scaling factor of 1.11 is required for signal restoration in the Arkansas basin. The situation is different in the Sacramento-San Joaquin basin where the six LSMs all indicated that the filtered TWS signal was dampened due to proximity of the ocean, with scaling factors from all LSMs exceeding 1.0, ranging from 2.67 (CLM4.0) to 3.66 (WGHM) (Figure 7). Basins with glacier melting also show relatively high scaling factor CVs, e.g., Yukon (0.13) ranging from 0.98 (WGHM) to 1.43 (Mosaic) (Figure S8). Different scaling factors suggest variations in both the seasonal cycles and long-term trends of TWS changes. Specific processing approaches are therefore imperative for these regions with intensive irrigation or glacial melting, generally unmodeled in a priori global models, to restore realistic TWS and GWS changes from GRACE and LSMs.

**Table 3.** Scaling Factors From GLDAS-1 Noah2.7 Published in *Landerer and Swenson* [2012], GLDAS-1 Noah2.7, Mosaic, VIC, and CLM2.0, and WGHM 2.2 in This Study<sup>a</sup>

Basin ID	Name	$k_{Landerer \& \ Swenson}$	$k_{Noah2.7}$	$k_{Mosaic}$	$k_{VIC}$	$k_{CLM2.0}$	$k_{WGHM2.2}$	$k_{CLM4.0}$	STNR
1	Amazon	1.11	1.19	1.21	1.17	1.23	1.11	1.15	79.1
2	Nile	NaN	0.74	0.77	0.55	0.67	0.89	1.21	17.3
3	Congo	1.24	1.23	1.22	1.25	1.24	1.20	1.20	26.6
4	Mississippi	1.11	1.13	1.06	1.07	1.11	1.11	1.05	36.2
5	Ob	1.08	1.09	1.11	1.18	1.09	1.16	1.06	27.0
6	Parana	1.38	1.40	1.14	1.49	1.47	1.35	1.27	22.7
7	Yenisei	1.19	1.18	1.07	1.10	1.08	1.00	1.05	20.5
8	Lena	1.21	1.33	1.31	1.29	1.25	1.26	1.12	27.0
9	Niger	0.94	0.81	0.85	0.75	0.73	0.82	0.88	17.1
10	Aral	NaN	1.11	1.05	1.09	1.02	1.11	1.04	28.2
11	Tamanraset	0.14	0.32	0.44	0.26	0.24	0.25	0.47	2.3
12	Changjiang	0.88	1.26	1.35	1.38	1.10	1.00	1.42	22.8
13	Amur	1.08	1.24	1.12	1.23	1.32	1.20	1.17	22.9
14	Mackenzie	1.15	1.18	1.15	1.14	1.07	1.21	0.99	19.1
15	Ganges	1.09	1.11	1.14	1.11	1.02	1.06	1.14	31.7
16	Volga	1.16	1.19	1.17	1.18	1.20	1.17	1.07	22.0
17	Zambezi	1.20	1.21	1.19	1.23	1.33	1.12	1.16	22.1
18	Missouri	0.89	0.90	0.87	0.86	0.92	0.95	0.80	25.5
19	Indus	1.34	1.26	1.08	0.99	1.12	1.05	1.29	11.5
20	Orinoco	1.26	1.33	1.35	1.28	1.15	1.29	1.36	41.6
21	Murray	1.02	1.13	0.63	1.17	1.00	0.96	1.36	12.5
22	Yukon	1.40	1.34	1.43	1.16	1.26	0.98	1.10	16.2
23	Senegal	NaN	0.64	0.73	0.65	0.56	0.68	0.78	9.7
24	Tigris	NaN	1.43	1.44	1.22	1.60	1.38	1.40	21.8
25	Colorado	1.21	1.15	1.07	1.06	1.14	1.09	0.88	9.6
26	Huanghe	Nan	1.11	0.93	0.84	0.91	1.12	1.07	11.9
27	Danube	1.17	1.18	1.19	1.16	1.18	1.20	1.28	18.9
28	Mekong	1.39	1.44	1.38	1.41	1.28	1.24	1.70	30.5
29	Columbia	1.36	1.37	1.17	1.35	1.39	1.25	1.17	20.3
30	Okavango	1.24	1.23	1.13	1.13	1.40	0.99	0.74	4.6
31	Kolyma	1.31	1.30	1.31	1.27	1.28	0.98	1.21	16.8
32	Arkansas	1.07	1.06	0.85	0.83	1.06	1.11	0.79	12.5
33	Dnieper	NaN	1.13	1.15	1.21	1.19	1.03	1.09	11.9
34	Zhujiang	NaN	1.36	1.43	1.36	1.22	1.09	1.53	15.6
35	Don	NaN	1.41	1.39	1.40	1.45	1.27	1.27	14.8
36	Limpopo	NaN	1.54	0.93	1.48	1.53	0.75	1.21	10.0
37	Irrawaddy	1.12	1.17	1.16	1.04	0.93	1.35	0.67	5.9
38	Indigirk	NaN	1.16	1.21	1.10	0.94	1.01	0.98	10.0
39	Pechora	NaN	1.31	1.29	1.33	1.37	1.37	1.30	11.0
40	Godavari	1.31	1.33	1.19	1.16	1.05	1.18	1.43	14.1
41	Ural	NaN	1.12	1.06	1.08	1.10	1.09	1.09	9.8
42	N. Davina	NaN	1.24	1.23	1.18	1.20	1.14	1.04	7.0
43	Huai	1.32	1.41	1.45	1.67	1.14	1.13	1.57	10.1
44	Fraser	1.66	1.69	1.49	1.59	1.53	1.33	1.41	15.5
45	Anadyr	1.66	1.69	1.66	1.63	1.58	1.94	1.49	13.6
46	Olenek	NaN	1.20	1.23	1.27	1.17	1.44	1.11	5.7
47	Chubut	1.93	1.90	1.71	2.48	1.39	1.23	2.15	8.4
48	Rufiji	1.25	1.27	1.02	1.27	1.42	1.18	1.34	13.1
49	Taz	1.35	1.36	1.29	1.39	1.26	1.51	1.21	6.6
50	Pyasina	1.59	1.62	1.59	1.51	1.50	1.68	1.29	7.7
51	Central Valley	3.22	2.81	2.74	3.41	2.70	3.66	2.67	14.2
52	Essequibo	1.18	1.18	1.22	1.25	1.15	1.39	1.41	8.0
53	Koksoak	1.69	1.73	1.54	1.56	1.64	1.77	1.37	5.3
54	Loire	1.51	1.58	1.56	1.61	1.56	1.47	1.72	7.8
55	Narmada	1.38	1.40	1.36	1.30	1.02	1.39	1.53	8.6
56	Flinders	1.55	1.54	1.62	1.73	1.62	1.37	1.01	4.7
57	Cunene	1.59	1.61	1.61	1.64	1.66	1.38	2.28	4.2
58	Douro	1.96	1.98	1.85	1.93	2.07	1.74	2.03	5.1
59	Barito	2.47	2.49	2.32	2.49	2.31	2.09	2.55	4.8
60	Gambia	1.46	1.48	1.25	1.39	1.16	1.00	1.68	7.9

<sup>a</sup>STNR represents the Signal-To-Noise-Ratio. NaN means no data.

Scaling factors derived from the Noah model in GLDAS-1 in this study differ slightly from those from *Landerer and Swenson* [2012] (Figure S9) showing a mean absolute percentage difference (MAPD) of 11%, which may be related to slightly different destripping (decorrelation) filters used in different studies. The point we



**Figure 6.** TWS anomalies from the scaling factor approaches for the period January 2003 to July 2013 using different LSMs for some dry basins, with showing monthly precipitation anomalies for the corresponding period.

are trying to make is that the additional LSMs examined in this study provide more comprehensive uncertainty estimates (CV) in scaling factors and that we need to pay special attention to regions with higher CVs. Furthermore, as stated by *Landerer and Swenson* [2012], TWS changes from LSMs tend to be dominated by the seasonal cycle of water storage variations, and may not be suitable for restoring secular trends in GRACE TWS change time series. Use of TWS anomalies from the scaling factor approach alone to quantify secular trends in groundwater depletion over areas with intensive groundwater pumping for irrigation may be inappropriate.

### 3.2. Temporal Variability in Scaling Factors

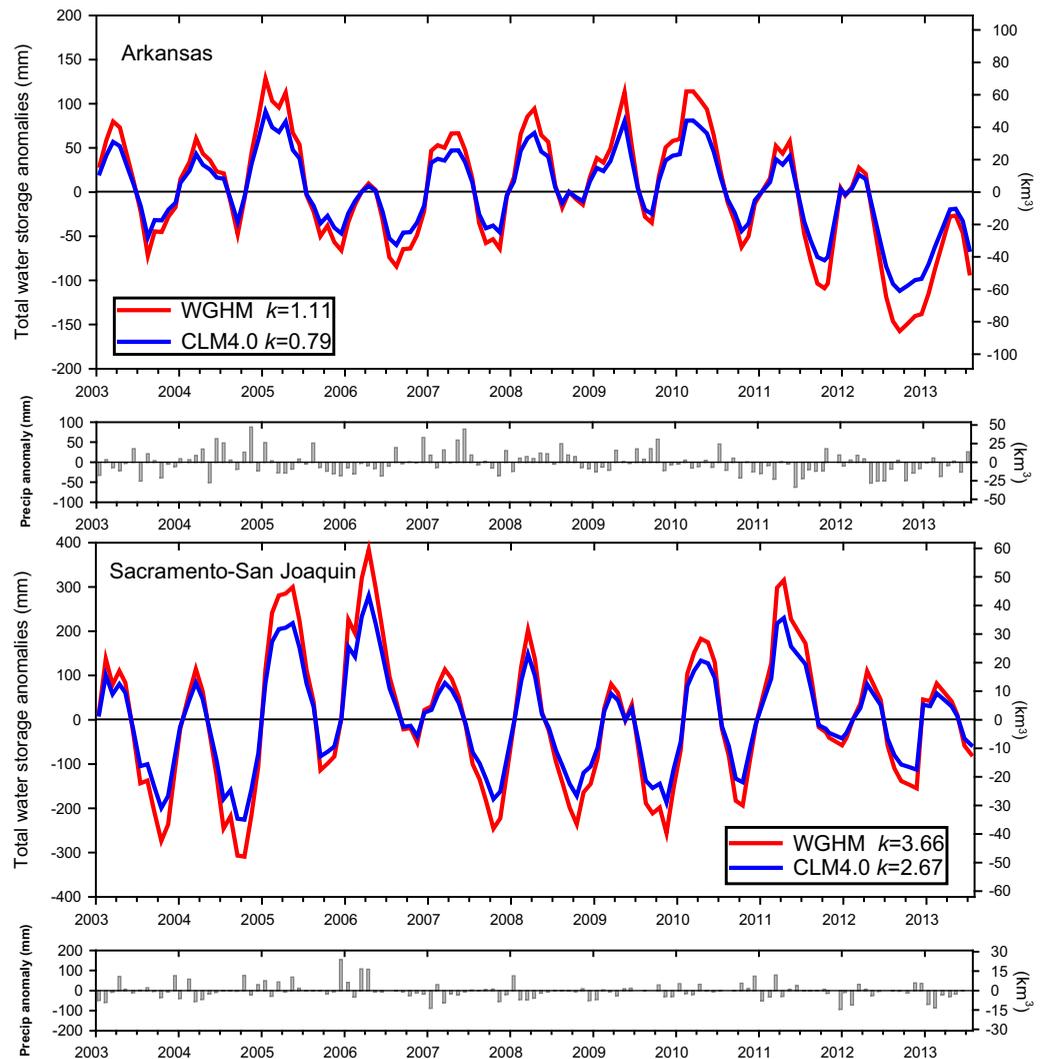
Temporal variability in scaling factors was computed from CVs of monthly scaling factor time series for January 2003 to December 2009 for each grid cell using GLDAS-1 Noah. Extreme values, i.e., <5% or >95%, of the monthly  $k$  samples were excluded from the analysis of CVs. Areas with negative CV values that often occur in relatively dry climate were also excluded from this analysis below.

#### 3.2.1. Global Scale at $1^\circ \times 1^\circ$ Spatial Resolution

From a global-scale perspective, 13.5% of the grid cells show CVs  $\leq 0.5$  and 42%  $\leq 1$ . Temporal variability in scaling factors was generally lower in humid areas, e.g., the equatorial region, eastern North America, and in some high-latitude regions, e.g., Europe and Russia (Figure 8). However, 58% of grid cells show CVs  $> 1.0$ , 30%  $> 2.0$ , and even 19.5%  $> 3.0$ . Higher temporal variability in scaling factors is generally found in dry areas where uncertainty in scaling factors from LSMs has been shown to be high (Figure 3), e.g., North Africa, the Middle East, North China, Mongolia, and southeast Australia. In addition, some high-latitude regions also display relatively high temporal variability in scaling factors, e.g., southern South America, northern North America, and east Russia. Furthermore, areas with intensive irrigation also have temporally variable scaling factors, e.g., the central High Plains (the upper Arkansas basin), Tigris and Euphrates areas, Indus basin, and North China Plain. High temporal variability in scaling factors may reflect: (1) local processes with respect to the large-scale mean that often occur in rivers or narrow mountain ranges, resulting in local model parameterization errors, and (2) the water cycle driven by episodic localized events (e.g., semi-arid climate), in contrast to temperate or tropical climates. Monthly scaling factors amplify local (space and time) variability in storage, and therefore are more sensitive to LSM errors.

#### 3.2.2. Basin Scale

Temporal variability in scaling factors at the basin scale is much lower than that at the grid cell described in section 3.2.1. Spatially averaging the monthly scaling factors acts as a low-pass filter, resulting in lower temporal variability. Results show 33 basins (55%) with CV values  $\leq 0.4$ , and 23 basins  $\leq 0.3$ . However, 12 basins (20%) have CV values  $> 0.7$ , and even 8 basins (13%) with CV values  $> 1$  (Figure 9).



**Figure 7.** TWS anomalies from the scaling factor approaches for the period January 2003 to July 2013 using different LSMs for some basins with intensive irrigation, with showing monthly precipitation anomalies for the corresponding period.

Examples of basins with low temporal variability in scaling factors in large basins and/or humid regions include the Zambezi, Amazon, Volga, Ob, Niger, Ganges, Mississippi, and Yenisei basins where CVs are  $\leq 0.15$  (Figure 9). Relatively lower variability in scaling factors indicates that the relative relationship of changes in TWS between a river basin and its surrounding region remains stable, including both relatively steady contributions from climate and human activities (e.g., human water use such as irrigation does not change substantially over the GRACE period). In this case, scaling factors derived from some LSM should be valuable in characterizing seasonal cycles in TWS variations. Relatively dry areas generally have high scaling factor variability, e.g., the Tamaraset (CV 3.94) and Aral (CV 1.57) (Figure 10 and Figure S10), attributed to precipitation variability showing less seasonality and predictability, likely resulting in large variability in relative differences in changes in TWS.

Alternatively, temporally uniform scaling factors in heavily irrigated basins, such as the Indus River basin (CV 1.34) and Sacramento-San Joaquin (CV 1.14) (Figure 10 and Figure S10) may be an artifact of the LSM (Noah) not considering human-induced GWS changes. In this case, separate scaling factors would be required to restore the natural variability and the long-term trends [Landerer and Swenson, 2012]. Again, the Yukon basin indicates relatively higher temporal variability (CV 1.4, Figure 10), which may be attributed to snow accumulation and melting processes that is time variable and may require use of a time varying scaling factor for signal restoration.

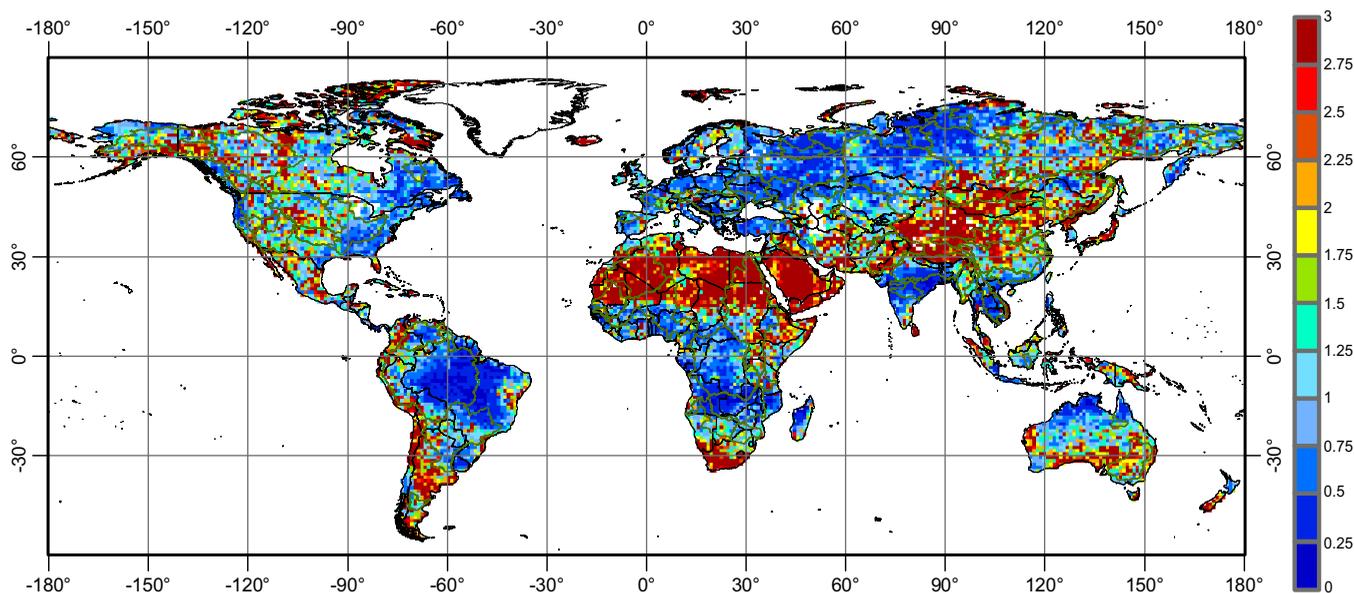


Figure 8. CVs of monthly scaling factors ( $1^\circ \times 1^\circ$ ) derived from GLDAS-1 Noah2.7 across the global land surface. CVs greater than 3 were set to 3.

### 3.3. Comparison of GRACE With WGHM Anomalies

#### 3.3.1. Statistical Results of Comparison

Statistics of the differences between GRACE TWS anomalies (i.e., CSR, JPL, and GFZ gridded products with CLM4.0 scaling factors and CSR SH coefficients with signal restoration using the additive correction and multiplicative correction approaches) and WGHM TWS anomalies for the 60 basins are shown in Figures 11 and 12 and Tables S1–S3. Correlations between GRACE and WGHM anomalies are high ( $r > 0.8$ ) over 40% (24) of the basins, moderate ( $r = 0.7–0.8$ ) over 23% (14) of the basins, and lower ( $r \leq 0.6$ ) over 37% (22) of these basins (Figure 11). The three gridded products (CSR, JPL, and GFZ) with CLM4.0 scaling factors and the CSR anomalies from the additive correction approach show similar orders of uncertainty with respect to WGHM TWS output, with medians of correlation coefficient, RMSD, and MAB of 0.8–0.84, 30–40, and 25–30 mm, respectively (Figure 12). However, GRACE TWS anomalies from the multiplicative correction approach resulted in the largest deviations from the WGHM TWS output, showing medians of  $r$  of 0.6, RMSD of 64 mm, and MAB of 50 mm. The multiplicative correction approach generally enhances amplitudes of TWS change signals over small basins (e.g.,  $< 200,000 \text{ km}^2$ ) that are mostly higher than the amplitudes of

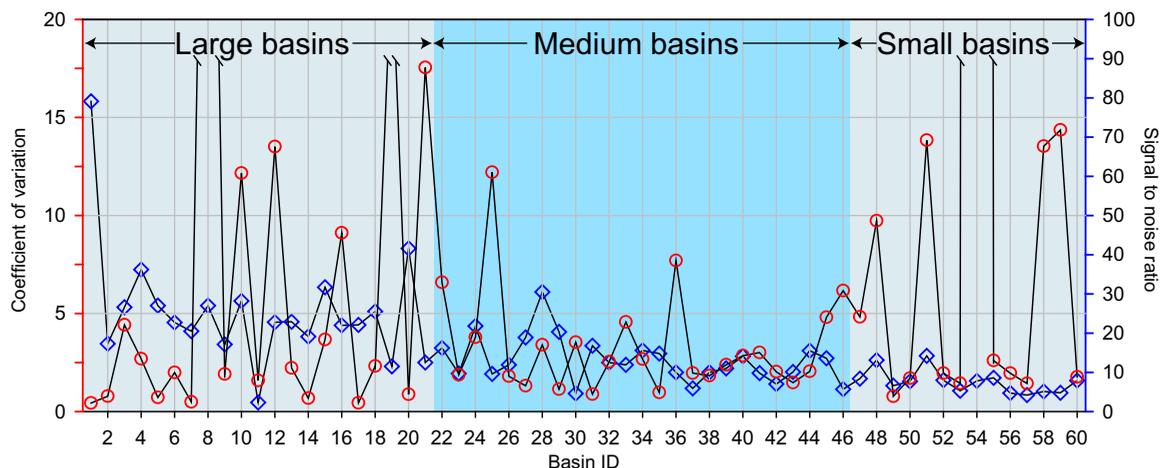
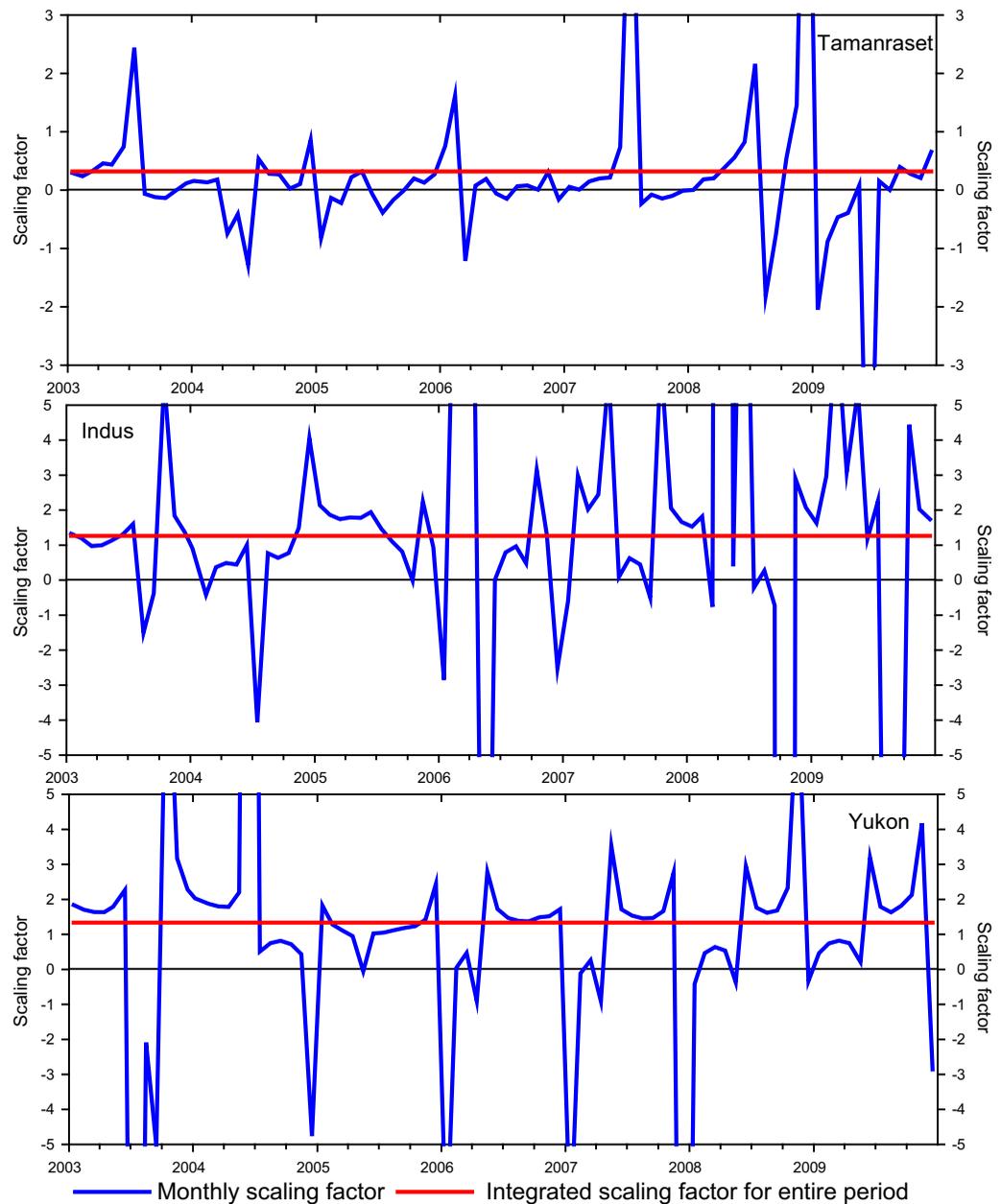


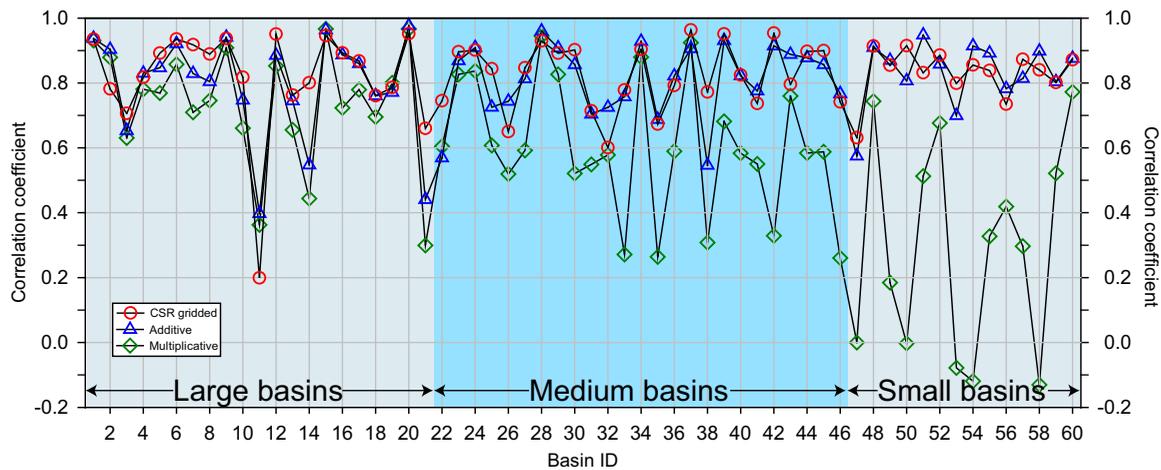
Figure 9. CVs of monthly scaling factors derived from GLDAS-1 Noah2.7 and STNR of GRACE signal for the 60 basins. Basin IDs and the corresponding basin names are presented in Table 1.



**Figure 10.** Basins with relatively high CVs of monthly scaling factor time series and its temporally constant scaling factors for the period January 2003 to December 2009 derived from GLDAS-1 Noah2.7.

TWS anomalies from other correction approaches and WGHM (Figure 11). This results in the TWS anomalies based on the multiplicative approach having higher uncertainty than other products relative to the WGHM model from a global perspective.

The primary difference between the multiplicative correction and the other two approaches is the assumption of uniformity in TWS changes within a basin, which is rarely met in reality because of the heterogeneity of water storage variations. Signal restoration from the scaling factor and additive correction approaches is based largely on the patterns of TWS variations from LSMs, as opposed to the multiplicative correction approach with a simple assumption of uniform TWS changes within a basin. This attribute tends to make TWS anomalies from the scaling factor and additive correction approaches closer to the output from the WGHM model that is taken as reference in this study. It should be noted that for relatively large basins, e.g., areas  $> 1,000,000 \text{ km}^2$  (basins ID number  $\leq 21$ , Figure 11), all of these approaches resulted in similar TWS



**Figure 11.** Correlation coefficients of TWS anomalies from the scaling factor approach with CLM4.0 scaling factors and CSR SH coefficients, additive correction approach, and multiplicative correction approach with TWS anomalies from the WGHM2.2 model for the 60 basins.

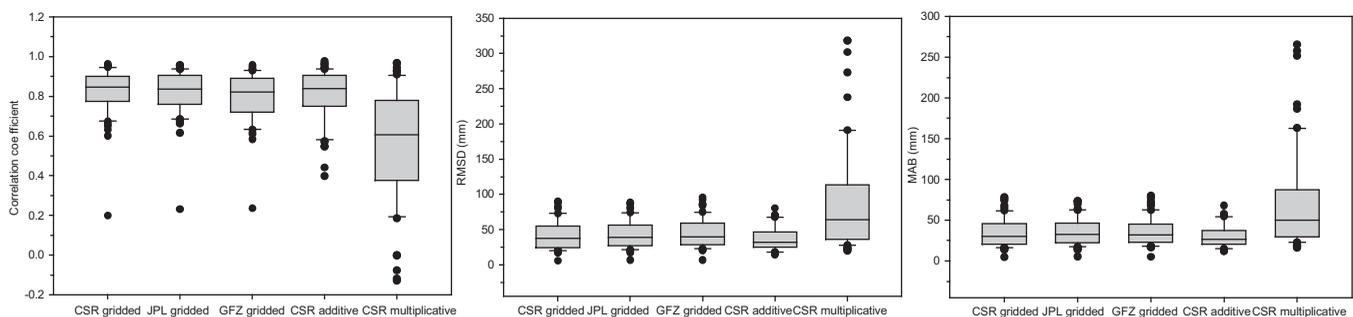
anomalies, showing mean correlation coefficient of  $\sim 0.8$ , RMSD of  $\sim 33$  mm, and MAB of  $\sim 27$  mm, though consistencies between GRACE TWS anomalies and WGHM output may differ to a varying degree. General similarity in large basins is most likely related to offsets between bias and leakage, reducing the importance of GRACE signal restoration.

TWS anomalies from the additive correction approach showed slightly lower RMSD of 31 mm and MAB of 24 mm relative to the CSR, JPL, and GFZ gridded products with signal restored using CLM4-based scaling factors. The GFZ gridded TWS product showed a slightly higher RMSD of 40 mm and MAB of 32 mm.

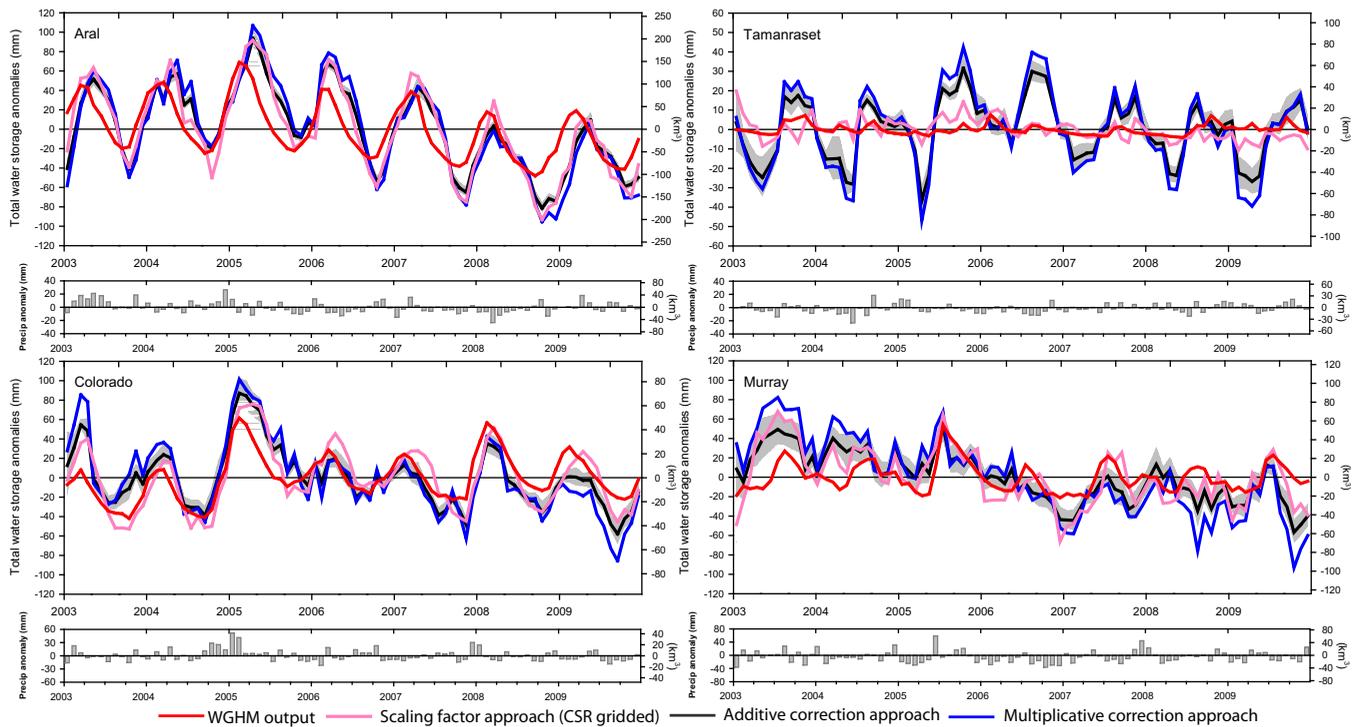
### 3.3.2. Comparison Over Basins With Varying Climates and Irrigation

By linking these statistics with the climate of the basins examined, we found that correlations between GRACE TWS anomalies from the five outputs (three gridded products with CLM4.0 scaling factors and additive and multiplicative correction approaches) and WGHM TWS output are generally higher in humid areas with an overall correlation coefficient of 0.84. Examples include the Amazon, Parana, Zambezi, and Orinoco basins that show mean correlation coefficients between GRACE TWS anomalies and WGHM TWS output of 0.93, 0.90, 0.84, and 0.97, respectively, and mean RMSD of 66.7, 25.0, 75.6, and 51.2 mm, respectively (Figure S11 and Tables S1–S3).

However, the arid and some semiarid basins have lower  $r$  (mean  $\sim 0.6$ ). Examples include Aral (semiarid) in Central Asia ( $r = 0.74$ , RMSD = 29.6 mm), Tamanraset (hyperarid) in North Africa ( $r = 0.32$ ; RMSD = 10.6 mm), Colorado (semiarid) in the Southwest U.S. ( $r = 0.72$ ; RMSD = 21.7 mm), and Murray-Darling (semiarid) in southeast Australia ( $r = 0.47$ ; 25.2 mm) (Figure 13 and Tables S1–S3). Differences in magnitudes of RMSD over these arid or semiarid basins relative to humid basins are attributed primarily to relatively lower seasonal (annual) amplitudes of TWS changes in arid and semiarid basins, e.g., Aral (46.3 mm),



**Figure 12.** Boxplots of correlation coefficients, root mean square difference (mm), and mean absolute bias (mm) of TWS anomalies from the scaling factor approach with CLM4.0 scaling factors and CSR SH coefficients, additive correction approach, and multiplicative correction approach with TWS anomalies from the WGHM2.2 model for the 60 basins.



**Figure 13.** TWS anomalies from the scaling factor approach with CLM4.0 scaling factors and CSR SH coefficients, the additive correction approach, and multiplicative correction approach, and from WGHM2.2 for some of the arid basins for the period January 2003 to December 2009, showing monthly precipitation anomalies for the corresponding period.

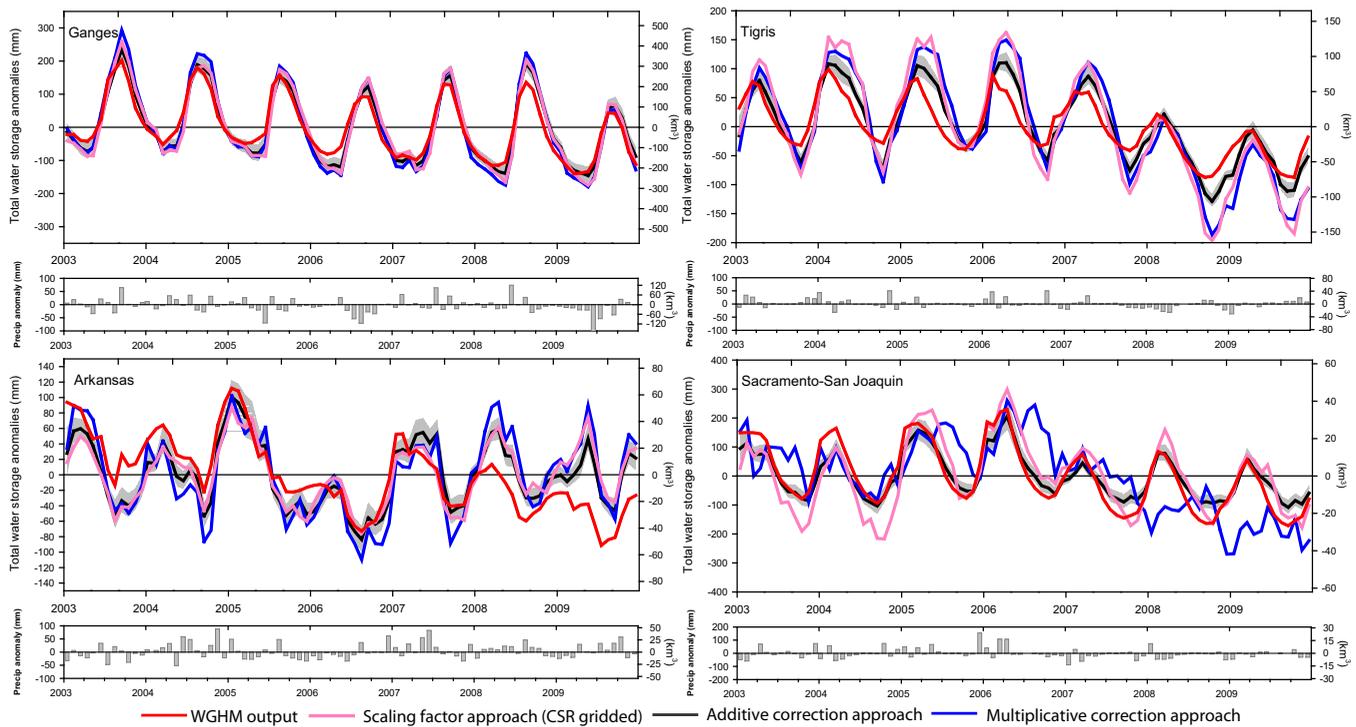
Tamanraset (2.2 mm), Colorado (37.3 mm), and Murray-Darling (19.9 mm) (Table 1). However, seasonal amplitudes of TWS changes over the Amazon (211.4 mm), Parana (75.6 mm), Zambezi (147 mm), and Orinoco (195 mm) are much higher than other regions globally, resulting in relatively large RMSD.

As far as basins in high-latitude regions are concerned, the mean  $r$  between GRACE TWS anomalies and WGHM output is also low ( $r = 0.63$ ; RMSD = 55.4 mm), e.g., Mackenzie ( $r = 0.60$ , RMSD = 33.7 mm), Yukon ( $r = 0.64$ ; RMSD = 64.4 mm), Koyama ( $r = 0.66$ ; RMSD = 37.9 mm), and Indigirka ( $r = 0.54$ ; RMSD = 37.8 mm) (Figure S12). Large differences in amplitude and timing of TWS variations over the high-latitude basins may be related to: (1) poor quality of LSMs, CLM, or GLDAS Noah used in the scaling factor and additive correction approaches to simulate hydrological processes in the cryosphere, and (2) poor performance of WGHM over these high-latitude regions. GRACE TWS anomalies are moderately correlated with WGHM TWS output over semiarid (mean  $r = 0.76$ ; RMSD = 32.7 mm) and subhumid regions (mean  $r = 0.72$ ; RMSD = 41.8 mm).

STNRs over the 60 river basins provide additional information to assess relationships between uncertainties in GRACE TWS anomalies and climate (Table 1). Humid regions have the highest STNRs (mean: 18) relative to arid (mean: 6) with subhumid regions (mean: 13) and semiarid regions (mean: 13) in between. These results suggest that GRACE TWS anomalies may be more accurate in relatively humid regions, based on comparisons with WGHM, and on the intrinsic attributes of GRACE signal, i.e., uncertainties in GRACE signal relative to its seasonal amplitude are lower in humid regions relative to arid or cold regions.

Apparently, basin size also partially determines the STNRs. Mean STNR values generally decrease with basin area, i.e., large basins ( $\geq 1,000,000 \text{ km}^2$ , mean: 26), medium basins ( $1,000,000\text{--}200,000 \text{ km}^2$ , mean: 13), and small basins ( $\leq 200,000 \text{ km}^2$ ; mean: 8). Therefore, use of GRACE TWS anomaly outputs may be more appropriate in humid and/or large basins than in arid (cold) and/or small basins.

As far as irrigation impacts on GRACE TWS signal restoration, basins with intensive irrigation examined in this study are found mostly in humid or subhumid regions, e.g., Ganges (humid), Arkansas (subhumid, Central High Plains aquifer), and the North Central Valley (subhumid), except for the Tigris (semiarid). Interestingly, correlations between GRACE TWS anomalies and WGHM output differ, e.g., 0.96 for Ganges (mean RMSD: 37.6 mm), 0.89 for the Tigris (mean RMSD: 47.7 mm), and 0.63 for the Arkansas (mean RMSD:



**Figure 14.** TWS anomalies from the scaling factor approach with CLM4.0 scaling factors and CSR SH coefficients, the additive correction approach, and multiplicative correction approach, and from WGHM2.2 for some of basins with intensive irrigation for the period January 2003 to Dec 2009, showing monthly precipitation anomalies for the corresponding period.

38.8 mm), and 0.76 for the Sacramento-San Joaquin (mean RMSD: 72.3 mm) (Figure 14). For the Ganges basin, TWS time series from GRACE and WGHM are highly consistent in terms of timing and amplitude (mean  $r = 0.96$ ; mean RMSD = 37.6 mm). For the Tigris basin, GRACE-derived TWS time series exhibit higher seasonal amplitudes than those of WGHM, with the scaling factor and multiplicative factor approaches generating higher amplitudes of TWS variation than the additive correction approach. It is also noted that the relatively high consistencies between GRACE and WGHM TWS anomalies in the Ganges (1,628,000 km<sup>2</sup>) and Tigris (823,000 km<sup>2</sup>) basins may be related to their large size. Signal restoration over large areas is not critical. For the Arkansas and Sacramento-San Joaquin basins, large differences in both timing and amplitude occur between the GRACE and WGHM TWS anomalies, especially with the multiplicative correction approach resulting in lower  $r$  values of only 0.58 and 0.51 than the scaling factor or additive correction approach. These results indicate that caution should be exercised in restoring GRACE signal over aquifers with intensive irrigation, especially over relatively small aquifers. The additive correction approach and gridded products based on CLM4.0 scaling factors provide more accurate TWS anomalies relative to the WGHM counterparts.

#### 4. Summary

Derivation of GRACE TWS anomaly requires removal of high-frequency noise through destriping (decorrelation), truncation, and low-pass filtering of high degree and order SH coefficients. Estimation of signal attenuation (bias) and signal interference from the area surrounding a study basin (leakage) requires (1) approximation to the true signal, generally a LSM output and/or (2) some hypotheses on the mass distribution within a basin (following LSM spatial distributions or homogeneity) to restore signal losses. We investigated the impact of LSM model errors (forcing and local process representation) and impact of approaches (scaling factor, additive, and multiplicative correction approaches) on restoration of GRACE TWS anomalies over the global land surface. Results are relatively independent of LSMs and approaches over most nonarid and/or large or medium river basins (area  $\geq 1,000,000$  km<sup>2</sup> or 1,000,000–200,000 km<sup>2</sup>). This implies that selection of LSMs for generating scaling factors and approaches for bias and leakage corrections for some of these areas is not critical. However, different approaches yield varying results for most arid and semiarid

regions, small regions (area < 200,000 km<sup>2</sup>), some humid regions, as well as regions with intensive irrigation due mostly to lack of relevant components (e.g., GWS changes) and irrigation in LSMs and limitations of LSMs in environments of high temporal and spatial variability (arid and semiarid), specifically:

1. Scaling factors are insensitive to selected LSMs for most nonarid basins; however, appreciable differences are found over most arid and semiarid regions (e.g., Tamaraset, Nile, and Murray-Darling), areas with intensive irrigation (e.g., Upper Arkansas and Sacramento-San Joaquin), and some humid basins adjacent to basins with large mass variability (e.g., Irrawaddy);
2. Temporal variability in scaling factors is relatively high ( $CV > 1$ ) over some arid and semiarid regions (e.g., Tamaraset and Aral), basins with intensive irrigation (e.g., Indus and Sacramento-San Joaquin), and basins with glacier melting (e.g., Yukon). A temporally variant scaling factor is highly dependent on the ability of the LSM to describe episodic events in semiarid regions; and
3. The additive correction approach and the gridded products with CLM4.0-based scaling factors more closely agree with WGHM output (taken as a reference) than the multiplicative correction approach. Large differences in TWS anomalies from the three approaches to restore GRACE signal are found in most arid and semiarid regions, areas with intensive irrigation, and relatively small basins. Therefore, the gridded products based on the scaling factor approach to restore signal losses over areas with intensive irrigation may not be suitable and the signal restoration method should be adapted. The multiplicative correction approach, in most cases, deviates from the other two approaches especially in smaller basins.

Areas with intensive irrigation and/or arid and semiarid regions have large variability in spatial, temporal, and model variability in scaling factors linked to misrepresented processes in LSMs. Different approaches for bias and leakage corrections yield largely different TWS anomalies with respect to the WGHM output, which may impact results of studies related to groundwater depletion and hydrological extremes, including droughts and floods. Accurate estimation of TWS and GWS anomalies over these regions should consider a priori knowledge (e.g., spatial patterns in groundwater depletion) through forward modeling schemes of GRACE data [e.g., Chen *et al.*, 2014; Feng *et al.*, 2013; Longuevergne *et al.*, 2013; Rodell *et al.*, 2009; Scanlon *et al.*, 2012b]. The gridded TWS anomaly products and regular processing for bias and leakage corrections may not be appropriate for these areas as the a priori LSM does not include these important processes. LSMs play an integral role in GRACE processing. Improving LSMs to simulate all storage compartments, their interactions and irrigation is very important for applications of GRACE to groundwater depletion and hydrological extreme studies. Future studies may consider multiple global hydrological models and the uncertainties among models examined in [Haddeland *et al.*, 2014; Schewe *et al.*, 2014]. Results of this study should be valuable in assessing uncertainties in GRACE-derived TWS anomalies and guide appropriate use of various approaches of correcting bias and leakage of GRACE signals.

#### Acknowledgments

This study was supported mostly by the BP project (CW1964449) with some additional support from the Jackson School of Geosciences at University of Texas at Austin and initial support from the NASA Project NNX09AN10G. 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 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/>); Gridded GRACE total water storage anomaly data are available at the Jet Propulsion Laboratory, California Institute of Technology (<http://grace.jpl.nasa.gov/data/gracemonthlymassgridsland/>); GLDAS-1 output is available at Goddard Earth Sciences Data and Information Services Center, NASA (<http://disc.sci.gsfc.nasa.gov/hydrology/data-holdings/>); and the global aridity index map is available at the Consultative Group for International Agricultural Research Consortium for Spatial Information (CGIAR CSI) (<http://www.cgiar-csi.org/data/global-aridity-and-pet-database>). Comments provided by four anonymous reviewers are highly appreciated. Publication is authorized by the Director, Bureau of Economic Geology, Jackson School of Geosciences, University of Texas at Austin.

#### References

- Alcamo, J., P. Doll, T. Henrichs, F. Kaspar, B. Lehner, T. Rosch, and S. Siebert (2003), Development and testing of the WaterGAP 2 global model of water use and availability, *Hydrol. Sci. J.*, 48(3), 317–337.
- Baur, O., M. Kuhn, and W. E. Featherstone (2009), GRACE-derived ice-mass variations over Greenland by accounting for leakage effects, *J. Geophys. Res.*, 114, B06407, doi:10.1029/2008JB006239.
- Cai, X. T., Z. L. Yang, C. H. David, G. Y. Niu, and M. Rodell (2014), Hydrological evaluation of the Noah-MP land surface model for the Mississippi River Basin, *J. Geophys. Res. Atmos.*, 119, 23–38, doi:10.1002/2013JD020792.
- Chen, J. L., C. R. Wilson, J. S. Famiglietti, and M. Rodell (2005), Spatial sensitivity of the Gravity Recovery and Climate Experiment (GRACE) time-variable gravity observations, *J. Geophys. Res.*, 110, B08408, doi:10.1029/2004JB003536.
- Chen, J. L., C. R. Wilson, and B. D. Tapley (2010a), The 2009 exceptional Amazon flood and interannual terrestrial water storage change observed by GRACE, *Water Resour. Res.*, 46, W12526, doi:10.1029/2010WR009383.
- Chen, J. L., C. R. Wilson, B. D. Tapley, L. Longuevergne, Z. L. Yang, and B. R. Scanlon (2010b), Recent La Plata basin drought conditions observed by satellite gravimetry, *J. Geophys. Res.*, 115, D22108, doi:10.1029/2010JD014689.
- Chen, J. L., C. R. Wilson, and B. D. Tapley (2013), Contribution of ice sheet and mountain glacier melt to recent sea level rise, *Nat. Geosci.*, 6(7), 549–552.
- Chen, J. L., J. Li, Z. Zhang, and S. Ni (2014), Long-term groundwater variations in Northwest India from satellite gravity measurements, *Global Planet. Change*, 116, 130–138, doi:10.1016/j.gloplacha.2014.02.007.
- Doll, P., F. Kaspar, and B. Lehner (2003), A global hydrological model for deriving water availability indicators: Model tuning and validation, *J. Hydrol.*, 270(1–2), 105–134.
- Eicker, A., M. Schumacher, J. Kusche, P. Döll, and H. Schmied (2014), Calibration/data assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model (WGHM) using an ensemble Kalman filter: First results, *Surv. Geophys.*, 35(6), 1285–1309.
- Feng, W., M. Zhong, J. M. Lemoine, R. Biancale, H. T. Hsu, and J. Xia (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, doi:10.1002/wrcr.20192.

- Fenoglio-Marc, L., J. Kusche, and M. Becker (2006), Mass variation in the Mediterranean Sea from GRACE and its validation by altimetry, steric and hydrologic fields, *Geophys. Res. Lett.*, *33*, L19606, doi:10.1029/2006GL026851.
- Forootan, E., R. Rietbroek, J. Kusche, M. A. Sharifi, J. L. Awange, M. Schmidt, P. Omondi, and J. Famiglietti (2014), Separation of large scale water storage patterns over Iran using GRACE, altimetry and hydrological data, *Remote Sens. Environ.*, *140*, 580–595.
- Gent, P. R., et al. (2011), The Community Climate System Model Version 4, *J. Clim.*, *24*(19), 4973–4991.
- Guo, J. Y., X. J. Duan, and C. K. Shum (2010), Non-isotropic Gaussian smoothing and leakage reduction for determining mass changes over land and ocean using GRACE data, *Geophys. J. Int.*, *181*(1), 290–302.
- Haddeland, I., et al. (2014), Global water resources affected by human interventions and climate change, *Proc. Natl. Acad. Sci. U. S. A.*, *111*(9), 3251–3256.
- Houborg, R., M. Rodell, B. L. Li, R. Reichle, and B. F. Zaitchik (2012), Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations, *Water Resour. Res.*, *48*, W07525, doi:10.1029/2011WR011291.
- Huffman, G. J., R. F. Adler, D. T. Bolvin, G. J. Gu, E. J. Nelkin, K. P. Bowman, Y. Hong, E. F. Stocker, and D. B. Wolff (2007), The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *J. Hydrometeorol.*, *8*(1), 38–55.
- Jacob, T., J. Wahr, W. T. Pfeffer, and S. Swenson (2012), Recent contributions of glaciers and ice caps to sea level rise, *Nature*, *482*(7386), 514–518.
- Klees, R., E. A. Zapreeva, H. C. Winsemius, and H. H. G. Savenije (2007), The bias in GRACE estimates of continental water storage variations, *Hydrol. Earth Syst. Sci.*, *11*(4), 1227–1241.
- Landerer, F. W., and S. C. Swenson (2012), Accuracy of scaled GRACE terrestrial water storage estimates, *Water Resour. Res.*, *48*, W04531, doi:10.1029/2011WR011453.
- Leblanc, M. J., P. Tregoning, G. Ramillien, S. O. Tweed, and A. Fakes (2009), Basin-scale, integrated observations of the early 21st century multiyear drought in southeast Australia, *Water Resour. Res.*, *45*, W04408, doi:10.1029/2008WR007333.
- Lo, M. H., J. S. Famiglietti, P. J. F. Yeh, and T. H. Syed (2010), Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data, *Water Resour. Res.*, *46*, W05517, doi:10.1029/2009WR007855.
- Long, D., B. R. Scanlon, L. Longuevergne, A.-Y. Sun, D. N. Fernando, and H. Save (2013), GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas, *Geophys. Res. Lett.*, *40*, 3395–3401, doi:10.1002/grl.50655.
- Long, D., L. Longuevergne, and B. R. Scanlon (2014), Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites, *Water Resour. Res.*, *50*, 1131–1151, doi:10.1002/2013WR014581.
- Longuevergne, L., B. R. Scanlon, and C. R. Wilson (2010), GRACE hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA, *Water Resour. Res.*, *46*, W11517, doi:10.1029/2009WR008564.
- Longuevergne, L., C. R. Wilson, B. R. Scanlon, and J. F. Crétaux (2013), GRACE water storage estimates for the Middle East and other regions with significant reservoir and lake storage, *Hydrol. Earth Syst. Sci.*, *17*(12), 4817–4830.
- Müller Schmied, H., S. Eisner, D. Franz, M. Wattenbach, F. T. Portmann, M. Flörke, and P. Döll (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*(9), 3511–3538, doi:10.5194/hess-18-3511-2014.
- Pokhrel, Y. N., Y. Fan, G. Miguez-Macho, P. J. F. Yeh, and S. C. Han (2013), The role of groundwater in the Amazon water cycle: 3. Influence on terrestrial water storage computations and comparison with GRACE, *J. Geophys. Res. Atmos.*, *118*, 3233–3244, doi:10.1002/jgrd.50335.
- Proulx, R. A., M. D. Knudson, A. Kirilenko, J. A. VanLooy, and X. D. Zhang (2013), Significance of surface water in the terrestrial water budget: A case study in the Prairie Coteau using GRACE, GLDAS, Landsat, and groundwater well data, *Water Resour. Res.*, *49*, 5756–5764, doi:10.1002/wrcr.20455.
- Reager, J. T., and J. S. Famiglietti (2009), Global terrestrial water storage capacity and flood potential using GRACE, *Geophys. Res. Lett.*, *36*, L23402, doi:10.1029/2009GL040826.
- Rodell, M., et al. (2004), The global land data assimilation system, *Bull. Am. Meteorol. Soc.*, *85*(3), 381–394.
- Rodell, M., I. Velicogna, and J. S. Famiglietti (2009), Satellite-based estimates of groundwater depletion in India, *Nature*, *460*(7258), 999–980.
- Save, H., S. Bettadpur, and B. Tapley (2012), Reducing errors in the GRACE gravity solutions using regularization, *J. Geod.*, *86*(9), 695–711.
- Scanlon, B. R., L. Longuevergne, and D. Long (2012a), Ground referencing GRACE satellite estimates of groundwater storage changes in the California Central Valley, USA, *Water Resour. Res.*, *48*, W04520, doi:10.1029/2011WR011312.
- Scanlon, B. R., C. C. Faunt, L. Longuevergne, R. C. Reedy, W. M. Alley, V. L. McGuire, and P. B. McMahon (2012b), Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley, *Proc. Natl. Acad. Sci. U. S. A.*, *109*(24), 9320–9325.
- Schewe, J., et al. (2014), Multimodel assessment of water scarcity under climate change, *Proc. Natl. Acad. Sci. U. S. A.*, *111*(9), 3245–3250.
- 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, doi:10.1029/2007GL030139.
- Sun, A. Y. (2013), Predicting groundwater level changes using GRACE data, *Water Resour. Res.*, *49*, 5900–5912, doi:10.1002/wrcr.20421.
- Swenson, S., and J. Wahr (2002), Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity, *J. Geophys. Res.*, *107*(B9), 2193, doi:10.1029/2001JB000576.
- Swenson, S., and J. Wahr (2006), Post-processing removal of correlated errors in GRACE data, *Geophys. Res. Lett.*, *33*, L08402, doi:10.1029/2005GL025285.
- Swenson, S., and J. Wahr (2007), Multi-sensor analysis of water storage variations of the Caspian Sea, *Geophys. Res. Lett.*, *34*, L16401, doi:10.1029/2007GL030733.
- Tiwari, V. M., J. Wahr, and S. Swenson (2009), Dwindling groundwater resources in northern India, from satellite gravity observations, *Geophys. Res. Lett.*, *36*, L18401, doi:10.1029/2009GL039401.
- Tourian, M. J., O. Elmi, Q. Chen, B. Devaraju, S. Roohi, and N. Sneeuw (2015), A spaceborne multisensor approach to monitor the desiccation of Lake Urmia in Iran, *Remote Sens. Environ.*, *156*, 349–360.
- Trabucco, A., and R. J. Zomer (2009), Global aridity index (global-aridity) and global potential evapotranspiration (global-PET) geospatial database, CGIAR-CSI, Washington, D. C. [Available at <http://www.csi.cgiar.org/>]
- Velicogna, I., and J. Wahr (2006), Measurements of time-variable gravity show mass loss in Antarctica, *Science*, *311*(5768), 1754–1756.
- Voss, K. A., J. S. Famiglietti, M. Lo, C. de Linage, M. Rodell, and S. C. Swenson (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, doi:10.1002/wrcr.20078.
- Wahr, J., M. Molenaar, and F. Bryan (1998), Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, *J. Geophys. Res.*, *103*(B12), 30,205–30,229.
- Wahr, J., S. Swenson, and I. Velicogna (2006), Accuracy of GRACE mass estimates, *Geophys. Res. Lett.*, *33*, L06401, doi:10.1029/2005GL025305.
- Wouters, B., D. Chambers, and E. J. O. Schrama (2008), GRACE observes small-scale mass loss in Greenland, *Geophys. Res. Lett.*, *35*, L20501, doi:10.1029/2008GL034816.