Assimilation of Gridded GRACE Terrestrial Water Storage Estimates in the North American Land Data Assimilation System

SUJAY V. KUMAR,[a] BENJAMIN F. ZAITCHIK,[b] CHRISTA D. PETERS-LIDARD,[a] MATTHEW RODELL,[a] ROLF REICHLE,[c] BAILING LI,[a,d] MICHAEL JASINSKI,[a] DAVID MOCKO,[a,c,e] AUGUSTO GETIRANA,[a,d] GABRIELLE DE LANNOY,[f] MICHAEL H. COSH,[g] CHRISTOPHER R. HAIN,[d,h] MARTHA ANDERSON,[g] KRISTI R. ARSENAULT,[a,i] YOULONG XIA,[j,k] AND MICHAEL EK[k]

[a] *Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland*
[b] *Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, Maryland*
[c] *Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland*
[d] *Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland*
[e] *Science Applications International Corporation, Beltsville, Maryland*
[f] *Department of Earth and Environmental Sciences, University of Leuven, Leuven, Belgium*
[g] *Hydrology and Remote Sensing Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland*
[h] *NOAA Center for Satellite Applications and Research, College Park, Maryland*
[i] *Science Applications International Corporation, McLean, Virginia*
[j] *IMSG at NCEP/EMC, College Park, Maryland*
[k] *Environmental Modeling Center, National Centers for Environmental Prediction, College Park, Maryland*

(Manuscript received 20 August 2015, in final form 27 February 2016)

ABSTRACT

The objective of the North American Land Data Assimilation System (NLDAS) is to provide best-available estimates of near-surface meteorological conditions and soil hydrological status for the continental United States. To support the ongoing efforts to develop data assimilation (DA) capabilities for NLDAS, the results of Gravity Recovery and Climate Experiment (GRACE) DA implemented in a manner consistent with NLDAS development are presented. Following previous work, GRACE terrestrial water storage (TWS) anomaly estimates are assimilated into the NASA Catchment land surface model using an ensemble smoother. In contrast to many earlier GRACE DA studies, a gridded GRACE TWS product is assimilated, spatially distributed GRACE error estimates are accounted for, and the impact that GRACE scaling factors have on assimilation is evaluated. Comparisons with quality-controlled in situ observations indicate that GRACE DA has a positive impact on the simulation of unconfined groundwater variability across the majority of the eastern United States and on the simulation of surface and root zone soil moisture across the country. Smaller improvements are seen in the simulation of snow depth, and the impact of GRACE DA on simulated river discharge and evapotranspiration is regionally variable. The use of GRACE scaling factors during assimilation improved DA results in the western United States but led to small degradations in the eastern United States. The study also found comparable performance between the use of gridded and basin-averaged GRACE observations in assimilation. Finally, the evaluations presented in the paper indicate that GRACE DA can be helpful in improving the representation of droughts.

1. Introduction

Terrestrial water storage (TWS), typically defined as the sum of soil moisture, surface water, groundwater, and snow, is an integrated measure of the hydrological cycle. The snow, surface water, and groundwater components of TWS are core water resource variables, relevant to monitoring and predicting hydrological drought and to making informed long-term water management decisions. Root zone soil moisture, meanwhile, is of central importance to agriculture and ecology. All TWS variables are known to impact land–atmosphere fluxes of heat and moisture in a manner that influences local climate and can have significant effects on numerical weather forecasts (Cohen and Entekhabi 1999; Koster et al. 2004; Maxwell et al. 2007).

Corresponding author address: Sujay Kumar, Hydrological Sciences Laboratory, NASA GSFC, Code 617, Greenbelt, MD 20771. E-mail: sujay.v.kumar@nasa.gov

DOI: 10.1175/JHM-D-15-0157.1

For all these reasons, an accurate representation of TWS and its components is a primary target for the North American Land Data Assimilation System (NLDAS; Mitchell et al. 2004). NLDAS is multi-institution effort focused on generating high-quality, spatially and temporally consistent land surface model datasets from best-available observations and model outputs. The NLDAS domain spans the continental United States (CONUS) at 1/8° spatial resolution and employs a high-quality forcing dataset that includes daily gauge-based precipitation analysis, bias-corrected shortwave radiation, and surface meteorology reanalysis. This forcing dataset is then used to drive four land surface models (LSMs) to generate hourly model outputs of land surface conditions including soil moisture, snow, runoff, streamflow, and land–atmosphere fluxes. NLDAS analyses support real-time numerical weather prediction, experimental drought monitoring, and retrospective studies of U.S. climate and hydrology.

Phase 2 of NLDAS (NLDAS-2; Xia et al. 2012b) includes several enhancements over phase 1. These include improvements to the forcing datasets and the ability to generate model products operationally in real time. The first two phases of the NLDAS project, however, do not employ the assimilation of remotely sensed datasets of land surface variables. As NLDAS development continues, efforts are focused on bridging this gap by enabling the assimilation of various hydrological remote sensing products. Kumar et al. (2014) demonstrated the assimilation of soil moisture and snow remote sensing retrievals into the Noah LSM in the NLDAS-2 configuration. Given the importance of TWS as a hydrological variable, there is a strong reason to include the assimilation of TWS estimates derived from the Gravity Recovery and Climate Experiment (GRACE; Tapley et al. 2004) in these efforts. Among the LSMs being targeted for inclusion in NLDAS at this time, the Catchment land surface model (CLSM; Koster et al. 2000) is the only one that simulates groundwater storage variations, which is a prerequisite for GRACE data assimilation (DA). This motivates our choice to make CLSM the sole focus of this study.

GRACE offers unprecedented ability to monitor changes in total, column-integrated TWS, from surface through groundwater. This capability has supported groundbreaking research on water cycle dynamics, water resources, and the cryosphere from basin to continental scales (e.g., Rodell et al. 2009; Velicogna 2009; Rodell et al. 2011). The spatial and temporal resolutions of the GRACE measurements, however, are problematic for use in hydrological applications at subbasin scale or submonthly time periods. The standard GRACE products are available on a monthly basis with an effective spatial resolution of no better than 150 000 km^2 at midlatitudes (Rowlands et al. 2005; Swenson et al. 2006). Further, they typically have a 2–4-month latency, which greatly limits their value for real-time applications.

Data assimilation has been used as an effective way to combine information provided by GRACE with the estimates from an LSM and to generate spatially and temporally disaggregated and continuous, low-latency TWS estimates. Zaitchik et al. (2008) first demonstrated the assimilation of GRACE observations into a land surface model over the Mississippi River basin. They found that ensemble-based assimilation of GRACE data led to improved correlations with observed groundwater storage variations and river flow at sub-GRACE spatial and temporal scales. Su et al. (2010) and Forman et al. (2012) examined the impact of GRACE assimilation on the simulation of snowpack over snow-dominated basins and found improvements in hydrologic state and flux estimations. Subsequent studies by Houborg et al. (2012) and Li et al. (2012) extended the assimilation approach of Zaitchik et al. (2008) to larger domains, that is, North America and Europe, respectively. They also demonstrated that GRACE-based drought indicators can be used for objective identification of droughts.

The core assimilation routine in these aforementioned studies ingested GRACE data that had been preprocessed by averaging up to river basin scales. While this approach simplified the task of dealing with spatially correlated errors, it also introduced artificial boundaries and did not make optimal use of subbasin-scale information contained in the GRACE TWS anomaly fields. A few prior studies (Su et al. 2010; Eicker et al. 2014; Tangdamrongsub et al. 2015) have explored the use of the gridded GRACE data product for DA. Among these, the two newer studies (Eicker et al. 2014; Tangdamrongsub et al. 2015) were conducted over a single river basin and employed GRACE DA with a hydrological model. Eicker et al. (2014) showed that assimilation of gridded and basin-averaged GRACE data provided comparable results. The gridded assimilation provided a closer match to GRACE observations when gridded GRACE products were aggregated at 5° resolution with spatially uncorrelated error assumptions. Eicker et al. (2014), however, did not provide independent evaluations of the gridded and basin-averaged GRACE DA approaches. Previous GRACE assimilation studies have been limited in compiling the best possible set of spatially complete, quality-controlled in situ observational data to evaluate the performance of the DA system. Most studies did perform a set of evaluations, but it is extremely

challenging to work with all relevant datasets in a systematic way, and in many regions independent validation data simply do not exist.

The objectives of the present study are threefold. First, we examine the utility of GRACE DA to contribute to model skill beyond what NLDAS-2 accomplishes on the basis of best-available forcing and parameter data quality. To accomplish this objective, the software framework, forcing data, and domain are all set to match NLDAS-2. CLSM, which we use as the land surface model into which GRACE observations are assimilated, is expected to be included in the next phase of NLDAS. Our second objective is to assess the utility of assimilating gridded GRACE data products along with their spatially distributed error estimates, instead of assimilating GRACE data that have been preprocessed to subjectively defined basins. The evaluation is conducted through a comparison of the gridded assimilation strategy to the basin-averaged assimilation approach employed in previous GRACE DA studies. The use of gridded GRACE products is an important step toward standardizing GRACE DA and facilitating its extension to other regions. The third objective of this study is to examine the impact of GRACE DA on key terrestrial water cycle components. We make use of the Land surface Verification Toolkit (LVT; Kumar et al. 2012) to perform the most extensive observation-based evaluation of model output that we found to be feasible for the NLDAS domain. This is complemented by quantitative comparisons to the operational U.S. Drought Monitor (USDM; Svoboda et al. 2002) in order to assess one likely application of an NLDAS configuration that includes GRACE DA.

2. Data and methods

a. GRACE terrestrial water storage

GRACE consists of two satellites following one another in near-polar orbit. The twin satellite system measures the spatiotemporal variations in the earth's gravity field based on perturbations to the orbits of those satellites. After removing the influence of atmospheric and oceanic circulations and glacial isostatic adjustment, the remaining signal on monthly to interannual time scales can be attributed to the variations of terrestrial water storage. Converting GRACE observations into estimates of TWS variations is not trivial, and methods for optimizing the accuracy and resolution of this conversion are still a subject of active research (e.g., Sakumura et al. 2014). In many cases, scientists extract time series of TWS change for specific regions of interest using geographically customized weighting schemes (e.g., Swenson and Wahr 2002). However, monthly gridded products are also made available to the broader research community. In particular, the NASA Jet Propulsion Laboratory (JPL) Tellus website (http://grace.jpl.nasa.gov/data/get-data/monthly-mass-grids-land/) distributes monthly GRACE TWS anomaly products on 1° horizontal-resolution grids (Landerer and Swenson 2012). While these gridded products are not necessarily optimal for all regions, they offer significant advantages of global coverage, public availability, and standardization. As a result, many in the hydrology community depend on these gridded products for their analyses. The gridded nature of the products is also an obvious advantage for DA systems.

In this study, we use the Tellus GRACE monthly mass grids optimized for land applications covering the period from January 2003 to January 2013. This product is based on the Release-05 (RL05) spherical harmonics fields produced by the University of Texas Center for Space Research (CSR), JPL, and German Research Centre for Geosciences (GFZ). We use the CSR version, but the products are generally similar. A number of filtering procedures (Swenson and Wahr 2006; Wahr et al. 2006) are applied to reduce measurement errors and to convert the data from the spectral domain to geographical coordinates. Tellus also provides a grid of multiplicative scaling coefficients aimed at restoring some of the signal loss due to filtering and truncation of the original GRACE spherical harmonics used in the derivation of GRACE TWS observations (Landerer and Swenson 2012). We convert the GRACE TWS anomalies to a ''total TWS'' in model space by adding the 2003–13 mean TWS from an open-loop (OL; no DA) CLSM integration to the monthly GRACE TWS anomalies on a gridcell-by-gridcell basis. These processed observations are then employed in the DA system.

b. CLSM

CLSM (Koster et al. 2000) is the terrestrial component of the atmospheric data assimilation and forecasting system at the NASA Global Modeling and Assimilation Office (GMAO). The basic modeling unit in CLSM is the intersection of a topographically based hydrologic catchment (or watershed) with gridded meteorological forcing. Unlike the layer-based approach used in most LSMs, CLSM simulates subsurface water storage using three prognostic bulk moisture variables: surface excess and root zone excess, which represent the excess or deficit soil moisture relative to equilibrium conditions for the top 2 cm and 1 m of soils, respectively, and catchment deficit that represents the amount of moisture that is required to bring the subsurface to

saturation. The state of these prognostics is used to determine the fraction of each modeling unit that is saturated, the fraction that is unsaturated but transpiring, and the fraction that is wilting. This allows for subgrid-scale representation of the influence that saturation state has on energy partitioning. Soil moisture in a 2-cm surface layer and a 1-m root zone layer are diagnosed from the prognostic bulk moisture variables. Groundwater is not directly modeled, but the equilibrium vertical distribution of soil moisture includes an implicit water table, located at the depth of the equilibrium saturation. This model feature is critical to GRACE DA, since it implies the presence of an unconfined aquifer that has time-varying storage (mass) of water. The CLSM configuration used in this article employs the soil depth to bedrock dataset used in the second Global Soil Wetness Project (GSWP-2). Similar to the strategy used in Houborg et al. (2012) and Li et al. (2012), the depth to bedrock used in the model was uniformly increased by 2 m, to increase the dynamic range of catchment deficit and thus avoid artificial limits on drydown during droughts. A three-layer snow model (Stieglitz et al. 2001) is used in CLSM to simulate snow conditions on the land surface. The simulated terrestrial water storage is obtained by subtracting the catchment deficit from the maximum available pore space of the catchment and by adding the surface and root zone excess terms. This total storage thus includes both the root zone soil moisture and the groundwater. The TWS is the sum of the total soil water storage, snow water equivalent, and canopy water storage; the groundwater storage is calculated by subtracting the root zone soil moisture, snow water equivalent, and canopy water storage from the TWS. Note also that in this study, CLSM is run on the regular latitude–longitude NLDAS grid, instead of on irregular catchments. All simulations use the CLSM Fortuna 2.5 version.

c. Model configuration

A domain configuration similar to the one used in the NLDAS (Mitchell et al. 2004; Xia et al. 2012b) is employed in this study. The NLDAS project employs a 1/8° regular latitude–longitude grid centered over the CONUS (25°–53°N, 125°–67°W). The model and DA integrations are conducted using the NASA Land Information System (LIS; Kumar et al. 2006; Peters-Lidard et al. 2007). The model simulations are forced with NLDAS-2 meteorological forcing data (Xia et al. 2012a), which includes daily gauge-based precipitation analysis, bias-corrected shortwave radiation, and surface meteorology analysis. The simulations are run with a 15-min time step and the model is spun up by running from 1979 to 2012 twice and then reinitializing the model in 1979 [following the advice of Rodell et al. (2005)]. Since the GRACE data are more reliable from 2003 onward, all evaluations are conducted during a time period of 2003–12.

Routed streamflow estimates are generated using the Hydrological Modeling and Analysis Platform (HyMAP; Getirana et al. 2012), which computes streamflow estimates by considering water surface dynamics and time delays in the surface runoff and baseflow fields from the LSM. HyMAP also includes formulations to model the interaction between rivers and floodplains, floodplain water flow among grid cells, and evaporation from open water and is run at the same spatial and temporal resolutions as that of the LSM. Note that other surface water sources such as lakes and wetlands are not modeled here.

d. Data assimilation method

Similar to prior GRACE DA studies, we employ a three-dimensional ensemble smoother algorithm for assimilating the GRACE data, which is described in detail in Zaitchik et al. (2008). In contrast to traditional DA scenarios where observations are instantaneous, GRACE observations are time-averaged TWS anomalies and are reported at a coarse temporal interval of a month. The smoothing approach temporally disaggregates the observations for each month into a finer, daily temporal scale of the land surface model.

The algorithm alternates between an ensemble forecast step and a data assimilation step. In the forecast step, an ensemble of model states is propagated forward in time using the LSM. In the update step (at time k), the model forecast is adjusted toward the observation based on the relative uncertainties, with appropriate weights expressed in the Kalman gain $\mathbf{K}_k$:

$$\mathbf{X}_k^{i+} - \mathbf{X}_k^{i-} = \mathbf{K}_k(\mathbf{Y}_k - \mathbf{H}_k\mathbf{X}_k^{i-}). \quad (1)$$

The state and observation vectors (suitably perturbed; Burgers et al. 1998) are represented by $\mathbf{X}_k$ and $\mathbf{Y}_k$, respectively. The linear observation operator $\mathbf{H}_k$ relates the model states to the observed variable. The superscripts $i-$ and $i+$ refer to the state estimates of the ith ensemble member before ($-$) and after ($+$) the update, respectively.

The sequence of the forecast and update steps is shown in Fig. 1. As noted above, the GRACE TWS anomalies are converted to absolute TWS values by adding the corresponding time-averaged TWS from the model open-loop simulation, and the 1° GRACE TWS observations are then interpolated to the 1/8° model resolution. During the assimilation integration, the model is propagated forward for a month without any assimilation (forecast step in Fig. 1). During this run, the

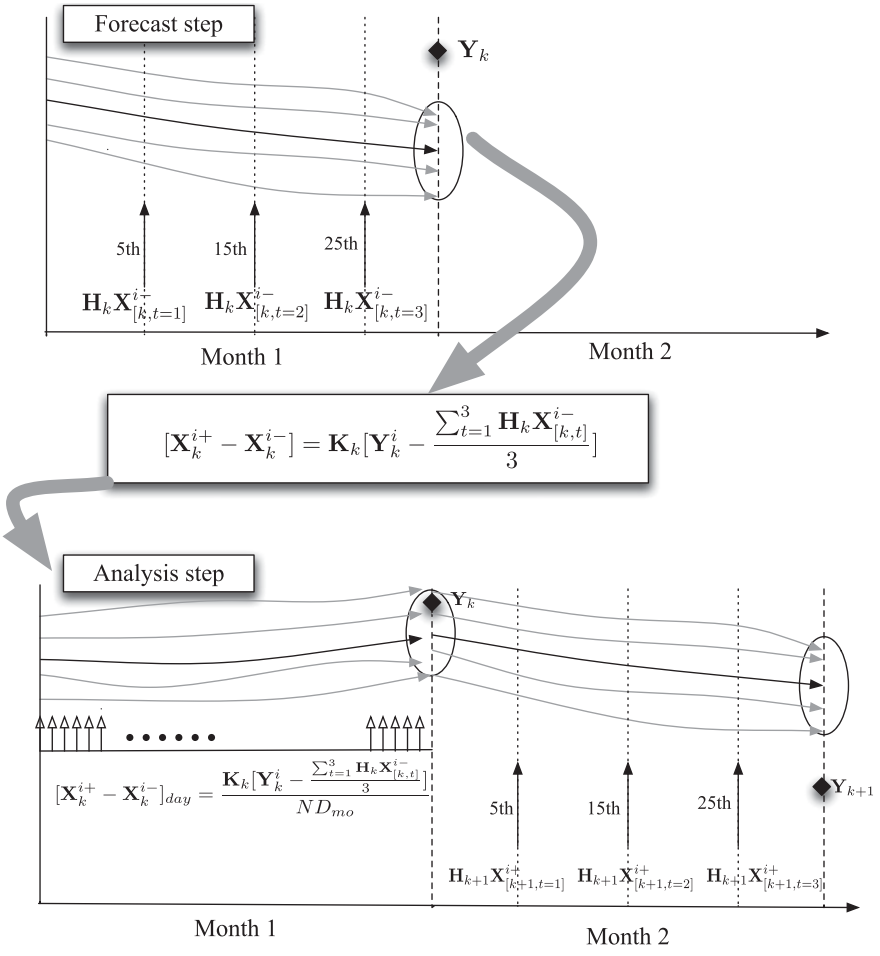

FIG. 1. Schematic of the sequence of the forecast and update steps of the ensemble smoother algorithm used for GRACE DA. During the forecast step, the observation operator stores the TWS values at the fifth, fifteenth, and twenty-fifth days of a month, for computing the analysis increments. During the analysis step, the ensemble is reinitialized to the beginning of the month and the DA increments are applied smoothly over each day.

TWS values at the fifth, fifteenth, and twenty-fifth days of the month are stored in memory for each grid point. The average of these three values is then used to compute the monthly model forecast of the GRACE TWS observation [i.e., $\mathbf{H}_k\mathbf{X}^{i-}_k$ in Eq. (1)], which completes the forecast step. The linear observation operator only performs a temporal aggregation, which is unlike the approach in Forman et al. (2012), where observations are kept at their coarse resolution and the observation operator also includes a spatial aggregation. This averaging approach is used to mimic the fact that GRACE monthly TWS estimates are informed by roughly three close overpasses each month. A true representation of GRACE views would require that we apply an inversion of GRACE product calculations to model predictions and perform the assimilation update in raw GRACE observation space, which imposes significant constraints on the modeling domain. Therefore, the current approximation is employed, consistent with prior GRACE DA studies. Next, the analysis increments $(\mathbf{X}^{i+}_k - \mathbf{X}^{i-}_k)$ for the beginning of the month are computed using Eq. (1).

The ensemble is then reinitialized at the beginning of the month, and during the second iteration, the analysis increments computed in Eq. (1) are applied evenly over each day of the month through dividing $(\mathbf{X}^{i+}_k - \mathbf{X}^{i-}_k)$ by the number of days in the month ($\mathrm{ND_{mo}}$). When the model integration for the current month is finished, the algorithm switches back to the forecast mode, storing TWS states from the three days in memory and repeating the above process.

The ensemble is generated by applying perturbations to the model states and input meteorological forcings. The parameters used for these perturbations are listed in Table 1. Zero-mean, normally distributed additive perturbations are applied to the downward longwave (LW)

TABLE 1. Parameters for perturbations to meteorological forcings and model prognostic variables in the assimilation experiments.

Variable	Perturbation type	Std dev	Cross correlations with perturbations in SW↓	LW↓	P		
Meteorological forcings							
Downward shortwave (SW↓)	Multiplicative	0.3	1.0	−0.5	−0.8		
Downward longwave (LW↓)	Additive	50 W m^{-2}	−0.5	1.0	0.5		
P	Multiplicative	0.50	−0.8	0.5	1.0		
CLSM TWS states			v_1	v_2	v_3	v_4	v_5
Catchment deficit (v_1)	Additive	0.05 mm	1.0	0.0	0.0	0.0	0.0
Surface excess (v_2)	Additive	0.02 mm	0.0	1.0	0.0	0.0	0.0
Snow water equivalent layer 1 (v_3)	Multiplicative	8×10^{-4}	0.0	0.0	1.0	0.9	0.8
Snow water equivalent layer 2 (v_4)	Multiplicative	8×10^{-4}	0.0	0.0	0.9	1.0	0.9
Snow water equivalent layer 3 (v_5)	Multiplicative	8×10^{-4}	0.0	0.0	0.8	0.9	1.0

radiation forcing, and lognormal multiplicative perturbations with a mean value of 1 are applied to the precipitation P and downward shortwave (SW) radiation fields. Time series correlations are imposed via a first-order autoregressive model [AR(1)] with a time scale of 24 h. The CLSM prognostic variables catchment deficit, surface excess, and the snow water equivalent for each of the three snow layers are perturbed with additive noise using the perturbation noise specified in Table 1. In model state perturbation instances, a horizontal error correlation of approximately 200 km was assumed. Furthermore, cross correlations are imposed on the variables (Reichle et al. 2007) as listed in Table 1. All model integrations use an ensemble size of 20. Quality-control checks on the physical limits of each variable are imposed to avoid possible unphysical values resulting from perturbations.

As noted earlier, an important difference in this study compared to previous GRACE DA studies is that here we use gridcell-by-gridcell assimilation of gridded GRACE hydrology products, whereas the prior studies employ river-basin-level assimilation. The use of gridcell-by-gridcell assimilation of GRACE presents certain challenges, since the 1° horizontal resolution of gridded GRACE products does not represent the true resolution of the data. The gridded products are derived from spherical harmonic representations of Earth's gravity field, and a Gaussian filter is applied during data processing to remove noise from high-degree Stokes coefficients (Wahr et al. 1998). This Gaussian smoothing means that both TWS anomaly estimates and errors are horizontally correlated. To mitigate sampling errors, a compact support function with a 300-km-radius is used to suppress spurious ensemble-derived correlations at larger separation distances (Reichle and Koster 2003; Gaspari and Cohn 1999).

Observation error estimates are another challenge for gridded GRACE data. Previous assimilation studies have used uniform observation error standard deviation of 20 mm for midlatitude studies, following Wahr et al. (2006). However, the gridded products are now provided with spatially distributed, temporally static error estimates (Landerer and Swenson 2012) that can be used in place of the uniform value. Figure 2 shows a map of the reported total measurement error of the GRACE data for the CONUS domain, interpolated to the NLDAS-2 resolution. An analysis of the spatial distribution of the error values in Fig. 2 indicates that over 80% of the domain, the total measurement error ranges between 0 and 40 mm. Higher errors are reported on the West Coast and in Florida, the lower Mississippi River, and parts of the southeastern United States. We performed simulations with distributed error estimates and with the uniform 20-mm value to match the previous studies.

Additional filtering procedures for the purposes of reducing the level of noise are applied to generate the gridded GRACE TWS data. However, these also lead to loss of signal, which becomes the dominant term in the error budget of the filtered data in some areas. To reduce the differences between the signal amplitudes of the original and filtered data, a multiplicative scale factor was developed (Landerer and Swenson 2012) and is distributed with the gridded GRACE TWS estimates. The main purpose of the scale factor is to restore signal amplitude lost during data processing, but because of the higher resolution of the hydrological model used to compute the scale factors on a 1° grid, they produce finescale features that are not present in the unscaled data. We apply the scaling factors prior to assimilating GRACE observations, but we also performed a simulation without scaling factors as a sensitivity test.

3. Results and discussion

This section first presents a quantitative assessment of the impact of assimilating GRACE TWS (DA-TWS) on various water budget components, including groundwater,

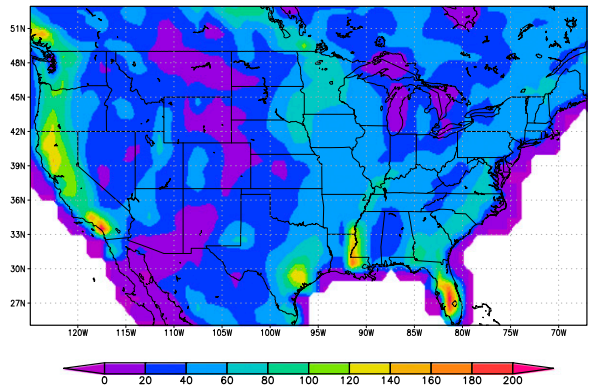

FIG. 2. Map of total measurement errors (mm) of GRACE TWS estimates.

soil moisture, streamflow, snow depth, and evapotranspiration (ET). The model simulations are evaluated by comparing against a number of independent datasets using the LVT (Kumar et al. 2012). The LVT is an integrated environment developed for the systematic evaluation of land model analyses, and it facilitates comparisons against a large suite of in situ, remotely sensed, and other model and reanalysis datasets using a variety of metrics.

The default GRACE DA configuration (DA-TWS) used in this section employs the distributed measurement errors and the scaling factors. Section 3h explores the sensitivity of the results to the use of the distributed error estimates and the scaling factors that are provided with the GRACE data, and section 3i provides a comparison of the DA approaches that employ gridded and basin-averaged GRACE data. The last part of this section focuses on the impact of assimilating TWS retrievals on drought estimates.

a. *Terrestrial water storage*

Figure 3 presents a time series of the daily averaged TWS from the OL and DA-TWS and the monthly GRACE TWS for six regions of the CONUS. The seasonality of the TWS is captured reasonably well in the OL integration, and GRACE DA leads to changes in the interannual variability of TWS estimates. In general, TWS in the DA simulation falls between that of OL and the GRACE observations, as one would expect. It is possible, however, for TWS in DA-TWS to fall outside of this range because of the long memory of TWS in the model. In addition, it is likely that the three-dimensional filter updates also contribute to this behavior. For example, if a neighboring innovation is large and has a sign opposite to and an estimated error less than that of the grid cell in question, the analysis may result in an update away from the observations at that grid cell. As seen in Fig. 3, the influence of DA is more impactful in the later years and over the Northeast, Midwest, Great Plains, and Southwest regions. In these regions, changes as large as 6%–9% in area-averaged daily mean TWS are observed. Larger differences between DA-TWS and OL are observed at the gridcell scales (not shown), indicating the significant influence of GRACE DA toward impacting the column-integrated water storage estimates. The influence of GRACE DA is encouraging, as it suggests that the unique information in GRACE observations can inform CLSM simulations even when the model uses high-quality NLDAS meteorological forcing data. The significant influence of DA on a simulation that already uses best-available meteorology and parameter data demonstrates that neither the model nor the inputs are perfect, and it is important to perform a thorough evaluation of how DA influences model performance judged against available evaluation data.

b. *Groundwater*

In situ groundwater level measurements from 128 monitoring wells were used for evaluation of simulated groundwater storage variations. We selected wells for inclusion that we determined to be installed in unconfined aquifers and not directly impacted by pumping or injections, based on an analysis of available metadata, published reports, and the temporal dynamics of the individual time series. Many of these wells were previously identified using similar criteria by Rodell et al. (2007) and Houborg et al. (2012). The seasonal climatology of each time series was removed before applying it for evaluation. We note that some monitoring wells are in coastal locations. These wells could be influenced by ocean level changes that are not represented in CLSM, but we did not find systematic coastal effects in our model evaluation.

Figure 4 shows changes in groundwater anomaly correlation R fields in the GRACE DA compared to the open-loop simulation. The anomaly R values at each grid point are computed by subtracting the monthly mean climatology values from the daily average raw data, so that they represent the deviations from the mean seasonal cycle. The anomaly R values are then calculated by comparing the anomalies of daily model water-table estimates and in situ groundwater level measurements from 128 monitoring wells across the United States. Here, we use anomaly R as the metric to exclude the skill contribution from correctly identifying the mean seasonal variation. Only grid points with at least 300 valid in situ measurements are chosen in the analysis. The statistical significance of the anomaly R differences is computed based on the Fisher's Z transform. As seen from Fig. 4, station coverage is sparse after the quality-control procedures employed here,

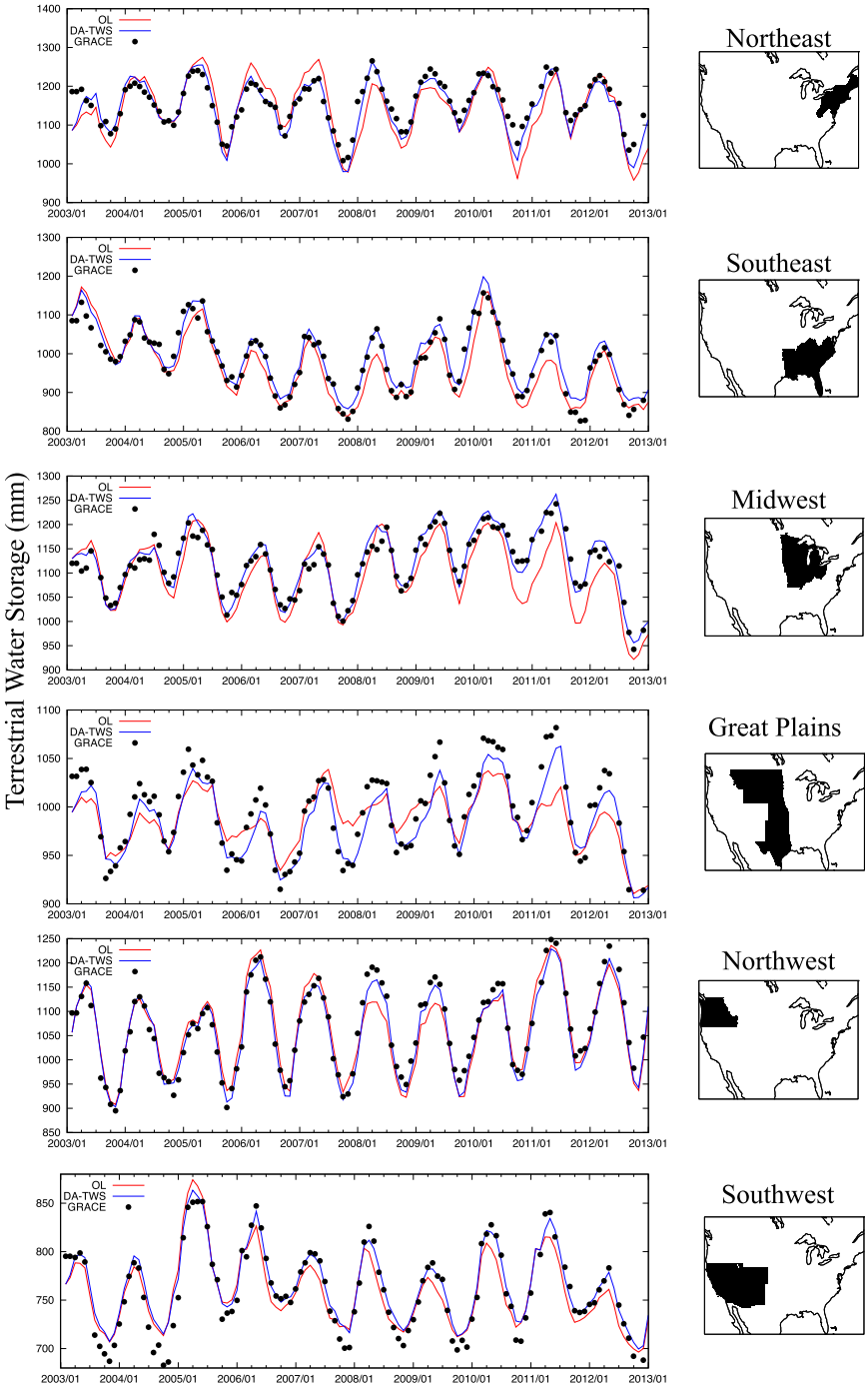

FIG. 3. A time series comparison of TWS estimates from the OL, DA-TWS, and GRACE observations over six major regions of the CONUS.

particularly in the western United States. In Fig. 4, the locations where the anomaly R differences are not statistically significant at the 95% confidence intervals are shown in gray shading.

There is an overall marginal improvement in the simulated groundwater estimates as a result of GRACE DA. The domain-averaged anomaly R from the OL simulation is 0.67, and it improves to a value of 0.69, which is barely statistically significant based on the 95% confidence intervals, with the assimilation of GRACE data. Figure 4 indicates that there are systematic improvements in the upper Mississippi and parts of the

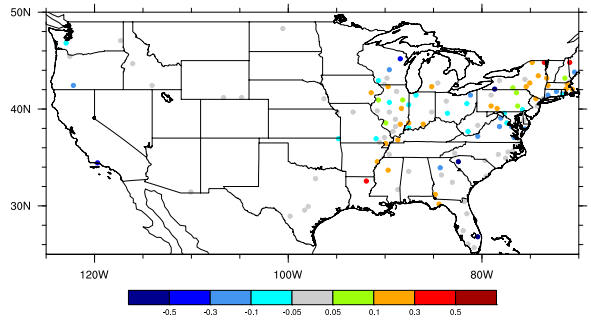

FIG. 4. Anomaly R differences of groundwater fields from DA-TWS relative to the OL integration. The warm colors indicate locations of improvement and cool colors indicate locations of degradation. The gray shading indicates locations where the anomaly R differences are not statistically significant.

Northeast, whereas the improvements are small in the southeastern United States and results are mixed for the few available sites in the western United States.

It is encouraging that DA-TWS improvements are observed in regions with relatively dense in situ data records (i.e., the Northeast and Midwest). However, the scarcity of quality evaluation sites in the western half of the country is a problem, because information on groundwater dynamics is likely to be most relevant in parts of the country that are from semiarid to arid and/or depend on groundwater for irrigation. Data available in the analysis presented in this paper are not adequate to support any conclusion on the value of GRACE DA in the western United States. Note also that CLSM uses a simple aquifer formulation to represent groundwater variations and is unlikely to accurately capture the groundwater dynamics in the central and western U.S. regions where the water table tends to be tens of meters deep. Use of more advanced groundwater models representative of such regions and synthetic DA studies similar to Forman and Reichle (2013) are needed to enhance the utility of GRACE DA in these regions. Such a study is left for a future work.

c. *Soil moisture*

The improvements in simulated soil moisture fields are evaluated by comparing against two reference datasets: 1) soil profile measurements from the USDA Soil Climate Analysis Network (SCAN; Schaefer et al. 2007) and 2) surface soil moisture measurements from four USDA Agricultural Research Service (ARS) experimental watersheds (Jackson et al. 2010). The quality-controlled version of the raw SCAN data (De Lannoy et al. 2014) is used to evaluate both surface and root zone soil moisture fields, as it provides profile measurements of soil moisture at depths of 5, 10, 20, 50, and 100 cm wherever possible. ARS data only include the surface soil moisture measurements, and the area-averaged estimates from individual sensor measurements at each ARS watershed are used in this study. Similar to Kumar et al. (2014), the root zone soil moisture is defined as the soil moisture content of the top 1 m of the soil column. The model root zone soil moisture is diagnosed from the bulk moisture variables, whereas the observation root zone soil moisture is computed as a suitably weighted vertical average over the observation layers.

Figure 5 maps the differences in anomaly R for the daily surface and root zone soil moisture fields [anomaly R(DA) minus anomaly R(OL)], from comparisons against both SCAN and ARS measurements. Similar to Fig. 4, locations shown in gray shading indicate that the anomaly R differences are not statistically significant (at the 95% confidence level) based on the Fisher's Z transform. Compared to SCAN data, there are statistically significant improvements in surface soil moisture anomaly R across many stations in the domain (Fig. 5a). The domain-averaged anomaly R increases from 0.57 for OL to 0.60 for the DA-TWS integration. Similar improvements are observed in the root zone soil moisture estimates, with a domain-averaged skill of 0.60 for OL improving to 0.63 with the assimilation of GRACE data. While GRACE DA improves correlation with observations at the majority of stations, especially in the eastern United States, degradations in the root zone soil moisture skill are observed at a few stations in the western United States and lower Mississippi River (Fig. 5b). The domain-averaged anomaly RMSE (computed as the RMSE of anomalies after removing the mean seasonal cycle) for the surface and root zone soil moisture for the OL are 0.047 and 0.035 $m^3 m^{-3}$, respectively, and they reduce to 0.045 and 0.034 $m^3 m^{-3}$ with DA-TWS. In this case, the domain-averaged improvement in the root zone soil moisture anomaly RMSE is not statistically significant. The comparison to the data from the four ARS stations also confirms these trends of improvements from DA-TWS. With DA, the anomaly R improves from 0.72 to 0.74 and the anomaly RMSE improves from 0.029 to 0.027 $m^3 m^{-3}$.

d. *Streamflow*

Here we use the two reference datasets used in Kumar et al. (2014) to evaluate the simulated streamflow estimates: 1) daily streamflow data from 2003 to 2012 obtained from the U.S. Geological Survey (USGS; http://nwis.waterdata.usgs.gov/nwis) over 572 small, unregulated basins and 2) monthly "naturalized" streamflow data at 19 major basin outlets (Mahanama et al. 2012), developed by removing water management effects. Similar to the approach used in Kumar et al. (2014), we use a normalized information contribution (NIC)

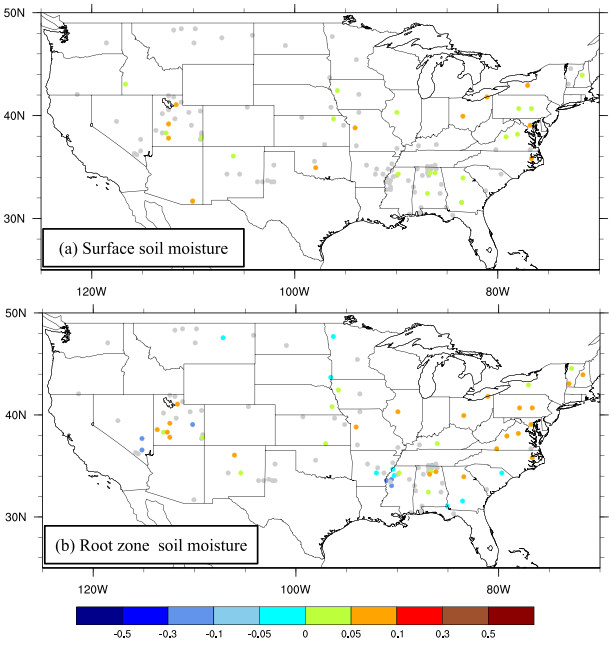

FIG. 5. Anomaly R differences of (a) surface and (b) root zone soil moisture fields from DA-TWS relative to the OL integration. Warm colors indicate locations of improvement and cool colors indicate locations of degradation.

metric to quantify the improvements and degradations in the simulated streamflow estimates from DA. This normalized approach is used because the magnitude of streamflow varies significantly across different basins. The NICs for RMSE, R, and Nash–Sutcliffe efficiency (NSE) are defined as follows:

$$\mathrm{NIC}_{\mathrm{RMSE}} = \frac{(\mathrm{RMSE}_o - \mathrm{RMSE}_a)}{\mathrm{RMSE}_o}, \tag{2}$$

$$\mathrm{NIC}_R = \frac{(R_a - R_o)}{(1 - R_o)}, \quad \text{and} \tag{3}$$

$$\mathrm{NIC}_{\mathrm{NSE}} = \frac{(\mathrm{NSE}_a - \mathrm{NSE}_o)}{(1 - \mathrm{NSE}_o)}, \tag{4}$$

where the subscripts o and a denote open loop and assimilation, respectively. Each NIC metric measures how much of the maximum possible skill improvement [which, in the case of R, is $(1 - R_o)$] is realized through DA [which, in the case of R, is $(R_a - R_o)$]. The sign of the NIC metric indicates if the assimilation leads to an improvement or degradation over the open loop, with positive and negative NIC values indicating improvements and degradations from data assimilation, respectively. For NIC = 0, the assimilation does not add any skill and for NIC = 1, the assimilation realizes the maximum possible skill improvement.

Figure 6 presents the maps of $\mathrm{NIC}_{\mathrm{RMSE}}$, NIC_R, and $\mathrm{NIC}_{\mathrm{NSE}}$ and their distribution across the basins. As indicated by the histograms of NIC metrics, the overall changes in streamflow due to DA-TWS are small, as most of the NIC values range from −0.05 to 0.05. Regionally, there are some improvements in the streamflow estimates (especially in the R comparisons) over parts of the upper and lower Mississippi basins. Some minor degradations are observed over parts of the California basin in the western United States.

The streamflow estimates were also evaluated at several large basin outlets where the modeled streamflow is compared against "naturalized" streamflow data (with water management effects removed), similar to the evaluations in Mahanama et al. (2012). Table 2 lists the details of the major basins examined in this study, and Fig. 7 presents a quantitative comparison of the influence of GRACE DA. The impact of GRACE DA is generally mixed, with some notable improvements over the upper Mississippi and Ohio Rivers and degradations at Garrison and Ft. Randall in terms of improving the magnitude of simulated streamflow. It must be noted that the absolute magnitude of discharge is much higher at the Ohio and the upper Mississippi Rivers, which are the two outlets where overall improvements of approximately 7% and 4% in RMSE are observed. The correlations are marginally improved in most basins, except at the upper Mississippi River, the Arkansas River near Ralston, Rio Puerco near Bernado, and Colorado River at Lees Ferry. It must also be noted that R values are quite low (<0.4) for both OL and DA-TWS in a number of these rivers, suggesting that the modeling system used in this study does not perform well in terms of capturing the discharge patterns of these rivers. In these situations, it is likely that some form of land surface model calibration or parameter optimization is required to achieve meaningful improvements in the representation of discharge.

e. *Snow depth*

The impact of GRACE assimilation on simulated snow depth is evaluated using three reference datasets: 1) daily in situ snow depth measurements from the Global Historical Climatology Network (GHCN; Menne et al. 2012) meteorological station network, 2) the spatially distributed daily snow depth analysis from the Canadian Meteorological Centre (CMC; Brown and Brasnett 2010), and 3) the NOAA National Weather Service's National Operational Hydrologic Remote Sensing Center (NOHRSC) Snow Data Assimilation System (SNODAS; Barrett 2003) outputs. The GHCN stations are chosen only if they report at least 3 months of valid data during the winter season (December–April) and if they report at least 2 years of data during

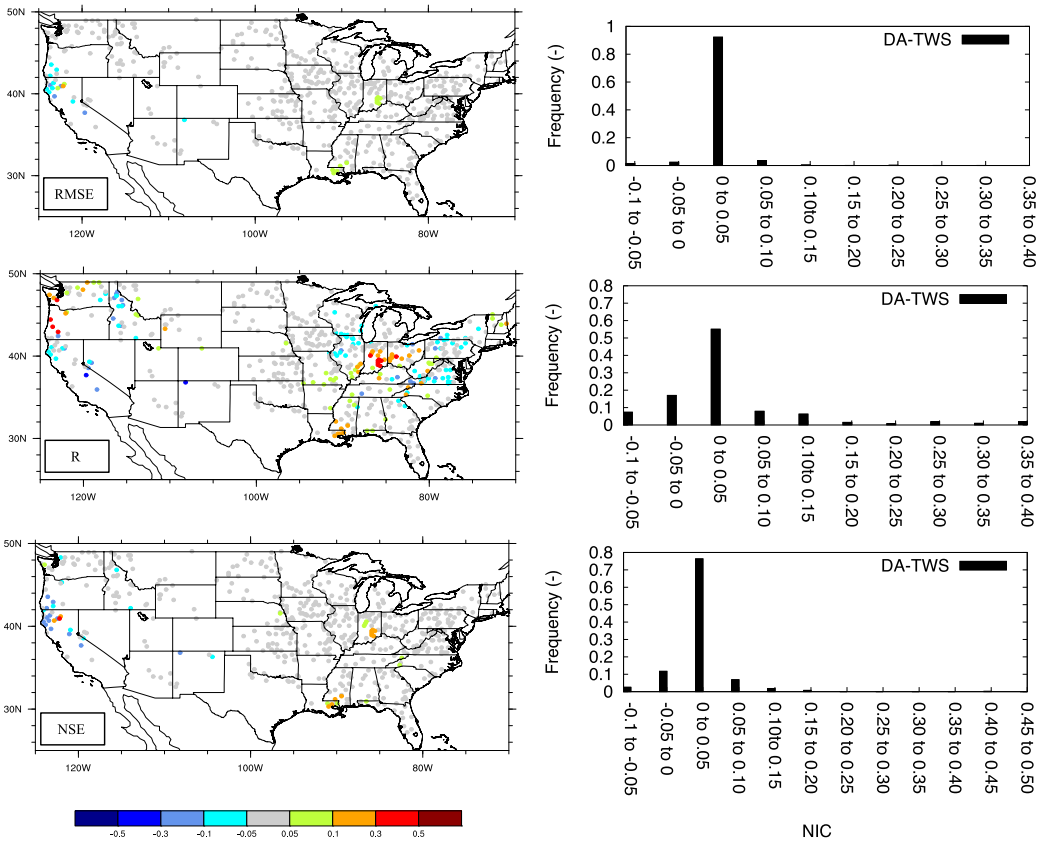

FIG. 6. (left) The streamflow NIC values for (top) RMSE, (middle) *R*, and (bottom) NSE from DA-TWS. The warm and cool colors represent areas of improvements and degradations, respectively. (right) The distribution of NIC values for each metric across the domain.

the 2003–13 time period. The CMC analysis is available daily at approximately 25-km spatial resolution globally from March 1998 onward. SNODAS data products are generated at 1-km spatial resolution beginning in October 2003 and at hourly temporal resolution over the CONUS. The SNODAS analyses are generated by combining the estimates from a snow model with satellite-derived, airborne, and ground-based observations of snow from surface synoptic observations, meteorological aviation reports, and special aviation reports acquired from the World Meteorological Organization (WMO). CMC analyses are also generated in a similar manner by combining the forecasts from a temperature index snow model with observed snow depth values (Brasnett 1999).

Figure 8 presents maps of RMSE differences compared to the above-mentioned three snow depth datasets. Similar to the comparisons in the previous sections, the RMSE differences are computed by subtracting RMSE of the DA integration from the RMSE of the open-loop simulation. In the figure, locations are shaded gray if the RMSE differences are not statistically significant at the 95% confidence level as indicated by the Student's *t* test.

Figure 8 indicates that the influence of GRACE DA on snow depth estimates is small, as changes in RMSE of snow depth fields are not statistically significant in most parts of the domain. The only regions where the RMSE differences are statistically significant are over parts of the Rocky Mountains, Sierra Nevada, and Cascades, which are all regions characterized by significant seasonal snowpack. Comparatively, the evaluation against GHCN data indicates larger improvements from GRACE DA whereas the changes in RMSE from GRACE DA in the CMC and SNODAS comparisons are very small. All three comparisons consistently indicate that DA-TWS provides improvements over these regions. Note that these improvements are still small, in the range of ±10 mm, when averaged across the entire simulation period.

f. Evapotranspiration

To quantify the impact of GRACE DA on ET simulation, we compare the model outputs against four ET

TABLE 2. List of the major basins.

Station No.	Station acronym	River name	Lat (N)	Lon (W)	Basin area (km^2)
1	PUE	Rio Puerco near Bernado	34.41°	106.85°	19 036
2	RAL	Arkansas River near Ralston	36.50°	98.73°	141 064
3	LEE	Colorado River at Lees Ferry	36.87°	111.58°	289 562
4	OHI	Ohio River at Metropolis	37.15°	88.74°	525 770
5	UPM	Upper Mississippi near Grafton	38.90°	90.30°	443 660
6	GUN	Gunnison River near Grand Junction	38.98°	108.45°	20 533
7	SBB	Sacramento River near Bend Bridge	40.29°	122.19°	23 051
8	GRE	Green River near Greendale	40.91°	109.42°	50 116
9	RAN	Missouri River at Ft. Randall Dam	43.07°	98.55°	682 465
10	MUS	Musselshell River near Moseby	46.99°	107.89°	20 321
11	GAR	Garrison Reservoir (Missouri River)	47.39°	101.39°	469 826
12	FTP	Missouri River at Fort Peck Reservoir	48.04°	106.36°	149 070

products: 1) the Atmosphere–Land Exchange Inverse (ALEXI; Anderson et al. 2007) model that uses time-differential land surface temperature data retrieved from geostationary satellite thermal band imagery to estimate ET via surface energy balance (available daily at 4-km spatial resolution), 2) gridded FLUXNET data (Jung et al. 2009) created by synthesizing FLUXNET tower data with meteorological forcings and vegetation information from interpolated station and satellite data (available monthly at ½° spatial resolution), 3) Moderate Resolution Imaging Spectroradiometer (MODIS)-based ET product from University of Washington (UW; Tang et al. 2009; available monthly at 5-km spatial resolution), and 4) MODIS-based ET product employing the Penman–Monteith equation (MOD16; Mu et al. 2011; available daily at 1-km spatial resolution). It should be noted that none of these products represent direct observations of ET. Note that all these products have random errors and biases of their own, and they are included here for the purpose of comparison and should not be considered "truth."

Figure 9 shows the monthly RMSE difference maps [RMSE(OL) minus RMSE(DA)] based on comparisons with these four reference datasets, at locations that are statistically significant (at the 95% confidence interval) indicated by the Student's t test. Similar to the snow depth evaluation, the impact of GRACE DA on ET estimates is small. The comparisons against ALEXI, FLUXNET, and UW show certain consistent patterns. For example, there are areas of decreased RMSE in the western and southern United States and increased RMSE near the upper Mississippi basin in these three comparisons. The RMSE difference map using MOD16, on the other hand, generally show that GRACE DA increased RMSE in most parts of the domain. Prior studies (Peters-Lidard et al. 2011) have shown that MOD16 data generally tend to underestimate ET.

g. *Improvements across water cycle components*

It must be stressed that despite the use of multiple ancillary datasets to evaluate various water budget components, it is still difficult to provide a consistent and integrated assessment of the overall improvement in the water cycle components because the stations for

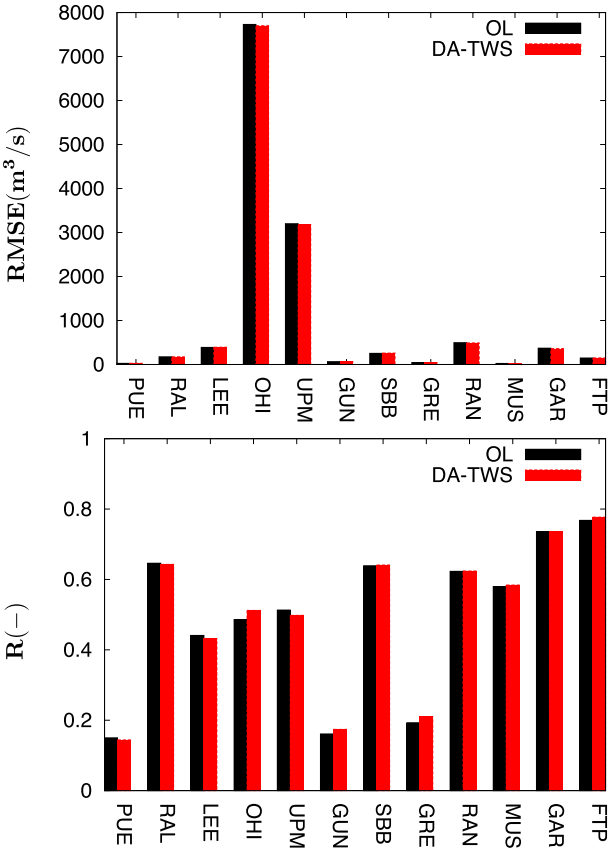

FIG. 7. A comparison of the RMSE and R of the streamflow estimates from the OL and DA-TWS at major basin outlets listed in Table 2.

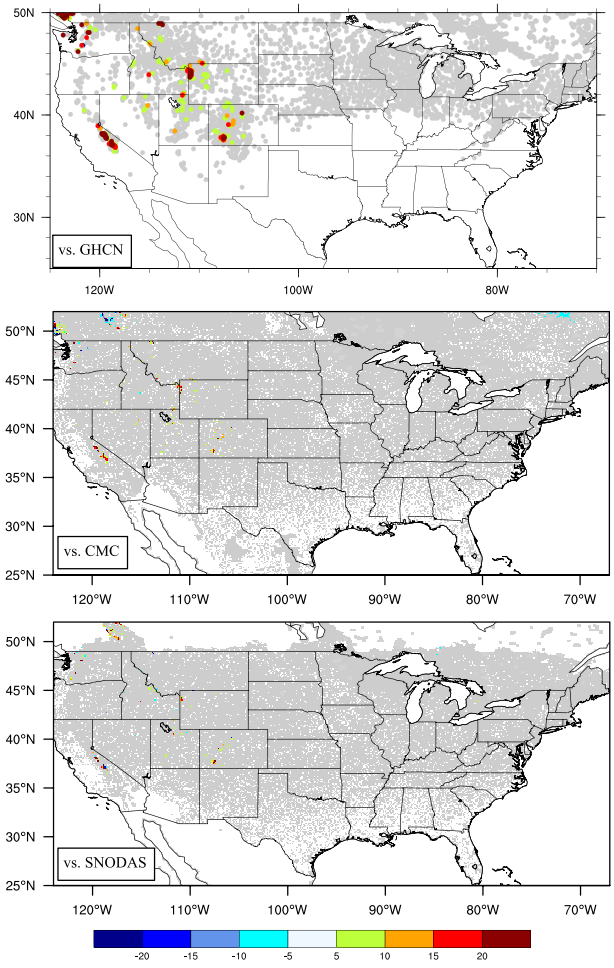

FIG. 8. RMSE differences (mm) of snow depth from DA-TWS relative to the OL integration, using in situ (top) GHCN, (middle) CMC, and (bottom) SNODAS products as reference datasets. The warm and cool colors show locations of improvements and degradations from GRACE DA, respectively. The gray shading indicates locations where the RMSE differences are not statistically significant.

different measurement types are not usually collocated. In the comparisons that involve in situ data, the representativeness errors are also likely to influence the results, given the differences in the spatial scales between a modeling grid and point-scale measurements. A few inconsistent patterns of improvements and degradations can be observed across the comparisons presented above. In particular, the groundwater and soil moisture comparisons indicate that there are improvements from GRACE DA in the upper Mississippi basin. The streamflow evaluation also indicates patterns of improvement over this region (best seen in the R comparisons in Fig. 6). The ET comparisons, however, show a general pattern of degradation, compared with ET estimates, in this region due to GRACE DA. Over the lower Mississippi basin, the groundwater and surface soil moisture improvements are generally positive whereas marginal degradations are observed in root zone soil moisture skills. Over this region, the streamflow estimates show improvements from GRACE DA, whereas ET estimates largely show degradations. Finally, over the mountainous western United States, there are improvements in soil moisture fields and snow depth, but little or no improvements in simulated streamflow and mixed results for ET. Nevertheless, the evaluation of the individual components of TWS shows generally encouraging trends of improvements from GRACE DA.

h. Influence of scaling factor and measurement errors

In this section, we investigate the influence of using the scale factors and the reported measurement errors in the DA integrations. Three DA configurations are compared here: 1) DA1, our baseline simulation, which uses the gridded GRACE TWS data with the scale factors and spatially distributed measurement errors (also known as DA-TWS in the previous subsections); 2) DA2, which uses the scale factors along with a spatially uniform measurement error with a Gaussian error variance of $20^2\,\text{mm}^2$; and 3) DA3, which uses a spatially uniform measurement error with a Gaussian error variance of $20^2\,\text{mm}^2$ but no scaling factors (Table 3). The comparison of DA2 and DA3 quantifies the impact of using the scaling factors, whereas the comparison of DA1 and DA3 quantifies the combined influence of using both the scaling factors and spatially distributed measurement errors.

Figure 10 presents a comparison of the three DA integrations. Figure 10a shows the domain-averaged monthly time series TWS from the three DA integrations and the OL simulation, for the whole CONUS domain. Generally, the influence of using the scaling factors and the measurement errors is small. As shown in Fig. 10a, the differences between the DA1 and DA2 integrations are smaller in early years compared to the differences of DA1 or DA2 relative to DA3. In later years, however, there are larger differences between DA1 and DA2. This suggests that in the CONUS domain averages, the influence of the use of the scaling factors is larger than that of the use of spatially distributed measurement errors in early years, whereas the influence of measurement errors becomes more significant in later years. Similar trends are observed in other water budget terms (not shown).

An independent quantitative evaluation of the three integrations is shown in Figs. 10b–g. Figures 10b and 10c present maps of differences in anomaly R of root zone soil moisture from DA1 and DA2 integrations relative to DA3, by comparing against the SCAN

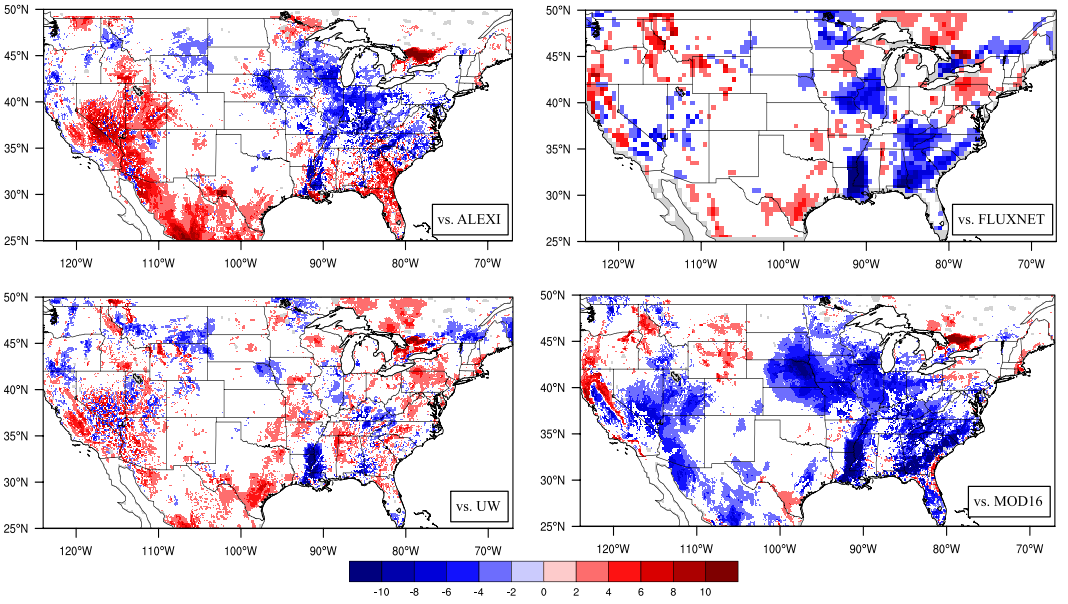

FIG. 9. RMSE differences ($W\,m^{-2}$) of ET from DA-TWS relative to the OL integration, using four reference datasets (ALEXI, gridded FLUXNET, UW, and MOD16). The warm colors represent the decreases in RMSE and cool colors represent increases in RMSE due to GRACE DA.

measurements. Similarly, Figs. 10d and 10e provide estimates of changes in groundwater anomaly R values (using the USGS station measurements as reference) in DA1 and DA2 relative to DA3. Finally, Figs. 10f and 10g present the NIC_{NSE} of streamflow estimates of DA1 and DA2 relative to DA3, using the same reference data employed in section 3d. Generally, the influence of the use of scaling factors is greater compared to the impact of using spatially distributed measurement errors on the skill improvements. The use of scaling factors improves the simulation of root zone soil moisture in parts of the western United States and the high plains, whereas the skills are marginally reduced in the eastern part of the modeling domain (Fig. 10c). Similar trends are seen in the groundwater comparisons where the added use of distributed measurement errors only improve the groundwater skills at a few stations (Fig. 10d) over that of the use of scaling factors (Fig. 10e). The use of these DA variants do not show significant impacts on the streamflow skills (Figs. 10f,g). At the domain-averaged scale, the differences in groundwater, root zone soil moisture, and streamflow skills of these DA variants are not statistically significant.

While benefits to the assimilated output of applying scale factors to the GRACE data were ambiguous in this case, it should be noted that the scale factors have previously been shown to restore signal lost because of GRACE data processing in certain regions, including the Pacific Northwest (Tang et al. 2010; Landerer and Swenson 2012). Hence, the results shown here should not discourage the use of the scale factors. Note also that the spatially distributed measurement errors (Fig. 2; used in DA1) are generally larger than the 20-mm error standard deviation assumption used in the DA3 simulation, especially over the eastern United States. Marginal degradations in the soil moisture and groundwater skills in the DA1 integration relative to DA3 in the eastern part of the domain suggests that use of lower error values (than the total measurement errors) may be more appropriate in these areas.

i. Comparison of DA strategies using gridded and basin-averaged GRACE data

As noted in the introduction, most prior GRACE DA studies employ preprocessed GRACE data averaged up to river basin scales. This section presents a comparison of using both gridded and basin-averaged GRACE data. The basin-averaged GRACE data are generated using the hydrological catchment delineations employed by Houborg et al. (2012). The DA integration using the basin-averaged GRACE data also employs scaling factors and spatially distributed measurement errors. The GRACE data are averaged up to the basin scales after applying the scaling factors. The basin-averaged GRACE DA integration (DA1b) is compared against our baseline simulation (DA1), which uses the gridded GRACE TWS data with the scale factors and spatially distributed measurement errors.

TABLE 3. Details of the DA variants to examine the impact of scaling factor, distributed measurement errors, and basin-level averaging.

Expt name	Scale factors used?	Observation error variance	Assimilated DA product
DA-TWS (DA1)	Yes	Spatially distributed (Fig. 2)	Gridded 1°
DA2	Yes	Spatially uniform (20^2 mm^2)	Gridded 1°
DA3	No	Spatially uniform (20^2 mm^2)	Gridded 1°
DA1b	Yes	Spatially distributed (Fig. 2)	Basin averaged

The comparison of the DA1b and DA1 integrations is presented in Fig. 11. Figure 11a shows the monthly, CONUS-averaged time series of TWS from DA1 and DA1b and indicates that the influence of both DA1 and DA1b are comparable. Similar to Fig. 10, an independent quantitative evaluation of the root zone soil moisture, groundwater, and streamflow fields from DA1 and DA1b is shown in Figs. 11b–d. The basin-averaged assimilation leads to marginally higher root zone soil moisture skills in many parts of the domain (Alabama, Mississippi, Utah, etc.). The comparison of groundwater fields, on the other hand, shows mixed results. For example, the DA1b integration has higher skills in parts of the upper Mississippi basin, whereas the DA1 integration performs better at locations such as Florida and the lower Mississippi. Finally, the DA1 and DA1b integrations show very small differences in the evaluation of the streamflow fields.

j. Evaluation of drought estimates

In this section, we quantify the potential of GRACE DA for improving drought estimation. Houborg et al. (2012) presented an objective methodology for developing GRACE-based drought indicators for the United States and Li et al. (2012) examined the use of GRACE DA for drought monitoring in western and central Europe. Here, we extend these earlier studies through quantitative comparisons against data from the USDM (Svoboda et al. 2002).

Estimates of drought are generated through percentile-based indices using TWS outputs from the OL and DA-TWS integrations. The percentile-based drought indices are computed in a manner similar to that used in the NLDAS drought monitoring system (Ek et al. 2011; Sheffield et al. 2012). The percentiles are computed as follows: using daily outputs from 34 years of OL model simulation (1979–2012), the TWS climatology is computed first, separately for each grid point. The climatology is generated by assembling the variable values for a particular calendar day across all 34 years, using a time window of 5 days to improve the sampling density. For example, 3 January climatology is assembled by using all values from 1 to 5 January, across all years, leading to $34 \times 5 = 170$ values for each calendar day. Once the climatology is assembled, the daily percentile values are computed by ranking each day's estimate against the climatology. Since the DA-TWS integration involves a shorter time period (10 years; 2003–12), compared to the OL, the TWS fields from DA-TWS run were scaled to match the OL climatology before computing the DA-TWS–based percentiles. The rescaling is performed using standard normal deviate-based approach, separately for each grid point, as shown in Eq. (5):

$$\theta_i' = \mu_i^o + (\theta_i - \mu_i^d)\frac{\sigma_i^o}{\sigma_i^d}, \tag{5}$$

where the subscript i denotes the grid point; θ_i is the TWS value from the DA-TWS run; θ_i' is the rescaled TWS value; μ_i^o and μ_i^d are the temporal TWS mean values from the OL and DA runs, respectively; and σ_i^o and σ_i^d are the standard deviations of TWS from the OL and DA runs, respectively. Once the TWS estimates from the DA run are rescaled, the percentiles from the DA run are generated by sampling from the OL climatology. The weekly drought percentage area values are produced for six different regions of the United States (South, Southeast, Northeast, Midwest, high plains, and West, as defined in the USDM) and for five drought categories of varying intensity: D_0 (abnormally dry, percentile $\leq$30%), D_1 (moderate drought, percentile $\leq$20%), D_2 (severe drought, percentile $\leq$10%), D_3 (extreme drought, percentile $\leq$5%), and D_4 (exceptional drought, percentile $\leq$2%).

Figure 12 presents a comparison of the spatial distribution of drought intensities for a number of representative cases in the years 2004, 2006, 2011, and 2012. The figure shows the drought percentiles from the OL and DA-TWS integration compared against the corresponding drought intensity map from the USDM archives (http://droughtmonitor.unl.edu/MapsAndData/MapArchive.aspx). In the 20–27 January 2004 case, the representation of the drought in the open-loop-based estimate differs significantly from USDM, as the OL simulation underestimates drought severity and fails to match the spatial locations of the affected areas. Though these limitations are not significantly improved in the DA-TWS–based estimates, over places such as Utah, Colorado, Nebraska, and Iowa, DA-TWS–based drought estimates are more closely aligned with the

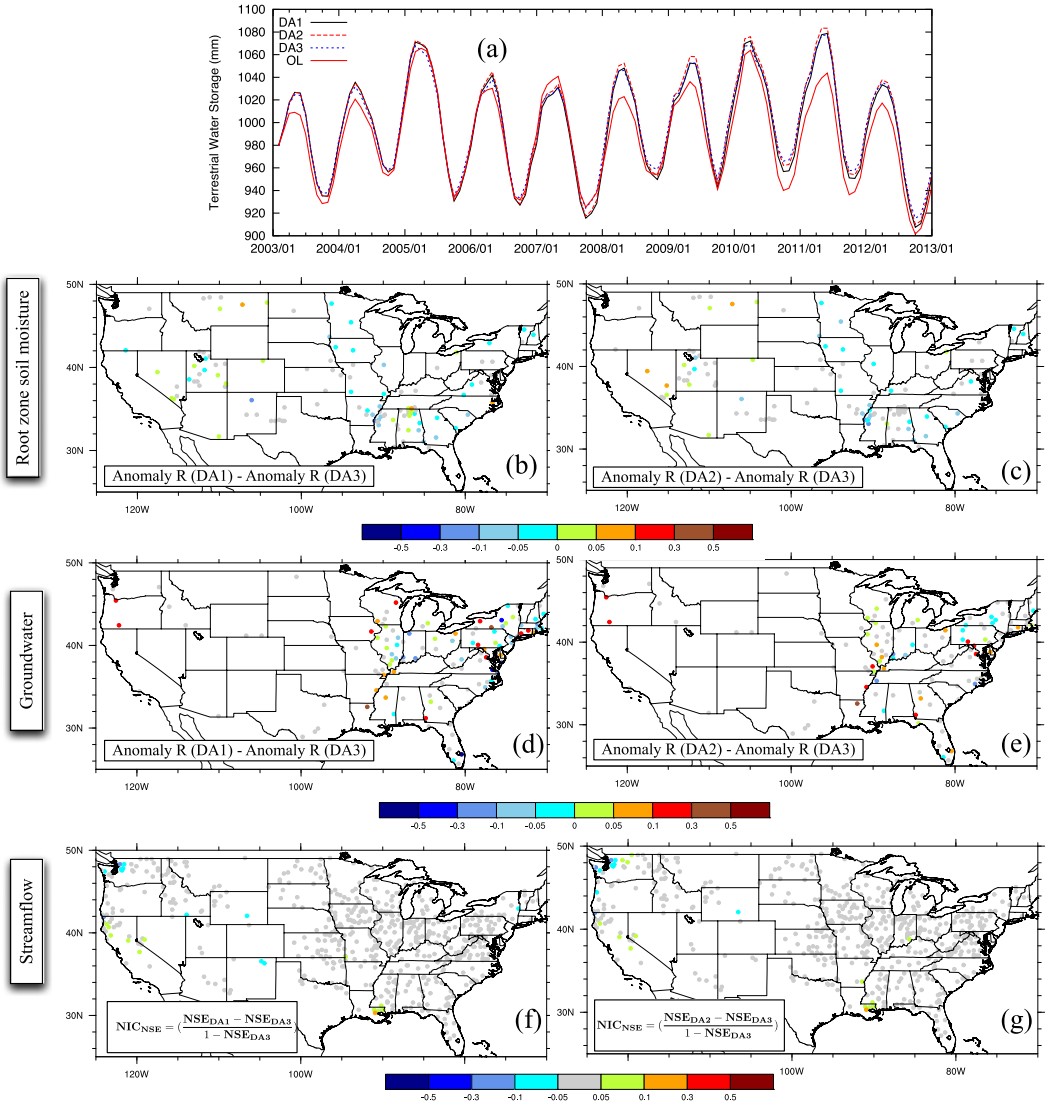

FIG. 10. A comparison of DA-TWS integrations that use no scaling factors and uniform measurement errors (DA3), that use scaling factors with uniform measurement errors (DA2), and that use scaling factors and spatially distributed measurement errors (DA1). (a) The time series of domain-averaged monthly TWS. (b),(c) The anomaly R differences of root zone soil moisture from DA1 and DA2 relative to DA3, respectively. (d),(e) The anomaly R differences of groundwater from DA1 and DA2 relative to DA3, respectively. (f),(g) The NIC_{NSE} of streamflow estimates from DA1 and DA2 compared to DA3, respectively.

USDM. The severe drought over South Dakota, however, is better represented in the OL estimate. In the 1–7 February 2006 case, the OL-based estimate matches USDM in the location of the extreme drought categories (D_3 and D_4) over northeastern Texas and southeastern Oklahoma and southeastern Arizona. It underestimates the D_0–D_2 categories over the southern plains (Kansas and Nebraska, for example) relative to USDM, while the DA-TWS–based estimates provide a closer match. The drought intensities over New Mexico and Louisiana are better matched in the DA-TWS estimates as well. The third example represents the onset of the 2011 Texas drought case, which is poorly represented in the OL estimates. The OL estimate does not capture the drought over the southern regions (New Mexico and Texas), whereas drought estimates over Indiana, Ohio, and Pennsylvania are overestimated relative to USDM. The DA-TWS simulation-based estimate matches the USDM intensity and spatial distribution of drought more closely over these regions. DA-TWS estimates, however, underestimate the moderate drought in the southeastern United States; 18–25 September 2012 is representative

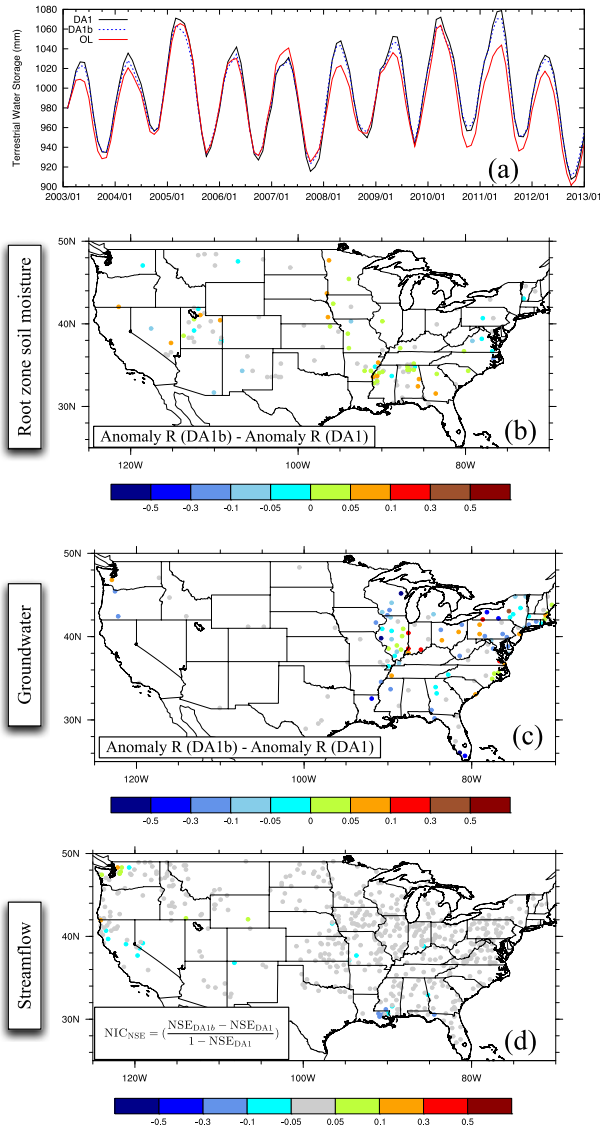

FIG. 11. A comparison of DA-TWS integrations that use gridded retrievals, scaling factors, and spatially distributed measurement errors (DA1) and basin-scale retrievals (DA1b). (a) The time series of domain-averaged monthly TWS. (b),(c) The anomaly R differences of root zone soil moisture and groundwater from DA1 relative to DA1b, respectively. (d) The NIC_{NSE} of streamflow estimates from DA1 compared to DA1b.

of a widespread drought that affected the central United States during that year. The OL-based estimates overestimate the severity over several areas including Minnesota, Iowa, and Wyoming and underestimate the extreme drought over Kansas and Oklahoma relative to USDM. The DA-TWS estimates have reduced bias in those regions (e.g., the representation of the extreme drought over Kansas and Oklahoma is notably improved in the DA-TWS estimate), but they fail to match the spatial patterns of USDM severe drought over Nebraska.

For a quantitative evaluation of the impact of GRACE DA on drought estimates, we compare the weekly percentage area of drought data from the USDM archives against the corresponding estimates generated from the model-integration-based drought indices. Figure 13 shows time series of weekly averages of drought area percentage computed from the TWS percentiles for the open loop and the DA-TWS integrations over the South region of the USDM. The figure also shows the R, RMSE, and bias of drought area percentages against USDM for different drought categories. In this region, the effect of cold season processes is small and soil moisture is a good proxy for drought. We focus on this region first, as our earlier results suggest that DA-TWS contributes more toward improving soil moisture representation compared to other moisture components such as snow. Overall, the skill of the open-loop simulation is high, especially for D_0–D_2 categories, as it provides a good match to the USDM drought area values. The impact of GRACE DA is mixed, as it leads to stronger agreement with USDM during some periods (e.g., 2006/07, 2011, and 2012), and reduced agreement in other periods (late 2008 and late 2009). For more severe drought categories (D_3 and D_4), similar trends can be seen, though the GRACE DA–based estimates show reduced agreement for the 2006/07 drought. The drought area representation is marginally improved by the GRACE DA–based integration in the late 2011 drought, whereas the drought area extent is better captured by the OL integration in the 2006/07 drought.

The quantitative statistics indicate that the aggregate impact of GRACE DA is to marginally reduce model agreement with USDM (especially in D_3 and D_4 categories) in this region as the R values are reduced and RMSE and biases are increased. Similar comparisons of the percent drought area comparisons were conducted over other regions (not shown), with GRACE DA leading to stronger agreement with USDM in the Midwest, West, and high plains and weaker agreement in the Southeast and Northeast. It must be stressed, however, that the percent drought area comparisons do not include direct comparisons of intensity and spatial locations of drought areas, which are often more important for drought monitoring applications. The individual case comparisons in Fig. 12 show that GRACE DA increases agreement with USDM for these cases.

The comparisons with USDM are informative insomuch as USDM is the most widely used drought monitor for the United States. They must be interpreted with some caution, however, since the USDM is not a conventional benchmark product. The weekly USDM product is produced by a rotating team of authors, and their maps are informed by both short- and long-term

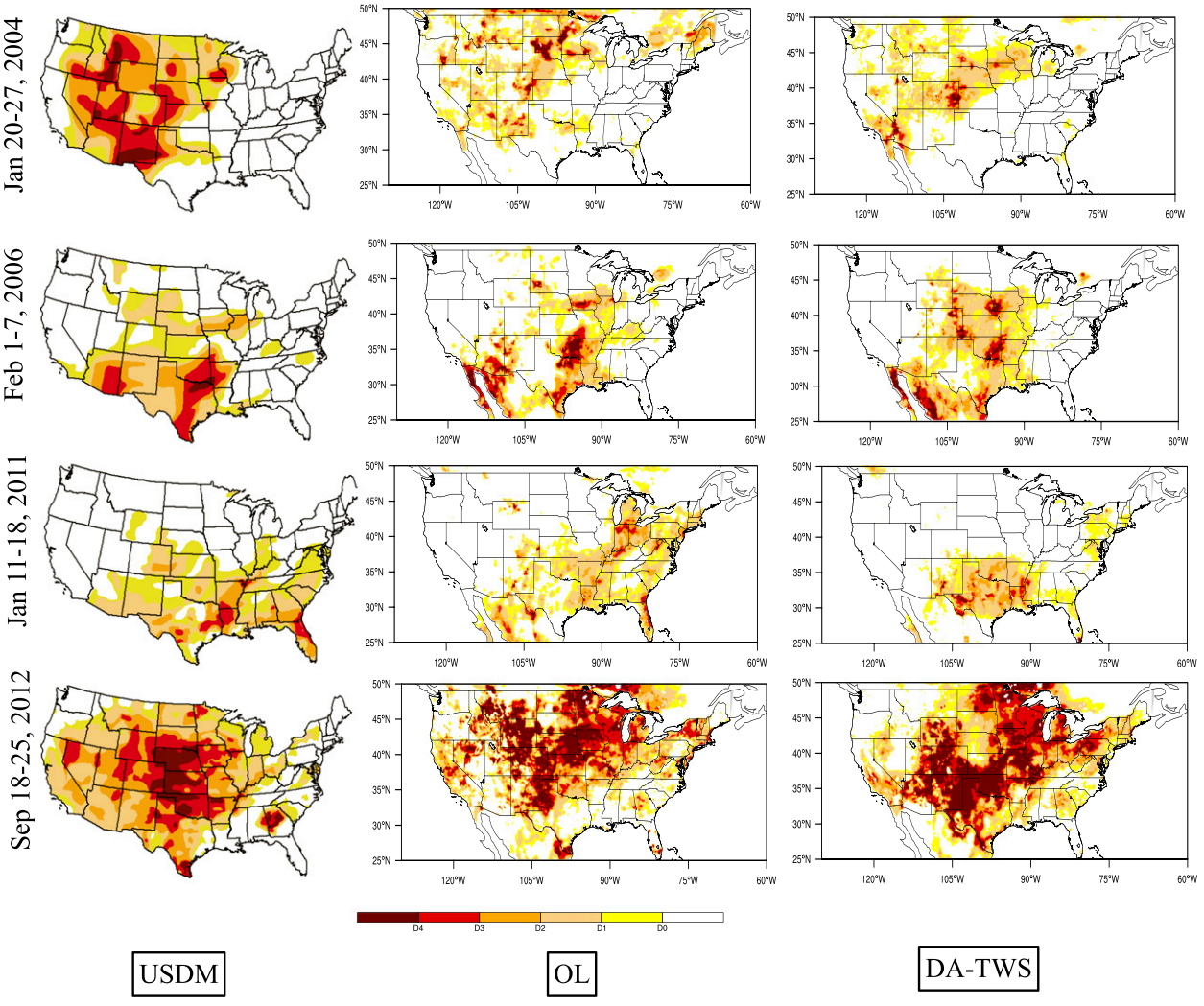

FIG. 12. Comparison of the drought percentile maps from (center) OL and (right) DA-TWS integrations against (left) the corresponding USDM estimate for four representative cases.

objective drought indicators (primarily precipitation based) and by subjective input from a network of local and regional experts (Svoboda et al. 2002). Agreement between a CLSM simulation and USDM can then be reasonably interpreted as an indicator of the simulation's realism, but areas of disagreement do not necessarily mean that the model or assimilation scheme are "wrong." The USDM product in any given week might be influenced by expert opinion of a drought's current impact on a region, while the modeling system simply tries to quantify soil moisture deficits. In addition, the USDM objective indicators tend to be focused on meteorology and its impacts on soil moisture, while GRACE DA is designed to improve simulation of unconfined groundwater as well.

Finally, there is some difficulty in comparing GRACE DA results to USDM after the fall of 2011, when GRACE DA–based drought indicators began to be provided to the USDM authors (Houborg et al. 2012). Therefore, agreement between the two products since that time might simply indicate that the USDM authors are making use of GRACE DA information when generating their drought maps. On the other hand, open-loop NLDAS model-based drought indicators are also used by the USDM authors. That said, the GRACE DA and NLDAS data comprise only one of about five categories of input data considered by the USDM authors. The question of whether GRACE DA has impacted USDM is a subject of active study, and early results are inconclusive.

4. Summary

This article examines the impact of assimilating gridded GRACE TWS data for improving land surface

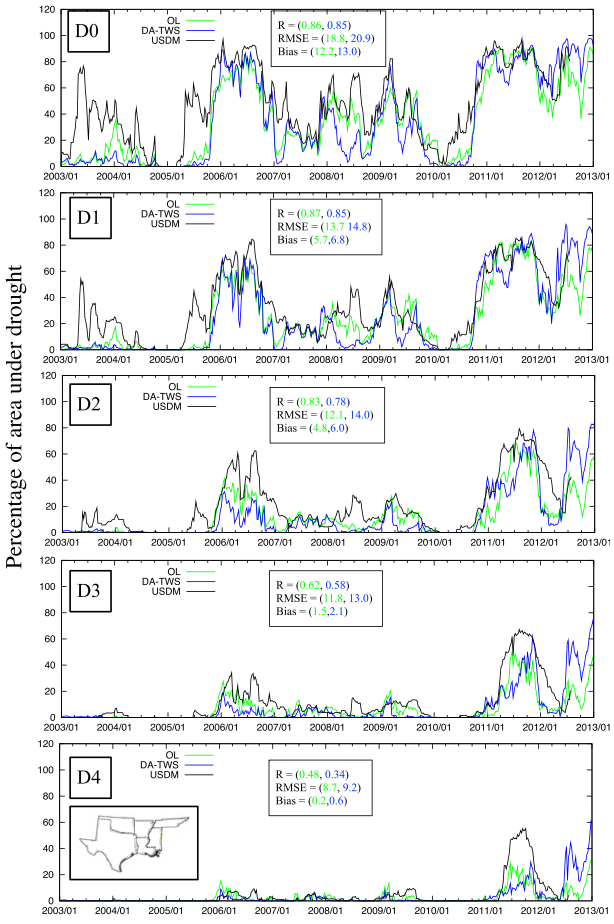

FIG. 13. Time series of the drought area percentage (based on TWS percentiles) from the OL and DA-TWS integrations for the southern United States for a time period of 2000–12. The definition of the South region used in the USDM is shown as an inset at bottom.

model estimates and their contribution toward improved estimation of droughts. The study is conducted over the CONUS using the NLDAS-2 domain configuration and datasets, with the CLSM. The DA integrations are conducted over a time period of 2003–12 using an ensemble smoother algorithm.

The model simulations and the added impact of DA on the land surface model estimates are evaluated using a wide range of independent datasets. The simulation of groundwater estimates was marginally improved by the GRACE DA in the comparisons against in situ USGS groundwater well data. The domain-averaged anomaly *R* skill for the open loop is 0.67, and it improves to 0.69 with GRACE DA. In the evaluations of surface and root zone soil moisture fields against the in situ SCAN and ARS measurements, improvements across the whole modeling domain were obtained. The domain-averaged anomaly *R* for the surface and root zone soil moisture fields for the OL simulation is 0.57 and 0.60, respectively. With GRACE DA, the domain-averaged anomaly *R* skills for the surface and root zone soil moisture fields improve to 0.60 and 0.63, respectively. Comparatively, the impact of GRACE DA on streamflow simulations was small, though improvements in the range of 4%–7% at the Ohio and upper Mississippi River outlets, two of the biggest rivers in terms of discharge magnitude, were obtained. Similar to prior studies such as Forman et al. (2012), the impact of GRACE DA on modeled snow fields was also small. Finally, the impact of GRACE DA on ET estimates was found to be mixed, with some patterns of improvements and degradations observed in the intercomparisons to the four reference datasets.

The NLDAS-2 forcing dataset, especially precipitation, is generally considered to be of high quality (Matsui et al. 2010). The use of NLDAS-2 data as input forcing to the LSMs leads to open-loop estimates with high skills. Consequently, prior studies that have examined the assimilation of soil moisture and snow datasets in the NLDAS-2 configuration have reported marginal success (Peters-Lidard et al. 2011; Kumar et al. 2014). In comparison, the results presented in this paper show more systematic improvements from GRACE DA in key water budget components (groundwater and soil moisture). Though the spatial and temporal resolution of GRACE is much coarser than other typical hydrological remote sensing datasets (soil moisture, snow, and land surface temperature), the fact that GRACE provides observations of the whole water column instead of a few top centimeters of the land surface is a likely contributor to the more significant improvements with GRACE DA observed in the results of this paper.

The study also examined the influence of using the scaling factors and spatially distributed measurement errors in the DA configurations. The use of scaling factors was more impactful than the use of spatially distributed measurement error standard deviations (instead of a uniform 20 mm error) in early years. In later years, the use of the spatially distributed measurement errors was found to be more impactful than the use of scaling factors. Interestingly, we found only small differences between DA of basin-scale GRACE observations and DA of gridded observations, and the two simulations performed comparably when compared to independent observations. This might be because the basins defined in this study are of a size that is on the same order of the true resolution of assimilated GRACE products, or only slightly larger. This means that assimilating on a grid-by-grid basis does not add significant information content relative to the basin average in most cases. Our result is in contrast to Eicker et al. (2014), who found modestly larger differences

between gridded and basin-averaged assimilation, but in experiments where the "basin" was the entire Mississippi basin and the "gridded" observations were at 5° rather than 1° resolution. This meant that their basin-averaged GRACE values were significantly larger than the true resolution of the GRACE TWS estimate.

The influence of DA on improving drought estimates was examined by generating percentiles from TWS fields. A quantitative evaluation of the area fraction in drought for five drought severity categories was examined by comparing the model-based estimates against corresponding USDM archived data. Positive impact from GRACE DA toward improving drought estimates was observed for several individual cases. In the percent drought area comparisons over the South region of USDM, GRACE DA increased CLSM agreement with USDM during the 2006/07, 2011, and 2012 droughts, but it decreased agreement for the late 2008 and late 2009 droughts. The operational USDM maps are drawn by authors who subjectively consider a variety of different drought indicators. Therefore, a single-variable-based (TWS) drought estimate would not be expected to match USDM perfectly. The comparisons presented in the article, however, suggest that GRACE DA can be a significant contributor toward developing a comprehensive, objectively blended drought analysis. We also expect this work to aid future multisensor assimilation studies to help better constrain TWS estimates.

Acknowledgments. Funding for this work was provided by the NASA Science Mission Directorate's Earth Science Division through the National Climate Assessment (NCA) project and NOAA's Climate Program Office MAPP program. Computing was supported by the resources at the NASA Center for Climate Simulation. The NLDAS-2 forcing data used in this effort were acquired as part of the activities of NASA's Science Mission Directorate and are archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC). The GRACE land data available at http://grace.jpl.nasa.gov were supported by the NASA MEaSUREs program.

REFERENCES

Anderson, M., J. Norman, J. Mecikalski, J. Otkin, and W. Kustas, 2007: A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. *J. Geophys. Res.*, **112**, D10117, doi:10.1029/2006JD007506.

Barrett, A., 2003: National Operational Hydrologic Remote Sensing Snow Data Assimilation System (SNODAS) products at NSIDC. Special Rep. 11, NSIDC, 19 pp. [Available online at https://nsidc.org/pubs/documents/special/nsidc_special_report_11.pdf.]

Brasnett, B., 1999: A global analysis for snow depth for numerical weather prediction. *J. Appl. Meteor.*, **38**, 726–740, doi:10.1175/1520-0450(1999)038<0726:AGAOSD>2.0.CO;2.

Brown, R., and B. Brasnett, 2010: Canadian Meteorological Centre (CMC) daily snow depth analysis data, version 1. National Snow and Ice Data Center Distributed Active Archive Center, accessed June 2014, doi:10.5067/W9FOYWH0EQZ3.

Burgers, G., P. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724, doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

Cohen, J., and D. Entekhabi, 1999: Eurasian snow cover variability and Northern Hemisphere climate predictability. *Geophys. Res. Lett.*, **26**, 345–348, doi:10.1029/1998GL900321.

De Lannoy, G., R. Koster, R. Reichle, S. Mahanama, and Q. Liu, 2014: An updated treatment of soil texture and associated hydraulic properties in a global land modeling system. *J. Adv. Model. Earth Syst.*, **6**, 957–979, doi:10.1002/2014MS000330.

Eicker, A., M. Schumacher, J. Kusche, P. Doll, 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**, 1285–1309, doi:10.1007/s10712-014-9309-8.

Ek, M., and Coauthors, 2011: North American Land Data Assimilation System Phase 2 (NLDAS-2): Its development and application to support US drought monitoring and prediction of National Integrated Drought Information System (NIDIS). GEWEX News, Vol. 21, International GEWEX Project Office, Silver Spring, MD, 7–9.

Forman, B., and R. Reichle, 2013: The spatial scale of model errors and assimilated retrievals in a terrestrial water storage assimilation system. *Water Resour. Res.*, **49**, 7457–7468, doi:10.1002/2012WR012885.

——, ——, and M. Rodell, 2012: Assimilation of terrestrial water storage from GRACE in a snow-dominated basin. *Water Resour. Res.*, **48**, W01507, doi:10.1029/2011WR011239.

Gaspari, G., and S. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, doi:10.1002/qj.49712555417.

Getirana, A., A. Boone, D. Yamazaki, B. Decharme, F. Papa, and N. Mognard, 2012: The Hydrological Modeling and Analysis Platform (HyMAP): Evaluation in the Amazon basin. *J. Hydrometeor.*, **13**, 1641–1665, doi:10.1175/JHM-D-12-021.1.

Houborg, R., M. Rodell, B. Li, R. Reichle, and B. 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.

Jackson, T., and Coauthors, 2010: Validation of Advanced Microwave Scanning Radiometer soil moisture products. *IEEE Trans. Geosci. Remote Sens.*, **48**, 4256–4272, doi:10.1109/TGRS.2010.2051035.

Jung, M., M. Reichstein, and A. Bondeau, 2009: Towards a global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, **6**, 2001–2003, doi:10.5194/bg-6-2001-2009.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *J. Geophys. Res.*, **105**, 24 809–24 822, doi:10.1029/2000JD900327.

——, and Coauthors, 2004: Realistic initialization of land surface states: Impacts on subseasonal forecast skill. *J. Hydrometeor.*, **5**, 1049–1063, doi:10.1175/JHM-387.1.

Kumar, S., and Coauthors, 2006: Land information system: An interoperable framework for high resolution land surface modeling. *Environ. Modell. Software*, **21**, 1402–1415, doi:10.1016/j.envsoft.2005.07.004.

——, C. Peters-Lidard, J. Santanello, K. Harrison, Y. Liu, and M. Shaw, 2012: Land surface Verification Toolkit (LVT)—A generalized framework for land surface model evaluation. *Geosci. Model Dev.*, **5**, 869–886, doi:10.5194/gmd-5-869-2012.

——, and Coauthors, 2014: Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. *J. Hydrometeor.*, **15**, 2446–2469, doi:10.1175/JHM-D-13-0132.1.

Landerer, F., and S. Swenson, 2012: Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour. Res.*, **48**, W04531, doi:10.1029/2011WR011453.

Li, B., M. Rodell, B. Zaitchik, R. Reichle, and R. Koster, 2012: Assimilation of GRACE terrestrial water storage into a land surface model: Evaluation and potential value for drought monitoring in western and central Europe. *J. Hydrol.*, **446–447**, 103–115, doi:10.1016/j.jhydrol.2012.04.035.

Mahanama, S., B. Livneh, R. Koster, R. Lettenmaier, and D. Reichle, 2012: Soil moisture, snow and seasonal streamflow forecasts in the United States. *J. Hydrometeor.*, **13**, 189–203, doi:10.1175/JHM-D-11-046.1.

Matsui, T., D. Mocko, M.-I. Lee, M. Suarez, and R. Pielke Sr., 2010: Ten-year climatology of summertime diurnal rainfall rate over the conterminous U.S. *Geophys. Res. Lett.*, **37**, L13807, doi:10.1029/2010GL044139.

Maxwell, R., F. K. Chow, and S. Kollet, 2007: The groundwater–land-surface–atmosphere connection: Soil moisture effects on the atmospheric boundary layer in fully coupled simulations. *Adv. Water Resour.*, **30**, 2447–2466, doi:10.1016/j.advwatres.2007.05.018.

Menne, M. J., I. Durre, R. Vose, B. Gleason, and T. Houston, 2012: An overview of the global historical climatology network-daily database. *J. Atmos. Oceanic Technol.*, **29**, 897–910, doi:10.1175/JTECH-D-11-00103.1.

Mitchell, K. E., and Coauthors, 2004: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.*, **109**, D07S90, doi:10.1029/2003JD003823.

Mu, Q., M. S. Zhao, and S. Running, 2011: Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.*, **115**, 1781–1800, doi:10.1016/j.rse.2011.02.019.

Peters-Lidard, C., and Coauthors, 2007: High-performance Earth system modeling with NASA/GSFC's Land Information System. *Innov. Syst. Software Eng.*, **3**, 157–165, doi:10.1007/s11334-007-0028-x.

——, S. Kumar, D. Mocko, and Y. Tian, 2011: Estimating evapotranspiration with land data assimilation systems. *Hydrol. Processes*, **25**, 3979–3992, doi:10.1002/hyp.8387.

Reichle, R., and R. Koster, 2003: Assessing the impact of horizontal correlations in background fields of soil moisture estimation. *J. Hydrometeor.*, **4**, 1229–1242, doi:10.1175/1525-7541(2003)004<1229:ATIOHE>2.0.CO;2.

——, ——, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe, 2007: Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *J. Geophys. Res.*, **112**, D09108, doi:10.1029/2006JD008033.

Rodell, M., P. Houser, A. Berg, and J. Famiglietti, 2005: Evaluation of 10 methods for initializing a land surface model. *J. Hydrometeor.*, **6**, 146–155, doi:10.1175/JHM414.1.

——, J. Chen, H. Kato, J. Famiglietti, J. Nigro, and C. Wilson, 2007: Estimating ground water storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeol. J.*, **15**, 159–166, doi:10.1007/s10040-006-0103-7.

——, I. Velicogna, and J. Famiglietti, 2009: Satellite-based estimates of groundwater depletion in India. *Nature*, **460**, 999–1002, doi:10.1038/nature08238.

——, E. McWilliams, J. Famiglietti, H. Beudoing, and J. Nigro, 2011: Estimating evapotranspiration using an observation based terrestrial water budget. *Hydrol. Processes*, **25**, 4082–4092, doi:10.1002/hyp.8369.

Rowlands, D. D., S. Luthcke, S. Klosko, F. Lemoine, D. Chinn, J. McCarthy, C. Cox, and O. Anderson, 2005: Resolving mass flux at high spatial and temporal resolution using GRACE intersatellite measurements. *Geophys. Res. Lett.*, **32**, L04310, doi:10.1029/2004GL021908.

Sakumura, C., S. Bettadpur, and S. Bruinsma, 2014: Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophys. Res. Lett.*, **41**, 1389–1397, doi:10.1002/2013GL058632.

Schaefer, G., M. Cosh, and T. Jackson, 2007: The USDA natural resources conservation service soil climate analysis network (SCAN). *J. Atmos. Oceanic Technol.*, **24**, 2073–2077, doi:10.1175/2007JTECHA930.1.

Sheffield, J., Y. Xia, L. Luo, E. Wood, M. Ek, K. Mitchell, and the NLDAS Team, 2012: Drought monitoring with the North American Land Data Assimilation System (NLDAS): A framework for merging model and satellite data for improved drought monitoring. *Remote Sensing of Drought: Innovative Monitoring Approaches*, B. Wardlow, M. Anderson, and J. Verdin, Eds., Taylor and Francis, 228–255.

Stieglitz, M., A. Ducharne, R. Koster, and M. Suarez, 2001: The impact of detailed snow physics on the simulation of snow cover and subsurface thermodynamics at continental scales. *J. Hydrometeor.*, **2**, 228–242, doi:10.1175/1525-7541(2001)002<0228:TIODSP>2.0.CO;2.

Su, H., Z.-L. Yang, R. Dickinson, C. Wilson, and G.-Y. Niu, 2010: Multisensor snow data assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment terrestrial water storage information. *J. Geophys. Res.*, **115**, D10104, doi:10.1029/2009JD013035.

Svoboda, M., and Coauthors, 2002: The Drought Monitor. *Bull. Amer. Meteor. Soc.*, **83**, 1181–1190, doi:10.1175/1520-0477(2002)083<1181:TDM>2.3.CO;2.

Swenson, S., and J. Wahr, 2002: Methods for inferring regional surface-mass anomalies from Gravity Recovery Climate Experiment (GRACE) measurements of time-variable gravity. *J. Geophys. Res.*, **107**, 2193, doi:10.1029/2001JB000576.

——, and ——, 2006: Post-processing removal of correlated errors in GRACE data. *Geophys. Res. Lett.*, **33**, L08402, doi:10.1029/2005GL025285.

——, P.-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, doi:10.1029/2006GL026962.

Tang, Q., S. Peterson, R. Cuenca, Y. Hagimoto, and D. Lettenmaier, 2009: Satellite-based near-real-time estimation of irrigated crop water consumption. *J. Geophys. Res.*, **114**, D05114, doi:10.1029/2008JD010854.

——, H. Gao, P. Yeh, T. Oki, F. Su, and D. Lettenmaier, 2010: Dynamics of terrestrial water storage change from satellite and surface observations and modeling. *J. Hydrometeor.*, **11**, 156–170, doi:10.1175/2009JHM1152.1.

Tangdamrongsub, N., S. Steele-Dunne, B. Gunter, P. Ditmar, and A. Weerts, 2015: Data assimilation of GRACE terrestrial water storage estimates into a regional hydrological model of the Rhine River basin. *Hydrol. Earth Syst. Sci.*, **19**, 2079–2100, doi:10.5194/hess-19-2079-2015.

Tapley, B., S. Bettadpur, M. Watkins, and C. Reigber, 2004: The Gravity Recovery and Climate Experiment: Mission overview and early results. *Geophys. Res. Lett.*, **31**, L09607, doi:10.1029/2004GL019920.

Velicogna, I., 2009: Increasing rates of ice mass loss from the Greenland and Antarctic Ice Sheets revealed by GRACE. *Geophys. Res. Lett.*, **36**, L19503, doi:10.1029/2009GL040222.

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**, 30 205–30 229, doi:10.1029/98JB02844.

——, S. Swenson, and I. Velicogna, 2006: Accuracy of GRACE mass estimates. *Geophys. Res. Lett.*, **33**, L06401, doi:10.1029/2005GL025305.

Xia, Y., M. Ek, H. Wei, and J. Meng, 2012a: Comparative analysis of relationships between NLDAS-2 forcings and model outputs. *Hydrol. Processes*, **26**, 467–474, doi:10.1002/hyp.8240.

——, and Coauthors, 2012b: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.*, **117**, D03109, doi:10.1029/2011JD016048.

Zaitchik, B., M. Rodell, and R. Reichle, 2008: Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River basin. *J. Hydrometeor.*, **9**, 535–548, doi:10.1175/2007JHM951.1.