**Multi-sensor Assimilation of SMOS Brightness Temperature and GRACE Terrestrial
Water Storage Observations for Soil Moisture and Shallow Groundwater Estimation**

Manuela Girotto^{1,2}, Rolf H. Reichle¹, Matthew Rodell³, Qing Liu^{1,4}, Sarith Mahanama^{1,4},
Gabriëlle J. M. De Lannoy⁵

¹ Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt,
MD, USA

² GESTAR, Universities Space Research Association, Columbia, MD 21044, USA

³ Hydrological Science Lab, NASA Goddard Space Flight Center, Greenbelt, MD, USA

⁴ Science Systems and Applications, Inc., Lanham, MD, USA

⁵ Department of Earth and Environmental Sciences, KU Leuven, Belgium

Corresponding Author: Manuela Girotto (manuela.girotto@nasa.gov)

Submitted to: Remote Sensing of the Environment

Date revised: March 30, 2019

19

20

21 **Highlights:**

- 22 1. Single-sensor assimilation of GRACE data improves groundwater estimates
- 23 2. Single-sensor assimilation of SMOS data improves surface soil moisture estimates
- 24 3. Joint assimilation of GRACE+SMOS data merges single-sensor assimilation benefits
- 25 4. Water storage increments from joint GRACE+SMOS assimilation are anti-correlated

26

27 **Keywords:**

28 GRACE, SMOS, soil moisture, terrestrial water storage, shallow groundwater, multi-sensor

29 data assimilation, multi-resolution data assimilation

30

31

Abstract

The Gravity Recovery and Climate Experiment (GRACE) mission provided monthly global estimates of the vertically integrated terrestrial water storage with about 300-400-km horizontal resolution between 2002-2017. Since 2009, the Soil Moisture and Ocean Salinity (SMOS) mission observes global L-band brightness temperatures, which are sensitive to near-surface soil moisture, with a revisit time of 1-3 days at a nominal 43-km spatial resolution. This work investigates if the multi-sensor assimilation of these observations into the Catchment land surface model can improve the estimation of 0-5 cm “surface” soil moisture, 0-100 cm “rootzone” soil moisture, and shallow (unconfined) groundwater levels. Single-sensor GRACE or SMOS assimilation and multi-sensor GRACE+SMOS assimilation experiments were performed over the continental U.S. for 6 years (July 2010 – June 2016). GRACE data assimilation mostly improves estimates of shallow groundwater, whereas SMOS data assimilation mainly improves estimates of surface soil moisture. The benefits introduced by the single-sensor assimilation are merged in the multi-sensor assimilation experiment, suggesting that better and more consistent soil moisture and groundwater estimates can be achieved when multiple observation types are assimilated. Interestingly, in the multi-sensor GRACE+SMOS experiment, the water storage increments introduced by the GRACE analysis and the SMOS analysis are anti-correlated. That is, when the GRACE assimilation increments remove water from the overall profile storage, the SMOS assimilation increments add water to it, and vice versa. This anti-correlation could be caused by the SMOS analysis trying to undo the increments from the GRACE analysis.

1. Introduction

The number of satellites observing land surface hydrological conditions has rapidly increased within the past decades (Lettenmaier et al., 2015; McCabe et al., 2017). The integration of these satellite data into land surface models may be the most promising approach for obtaining accurate, complete and consistent time series of land surface variables across the globe. Current data assimilation techniques are theoretically able to merge multiple types of observations coherently into a dynamical physically-based model, but only a handful of applications have attempted to develop and use a multi-variate analysis using multiple sensor observations (e.g., Kumar et al., 2018; Tian et al., 2017; Renzullo et al., 2008), and the majority of the modern era satellite observations are not yet fully exploited in a comprehensive manner. Many science efforts have focused on single-sensor assimilation experiments geared toward the estimation of a single hydrological process such as soil moisture (Bolten et al., 2009; De Lannoy & Reichle, 2016a, 2016b; Lievens et al., 2015; Reichle et al., 2007) snow (De Lannoy et al., 2012; Dziubanski & Franz, 2016; Giroto et al., 2014; Kumar et al., 2014; Zaitchik & Rodell, 2009) or terrestrial water storage (Giroto et al., 2016; Houborg et al., 2012; Kumar et al., 2016; Li et al., 2012; Su et al., 2010; Tangdamrongsub et al., 2015; Zaitchik et al., 2008). A more comprehensive understanding of the hydrological cycle should be achieved when estimates are constrained by complementary information from a broad range of observations and models (Pan et al., 2012; Rodell et al., 2015).

This paper specifically aims to estimate water storage components across the soil column as a whole, from the surface soil moisture down to the shallow (unconfined) groundwater (typically at depths less than 4 m). Recent microwave and gravity satellite missions are able to monitor these variables. Examples include the Soil Moisture Active Passive (SMAP, Entekhabi et al., 2010a) mission, the Soil Moisture and Ocean Salinity

(SMOS, Kerr et al., 2001) mission, and the Gravity Recovery and Climate Experiment (GRACE, Tapley, et al. 2004) mission. The latter provides global estimates of terrestrial water storage (TWS), defined as the sum of groundwater, soil moisture, snow water equivalent, surface water, ice, and water in biomass. The TWS retrievals have monthly temporal resolution and a spatial (horizontal) resolution of approximately 300-400-km at mid-latitudes (Landerer and Swenson, 2012). In contrast, SMOS observes brightness temperature (T_b) at L-band frequency (i.e., 1.4 GHz), which is sensitive to soil moisture and temperature in the top few centimeters of the soil, at much finer temporal and spatial resolutions (every 1-3 days and at 43-km; Kerr et al., 2001).

Previous data assimilation efforts have shown how these satellite missions, individually, can provide valuable information regarding different aspects of the water cycle. For example, the assimilation of near surface soil moisture retrievals or brightness temperature observations from the SMOS and SMAP missions yields improvements in modeled soil moisture estimates (e.g., De Lannoy & Reichle, 2016a,b; Reichle et al., 2007, 2017a,b). In some cases, the assimilation of soil moisture information also produces improved estimates of evaporation, and runoff (Brocca et al., 2010; Draper et al., 2011; Lievens et al., 2015; Reichle & Koster, 2005; Renzullo et al., 2014; Walker et al., 2001; Crow et al. 2017) or facilitates the evaluation of land model runoff processes (Crow et al. 2018). Further, the assimilation of GRACE TWS observations has proven valuable for groundwater estimation (e.g., Zaitchik et al., 2008; Girotto et al., 2016), drought monitoring (e.g., Houborg et al., 2012; Li et al., 2012; Schumacher et al., 2018), evapotranspiration applications (e.g., Kumar et al., 2016; Girotto et al., 2017), assessment of flood potential (Reager et al., 2015), streamflow estimation (e.g., Tangdamrongsub et al., 2015), estimation of snow water equivalent (e.g., Forman et al. 2012), identification of human driven hydrological

mechanisms (e.g., Felfelani et al., 2017; Girotto et al., 2017), and improving the global water budget (e.g., van Dijk et al., 2014).

The hypothesis for the present work is that an unprecedented accuracy in surface soil moisture, rootzone soil moisture and shallow groundwater estimates can be obtained through the multi-sensor assimilation of gravity (GRACE) and passive microwave (SMOS) observations. These two types of observations are complementary in terms of horizontal, vertical and temporal support. At the same time, combining observations from such different sensors presents a challenge owing to the unique disparity in their spatial and temporal scales.

Here, we perform multi-sensor and multi-scale data assimilation using the land component of the NASA Goddard Earth Observing System (GEOS), while taking advantage of the previously developed (single-sensor) assimilation techniques. The paper aims at 1) quantifying the improvements of hydrological estimates via single-sensor and multi-sensor assimilation of GRACE TWS and SMOS Tb observations through validation with in-situ measurements; and 2) understanding the impact of each assimilation system on the soil moisture profile estimates through an analysis of the associated assimilation diagnostics. These objectives provide the structural sub-headings used for the Results and Discussions sections.

2. Background

GRACE TWS solutions are provided as a temporally averaged, typically monthly, measurement of the entire column of terrestrial water. That is, TWS encompasses snow, groundwater, soil moisture, and surface water in lakes, wetlands, and rivers. Most other satellite data, such as those obtained from SMOS or SMAP, provide instantaneous electromagnetic observations that are directly related to a specific hydrological variable.

Furthermore, the spatial scale of GRACE TWS solutions is on the order of ~300 km which is typically much larger than other remotely sensed observations used in hydrology. Because of the uniqueness of the spatial and temporal resolutions of GRACE it is not trivial to merge GRACE data with other remotely sensed observations. In recent years, a few studies merged GRACE TWS observations with other satellite observations (**Table 1**).

First, Su et al., (2010) performed a one-dimensional (1D) assimilation of GRACE along with MODIS snow cover fraction information, whereby each model grid cell is updated only with local information and independently from all other grid cells. They focused on improving snow estimates over North America and found that the multi-sensor approach can provide significant improvements over a MODIS-only assimilation approach. More recently, Tian et al. (2017) jointly assimilated GRACE TWS and SMOS soil moisture retrievals over Australia using a 1D smoother assimilation technique for the period January 2010 - December 2013. In their results, groundwater estimates degraded when SMOS was assimilated. Nevertheless, overall the multi-sensor assimilation performed better than the GRACE or SMOS single sensor assimilation.

One approach to address the unique spatial and temporal resolutions of GRACE data is to interpolate and disaggregate the GRACE observations to scales that are more similar to those of other satellite missions. For example, Khaki and Awange (2019) disaggregated the GRACE observations to 5 days and aggregated the SMOS and AMSR-E observations to the same temporal resolution. These pre-processing steps allowed the authors to apply classical ensemble assimilation techniques over South America to investigate surface and subsurface hydrology. They found that more accurate estimates of groundwater and soil moisture variations are achieved via multi-sensor assimilation. Similarly, Zhao and Yang (2018) assimilated GRACE observations, along with AMSR-E brightness temperature and MODIS snow cover fraction information. In their work, they use a daily version of the GRACE

observations (Sakumura et al., 2016) with a spatial resolution of 0.5 degrees. They were able to complete a global land multi-sensor assimilation using a deterministic assimilation approach. They found that the assimilation of MODIS snow cover fraction improves snow estimation in the mid- and high-latitudes, while the AMSR-E assimilation improves global estimates of soil moisture and snow. The assimilation of GRACE was found to degrade soil moisture but improve snow estimation in high-latitudes.

In our work, we avoid the pre-processing of the satellite data and the associated increase in the noise in the assimilated observations. Thus, our approach is similar to the one taken by Tian et al., (2017). Unlike Tian et al. (2017), however, we apply a spatially distributed (3D) sequential assimilation scheme to spatially and vertically distribute the weights given to the satellite observations in the analysis. We focus our analysis on the Contiguous US (CONUS) for the period from July 2010 to June 2016. This period corresponds to maximum number (6) of full years for which both SMOS and GRACE observations of sufficient quality are available. Further, in our system we assimilate the SMOS radiances (brightness temperatures, T_b) instead of the soil moisture retrievals. A key disadvantage of a system that assimilates SM retrievals (such as in Tian et al., 2017) is that the SM retrievals may be produced with inconsistent ancillary data, such as for example soil temperature simulated by another model than that used in the assimilation system (De Lannoy and Reichle, 2016a).

1. < Table 1 here please >

3. Data and Model

3.1 Assimilated Satellite Observations

We assimilate passive microwave observations (T_b) from the SMOS mission and TWS retrievals from the GRACE mission. Specific details about the observations and the single-

sensor data assimilation are described in Girotto et al. (2016) for GRACE TWS assimilation and in De Lannoy & Reichle (2016a) for SMOS Tb assimilation.

3.1.1. GRACE TWS Observations

We use the level-3, monthly GRACE TWS (unscaled) product based on the RL05 spherical harmonics, which is available from the Jet Propulsion Laboratory (<http://grace.jpl.nasa.gov>; Landerer and Swenson, 2012). GRACE TWS observations are provided on a grid with $1^\circ \times 1^\circ$ spacing, although errors in neighboring observations are strongly correlated because the effective spatial resolution of the GRACE TWS observations is in the order of 300-400-km, with vertical accuracy ~ 10 -100 mm (Wahr et al., 2006; Swenson et al., 2006). Therefore, the calculation of error covariances for the update step (explained later in Section 3.1) at the 1° scale of the TWS data product is not well-conditioned. Previous studies have overcome this issue by aggregating the gridded GRACE TWS observations to between $2^\circ \times 2^\circ$ and $5^\circ \times 5^\circ$ (e.g., Eicker et al., 2014; Khaki et al., 2017), while still considering the error correlation of the aggregated observations. In our study, we thin (rather than aggregate) the $1^\circ \times 1^\circ$ gridded product; that is, we assimilate only one $1^\circ \times 1^\circ$ observation out of every three in the latitudinal and longitudinal directions. Put differently, the spacing of the assimilated observations is 3° . We use a spatial correlation length (radius) of 2° for the observation error (section 4.1). Thus, the thinning approach captures the relevant information provided in the original observational dataset while ensuring a numerically robust analysis solution (i.e., a set of well-conditioned matrices for the calculation of the Kalman Gain (section 4.1). Throughout the remainder of the paper, we refer to the thinned GRACE TWS data simply as “GRACE observations”.

3.1.2. SMOS Tb Observations

The SMOS Tb data we use in this work include both horizontally and vertically polarized data, filtered and interpolated (fitted) to a 40° incidence angle based on the multi-angular Level 1 SMOS (SCLF1C) observations from a given overpass and location. Fitting details are provided in De Lannoy et al. (2015). Data are provided on a 36km Equal Area Scalable Earth (EASE) version 2 grid (Brodzik et al., 2012). We removed SMOS data that are potentially affected by L-band radio frequency interference (RFI) or emission contributions from open water as described by De Lannoy et al. (2015). Observation error characteristics are described in Section 4.1. Further details of the SMOS preprocessing and quality control are described in De Lannoy & Reichle (2016a). Throughout the remainder of the paper, we will refer to the fitted and quality controlled Tb observations generally as “SMOS observations”.

3.2. Land Surface Model, Forcing Data, and Study Domain

We use the land component of the GEOS modeling and data assimilation framework: the Catchment land surface model (CLSM) (Koster et al., 2000). CLSM has been used in previous work for both TWS and soil moisture assimilation (e.g., Zaitchik et al., 2008; Giroto et al., 2016; De Lannoy & Reichle, 2016b, 2016a; Reichle et al., 2017a; 2017b). CLSM is able to represent shallow groundwater storage changes, for this reason it has been targeted for GRACE TWS data assimilation studies.

The three CLSM prognostic variables catchment deficit (*catdef*), rootzone excess (*rzexc*), and surface excess (*srfexc*) define the soil moisture profile. These prognostic variables are used in the assimilation state vector (Section 4.2 and Section 4.3). Rootzone excess is defined as the amount of water in the rootzone layer (0–100 cm) in excess of the amount that would be stored if the soil moisture profile was in equilibrium (Ducharne et al. 2000). Surface excess is correspondingly defined for the surface layer (0–5 cm). The model

prognostic catchment deficit (*catdef*) is “the average depth of water that would need to be added in order to bring the catchment to saturation”, Ducharne et al. (2000). It is complementary to the modeled unconfined mean groundwater table depth, which is the reason why in this writeup *catdef* is directly associated with the modeled shallow groundwater (Giroto et al., 2016). Throughout the remainder of the paper, we will often refer to the modeled shallow groundwater generally as “groundwater”. Other model prognostic variables that are used in the assimilation schemes (Section 4.2 and Section 4.3) are the layer-1 ground heat content (*ght1*) and the surface skin temperature. The *ght1* controls the layer-1 soil temperature, whereas (for snow-free conditions) the surface skin temperature is obtained from the temperatures across the saturated (*tc1*), unsaturated (*tc2*), and wilting (*tc4*) sub-tile areas (De Lannoy et al., 2016b). A tau-omega radiative transfer model is used to diagnose L-band brightness temperatures based on CLSM simulations of soil moisture and temperature (De Lannoy et al., 2013; De Lannoy et al., 2014) using similar parameters and ancillary inputs as in De Lannoy & Reichle (2016b, 2016a).

In this paper we use surface meteorological forcing data from the Modern Era Retrospective Analysis for Research Application version 2 (MERRA-2; Gelaro et al., 2017). The precipitation forcing used in this work is generated by the system’s Atmospheric General Circulation Model (AGCM) following the assimilation of atmospheric temperature, humidity, and wind observations, among others. That is, the precipitation is not corrected with the gauge-based product (Reichle et al., 2017c). This means that the improvements from the data assimilation experiments are representative of those that could be obtained in regions where the coverage with precipitation gauges is poor, which includes most of world outside of CONUS, western Europe and small regions in South America, and Asia (Koster et al., 2016).

Our study domain is the CONUS, and the study period covers 1 July 2010 to 1 July 2016, when both GRACE and SMOS provided high quality observations. The model grid is the 36-km EASEv2.

3.3. Model Perturbation Setup

Forecast errors are modeled by adding random fields, or perturbations, to the forcing and model prognostic variables of each ensemble member. These perturbation fields are correlated in space as well as in time (e.g., Reichle et al., 2007). The GRACE and SMOS observations have different spatial and temporal resolutions (Section 3.1) and therefore observe processes at different scales. The land model inherently simulates processes at various spatial and temporal scales. In previous single-sensor assimilation experiments (e.g., De Lannoy & Reichle, 2016b, 2016a; Girotto et al., 2016), the spatial and temporal model error correlation scales were chosen relative to the particular characteristics of the observations. For example, De Lannoy & Reichle (2016b, 2016a) assimilated SMOS observations using model error perturbation spatial scales of 0.5° and sub-daily temporal correlations, whereas Girotto et al. (2016, 2017) assimilated GRACE observations using model error perturbation spatial scales of 2° and daily to three-day temporal error correlations. The larger spatial and longer temporal scales of the perturbations used in the GRACE TWS assimilation ensured a reasonable distribution of the coarse-scale GRACE observational information to the finer-scale resolution of the land surface model.

We revised our earlier perturbation schemes for single-sensor assimilation to better reflect errors at multiple scales. For the multi-sensor GRACE + SMOS assimilation experiment presented here, we use a two-scale perturbation configuration in which some of the model prognostics and forcings are perturbed using the settings from the single-sensor

SMOS Tb assimilation while the remainder of the model prognostics and forcings are perturbed using the settings from the single-sensor GRACE TWS assimilation. Specifically, the model prognostics *swe* (for snow water equivalent) and *catdef* are also perturbed with the longer spatial error correlation length of 2° and a temporal correlation of one day (Giroto et al., 2016, 2017). In contrast, the model prognostic *srfexc* is perturbed with a shorter spatial error correlation length of 0.5° and temporal correlation of 3 hours (De Lannoy & Reichle, 2016b, 2016a). The model prognostics *rzexc*, *tc1*, *tc2*, *tc4*, and *ght* are not perturbed in line with previous studies. Precipitation, solar and longwave radiation are perturbed with the larger (2°) spatial perturbation scales that represent mesoscale perturbation characteristics. We use twenty-four ensemble members to generate the ensemble spread (Giroto et al., 2016). We perturb the forcing and model prognostic using the same perturbations for all of the experiment cases described in Section 4. **Table 2** summarizes the perturbation parameters.

< Table 2 here please >

3.4. Validation Datasets

We use independent in-situ measurements of soil moisture, groundwater and runoff to validate the data assimilation experiments. Details regarding the validation datasets and calculation of statistics can be found in Giroto et al. (2016); only key points and differences are included below.

Rootzone and surface soil moisture in-situ measurements come from both sparse network and grid-cell average measurements. The grid-cell-scale (EASEv2 at 36-km) are referred to as “Cal/Val” and are collected by the U.S. Department of Agriculture (Cosh et al., 2008; Entekhabi et al., 2014; Jackson et al., 2010). We use a total of 4 Cal/Val sites for surface soil moisture and 2 Cal/Val sites for rootzone soil moisture. Sparse network data

were obtained from two networks over the U.S., the Soil Climate Analysis Network (SCAN) (Schaefer et al., 2007) and the U.S. Climate Reference Network (USCRN) (Diamond et al., 2013). We retained measurements from a total of 52 SCAN sites and 41 USCRN sites for the validation of surface soil moisture, and from 46 SCAN sites and 38 USCRN sites for rootzone soil moisture. For groundwater, observations were collected by the U.S. Geological Survey (USGS), and by the Illinois State Water Survey (<http://www.isws.illinois.edu/warm>). We use a total of 157 groundwater monitoring wells. Runoff measurements were obtained for 238 unregulated medium size river basins in the CONUS from the USGS. For each basin, the observed river discharge was normalized by the basin area to convert the discharge into units of millimeters per day (Koster et al., 2018). We compare these observations to the basin average modeled estimates of runoff.

All available measurements within the experiment period (1 July 2010 – 1 July 2016) are used and the validation is performed on monthly averaged time series. The statistical skill metrics include the correlation coefficient (R) and the unbiased root-mean-square difference ($ubRMSD$) (i.e., $ubRMSD^2 = RMSD^2 - bias^2$; Entekhabi et al., 2010b). The skill metrics and the 95% confidence intervals (CI) are calculated at the individual sites. The domain averages of metrics (and confidence intervals) are obtained from grouping the individual sites into clusters (based on distance), and from averaging across each cluster first and then across all cluster averages (De Lannoy & Reichle, 2016b). Owing to the monthly resolution of the validation and the small number of Cal/Val sites, the statistical confidence intervals for these sites are very large and not shown in the results.

4. Methods

To meet our objectives of (i) improving hydrological estimates via SMOS and GRACE data assimilation and of (ii) understanding the associated assimilation diagnostics, we

designed four experiments. The first experiment consists of an ensemble open-loop (or model-only) experiment in which perturbations are applied but no satellite observations are assimilated. The three data assimilation experiments, summarized in **Table 4**, include one case with single-sensor GRACE assimilation (GRACE DA; Section 4.2), one case with single-sensor SMOS assimilation (SMOS DA; Section 4.3), and one case in which both types of observations are assimilated jointly (GRACE+SMOS DA; Section 4.4).

< Table 4 here please >

4.1. Data Assimilation

All assimilation experiments are performed using a spatially distributed (3D) ensemble Kalman filter (EnKF) (De Lannoy et al., 2010; Reichle & Koster, 2003). The “3D” notation refers to the fact that the filter distributes information horizontally as well as vertically. The 3D filter updates the model estimates in each 36-km model grid cell by assimilating multiple satellite observations within a larger radius of influence (see below) while exploiting spatial error covariance structures of both the model and observations.

The ensemble of $N = 24$ members is constructed by integrating the model state vector ($\mathbf{x}_{i,t}^j$) in time while applying perturbations to the CLSM prognostic variables and forcings (Section 3.3). The subscript t represents time, i designates the 36-km model grid cell and j indicates ensemble members ($j = 1, \dots, N$). To generate an ensemble for the first update step, the model was spun-up in ensemble mode for the 6 months (1 January to 1 July 2010) prior to the start of the assimilation. Depending on the assimilation experiment, increments are only calculated for a subset of the state vector ($\mathbf{x}_{i,t}^j$) for a single grid cell i (Section 4.2-4.4).

The state vector update (or increment, $\Delta\mathbf{x}$) of each ensemble member is calculated using the standard ensemble Kalman filter equation:

$$\Delta \mathbf{x}_{i,t}^j = \mathbf{K}_{I \rightarrow i, T \rightarrow t} [\mathbf{Z}_{I,T}^j - \mathbf{M}(\mathbf{x}^j)|_{I,T}] \quad \text{Equation (1)}$$

Where $\mathbf{K}$ is the Kalman gain and the vector $\mathbf{Z}_{I,T}^j$ contains perturbed observations from GRACE or SMOS (Burgers et al., 1998). **Table 3** lists the standard deviations and spatial correlation characteristics used to obtain these perturbations. The vector $\mathbf{M}(\cdot)$ is the collection of the observation predictions corresponding to the GRACE or SMOS observations. The subscript T indicates the temporal resolution of the observations; the subscript I is the set of observations within the influence radius r_I (i.e., 6° and 1° for GRACE and SMOS observations respectively, **Table 3**) from which observations (at their native spatial resolution) are collected to update the state vector at grid cell i . In a nutshell, **Equation 1** translates the vector of observation-minus-forecast residuals (i.e., innovations $[\mathbf{Z}_{I,T}^j - \mathbf{M}(\mathbf{x}^j)|_{I,T}]$) into increments (or updates) to the model prognostic state at the model spatial and temporal resolution (i, t) . The increments are computed based on the partitioning given by the Kalman gain ($\mathbf{K}_{I \rightarrow i, T \rightarrow t}$) with consideration of the relative uncertainties of the model forecast and the observations:

$$\mathbf{K}_{I \rightarrow i, T \rightarrow t} = \mathbf{C}_{\mathbf{xM}|_{I \rightarrow i, T \rightarrow t}} [\mathbf{R}_{I,T} + \mathbf{C}_{\mathbf{MM}|_{I,T}}]^{-1} \quad \text{Equation (2)}$$

Where the covariance $\mathbf{C}_{\mathbf{xM}|_{I \rightarrow i, T \rightarrow t}}$ is the error covariance computed between the forecast CLSM prognostic state ($\mathbf{x}_{i,t}$, at model spatial and temporal resolution i, t) and the forecast observation predictions ($\mathbf{M}(\cdot)|_{I,T}$, at observation temporal resolution T , with an observation-specific spatial resolution and collected within an influence radius r_I). The $\mathbf{C}_{\mathbf{MM}|_{I,T}}$ is the sample error covariance of the observation predictions. Again, the dimensions of the gain matrix ($\mathbf{K}$) differ depending on whether GRACE or SMOS data are assimilated. Details of each assimilation system are reported in the next three sections and are summarized in **Table 3**.

< Table 3 here please >

4.2. GRACE TWS Assimilation (GRACE DA)

The assimilation of GRACE TWS observations (GRACE DA) generally follows Giroto et al. (2016). Only key points and differences from the previously described work are reported in this Section. The assimilation increments are computed for the following CLSM variables:

$$\mathbf{x}_{i,t}^j = [\text{catdef}, \text{swe}]_{i,t}^{j\prime} \quad \text{Equation (3)}$$

where “ $\prime$ ” denotes the vector transpose. In previous GRACE assimilation studies (Forman, et al., 2012, Giroto et al., 2016, 2017; Kumar et al., 2016) *rzexc* was included as a prognostic state variable, but it only received negligible increments because it mostly varies on time scales shorter than that of the monthly TWS observations (Giroto et al., 2016). The snow water equivalent (*swe*) is updated but not further discussed, as we focus on water storage in the soil column.

The observation vector $\mathbf{Z}_{I,T}^j$ contains perturbed (Burgers et al., 1998) monthly, thinned TWS retrievals (at their native resolution of 300-400-km; Section 3.1.1). The observation errors account for instrument errors and representativeness errors (such as model or geolocation imperfections). We assume a spatially constant observation error variance of 15^2 mm^2 , correlation length of 2° and observation localization radius (or radius of influence) of $r_l=6^\circ$. Prior to assimilation, the TWS observations ($\mathbf{Z}_{I,T}^j$) are rescaled to the mean and standard deviation of the modeled TWS (Giroto et al., 2016) to ensure an unbiased assimilation system. Note that the a priori scaling approach does not imply that the climatology of the model is more correct than that of the observations. It is simply a convenient way of addressing the need for climatological consistency between observations and simulations in a bias free data assimilation systems (Giroto et al., 2016).

The vector $\mathbf{M}(\cdot)|_{I,T}$ contains the TWS model estimates that correspond to the

observations $\mathbf{Z}_{I,T}$. These observation predictions are computed on the (thinned) TWS GRACE grid (i.e., every 3°) from the modeled monthly TWS estimates from the 36-km model grid, using a Gaussian spatial averaging function having a 300-km full width at half maximum (to approximate the spatial resolution of the GRACE observations). The vector $\mathbf{M}(\cdot)|_{I,T}$ collects all of the observation predictions within an influence radius of $r_I=6^\circ$ around the state vector at location i . Temporally, the model estimates at the 7.5-minute simulation time step are aggregated to monthly TWS estimates. This is the reason why the GRACE assimilation method is a “two-step” scheme in which the same month executes the CLSM integration twice: first (Run 1) to collect monthly TWS observation-minus-forecast residuals (i.e., innovations), and a second time (Run 2) to update the CLSM prognostic states by applying the increments derived from the observation-minus-forecast residuals analysis obtained in the first integration. The calculation and application of increments follows three steps as described in Girotto et al. (2016). We first compute $\Delta\mathbf{x}$ values (**Equation 1**) for each day of the month; then, we take the average of the daily increments; finally, we apply this monthly averaged increment at the beginning of the month for Run 2 (i.e., effectively adjusting the initial model conditions for that month). A summary of GRACE DA parameters is provided in **Table 3**.

4.3. SMOS Tb Assimilation (SMOS DA)

Details of the assimilation of SMOS (SMOS DA) are described in De Lannoy & Reichle (2016a). Thus, only key points and differences from the previously described work are included in this section. For SMOS Tb assimilation, the increments are calculated for the following CLSM prognostic variables:

$$\mathbf{x}_{i,t}^j = [srfexc, rzexc, tc1, tc2, tc4, ght1]_{i,t}^j, \quad \text{Equation (4)}$$

Note that the CLSM variable “*catdef*” is not included in this state vector, unlike in (De Lannoy & Reichle, 2016a). Here, we exclude *catdef* from the SMOS state vector because *catdef* is more representative of the terrestrial water storage and is therefore included in the state vector for the GRACE assimilation. In any case, we expect SMOS observations to be mostly affecting the surface and the rootzone soil moisture.

The observation vector $\mathbf{Z}_{I,T}^j$ contains perturbed (Burgers et al., 1998) SMOS observations at both horizontal and vertical polarization. We consider a spatially and temporally constant observation error variance of 5^2 K^2 for both polarizations, a spatial error correlation length of 0.25° and an observation localization radius $r_l=1^\circ$. Prior to assimilation, the persistent seasonally varying Tb bias between the model observation predictions and the SMOS observations is removed from the innovations, to ensure an unbiased assimilation system (De Lannoy & Reichle, 2016a).

The observation operator translates the model CLSM state variables ($\mathbf{x}_{i,t}^j$ of **Equation 4**) into the Tb observation predictions $\mathbf{M}(\cdot)|_{I,T}$. using a calibrated tau-omega radiative transfer model (Section 3.2). Observation-minus-forecast residuals are collected within time windows of three hours (if any are available in either ascending or descending overpasses) and they are used to update the central model time step instantaneously. A summary of SMOS DA parameters is provided in **Table 3**.

4.4. Multi-sensor GRACE and SMOS Data Assimilation (GRACE+SMOS DA)

The multi-sensor assimilation (GRACE+SMOS DA) combines the concepts of the single-sensor GRACE DA (Section 4.2) and SMOS DA (Section 4.3) as illustrated in the simplified flowchart of Figure 1. Following the two-step procedure of the GRACE DA, we first compute the increments from the GRACE TWS analysis after integrating the model for

one month, then rewind the model and restart it with an updated initial condition based on the TWS analysis, and finally assimilate the SMOS Tb observations.

More specifically, we first conduct a one-month forecast ensemble integration, Run 1, which serves only to produce the GRACE TWS predictions needed for the calculation of the TWS increments. During Run 1, the SMOS Tb observations are provisionally assimilated as in the single-sensor Tb assimilation case. This provisional SMOS Tb assimilation ensures that the GRACE TWS analysis only corrects for errors in the modeled TWS that are not eventually corrected by the assimilation of SMOS Tb observations in Run 2 (see below). As in the single-sensor TWS assimilation case, the GRACE TWS analysis is conducted after Run 1. That is, we first compute the TWS observation predictions (i.e., the sum of the modeled groundwater, surface and rootzone soil moisture, snow, and water stored in the canopy) by aggregating the modeled TWS to the spatial and temporal resolution of the GRACE TWS observations (i.e., using a monthly temporal aggregation and a Gaussian smoothing spatial average function with a 300-km full width at half maximum ground distance). Thereafter, we calculate the increments that are generated by the GRACE TWS analysis.

The next step is to rewind the model to the beginning of the month and conduct a second simulation of the same month (Run 2). To this end, we first apply the GRACE increments as in the single-sensor GRACE DA. That is, we start Run 2 from the initial conditions that was used in Run 1 but updated with the increments from the GRACE TWS analysis. Note that this initial condition has seen the SMOS Tb observations only indirectly through their impact on the increments generated in the TWS analysis. Finally, we reintegrate the model for the same month while performing the SMOS DA throughout the month (Run 2).

The two-step procedure described above is then repeated for the following month, and so on. For each month, the calculation of the TWS and Tb analysis increments follows

477 exactly the same procedure described for the single-sensor GRACE DA and SMOS DA
478 assimilation (Sections 4.2. and 4.3). The only differences between the single-sensor
479 assimilation and the multi-sensor assimilation are that in the multi-sensor system (i) the
480 GRACE TWS and SMOS Tb observations are jointly assimilated, with the analysis of one
481 observation type impacting all subsequent analyses (including those of the other
482 observation type) through the model dynamics; and (ii) the SMOS DA is executed during
483 Run 1 as well as during Run 2.

484 < Figure 1 here please >

5. Results

5.1. Validation with In-situ Measurements

In this section, we quantify the improvements in hydrological estimates introduced by the various satellite data assimilation experiments through validation with monthly averaged in situ measurements (statistics obtained from daily timeseries lead to similar results). **Figure 2**, column one, shows maps for the correlation skill vs. independent measurements of surface soil moisture, rootzone soil moisture, groundwater, runoff, and TWS for the open-loop case. Skill differences for the assimilation cases compared to the open-loop (i.e., assimilation skill minus open-loop skill) are shown in **Figure 2**, columns two-to-four. Note that the skill against GRACE TWS is an independent validation for the open-loop and the single-sensor SMOS DA case but not for the GRACE DA and GRACE+SMOS DA cases, where TWS observations are also used in the assimilation process. The summarized bulk R and $ubRMSD$ values for the in-situ validation are shown in **Table 5** and visualized in **Figure 3**.

5.1.1. Open-Loop Model Performance

At most of the surface and rootzone soil moisture locations, the model has a positive R skill, except for a -0.64 value corresponding to an outlier for rootzone soil moisture skill at a USCRN location near Albuquerque, New Mexico (**Figure 2b**). The skill values are generally higher for the surface soil moisture than for the rootzone. In fact, the surface soil moisture domain-averaged R open-loop (or model-only) skill values are $0.73(\pm 0.03)$ at the sparse network and 0.82 at the Cal/Val sites (**Table 5**). Note that the 95% confidence intervals (CI) are provided in parenthesis for select metrics (Section 3.4). For the rootzone soil moisture, the skill values are $0.68(\pm 0.03)$ at the sparse network and 0.77 at the Cal/Val sites (**Table 5**). Similarly, the average surface soil moisture $ubRMSD$ values are $0.035(\pm 0.002)$ m^3m^{-3} at

the sparse network sites and $0.027 \text{ m}^3\text{m}^{-3}$ at the Cal/Val sites (**Table 5**). For the rootzone soil moisture, these are $0.035(\pm 0.002) \text{ m}^3\text{m}^{-3}$ and $0.027 \text{ m}^3\text{m}^{-3}$ (**Table 5**). There is no obvious connection between the spatial pattern in the R skill and the patterns in terrain or climate, but the correlation patterns for the surface soil moisture correspond to those for the rootzone soil moisture (e.g., the patterns of light blue markers in **Figure 2a** and dark blue markers in **Figure 2b** are similar).

The average R (CI) and $ubRMSD$ (CI) values are $0.63(\pm 0.01)$ and $57.8(\pm 1.00) \text{ mm}$, respectively, for the groundwater validation (**Table 5**). The eastern U.S. has generally high groundwater correlation skill, except for some lower skill values in the northeastern U.S. (New England). Skill values are also lower in the western U.S. (**Figure 2c**). In New England, model and groundwater observations have different seasonal modalities (Giroto et al., 2016). In the western U.S., the aquifers tend to be deeper so it is likely that the shallow groundwater scheme of the model is unable to reproduce the water table dynamics accurately.

For runoff, the average R (CI) and $ubRMSD$ (CI) are $0.62(\pm 0.02)$ and $0.82(\pm 0.02) \text{ mm/day}$, respectively (**Table 5**). The correlations are higher in the eastern U.S. and in the western coastal regions. Lower values of runoff R are seen in the mid-western U.S. where runoff magnitudes and temporal variabilities are small, thus making it more difficult for the model to predict runoff dynamics (**Figure 2d**). Finally, the average TWS R and $ubRMSD$ values are 0.72 and 46.8 mm , respectively (**Table 5**), with lower model skill values in regions where the seasonal cycle is the weakest (e.g., blue areas in the south west regions, Arizona and New Mexico, **Figure 2e**).

5.1.2. GRACE TWS Assimilation (GRACE DA)

More than half of the soil moisture sites (both surface and rootzone) experience degradation due to GRACE DA (i.e., the red colored locations outnumber the blue colored ones in **Figures 2f** and **2g**). The average surface soil moisture skill values are similar to those of the open-loop experiment (**Table 5**). The GRACE DA improves most of the groundwater sites (**Figure 2h, Table 5**), although some degradation is visible, for example in Arizona and in the New England region. For runoff, the average R (CI) and $ubRMSD$ (CI) skill values are $0.61(\pm 0.02)$ and $0.76(\pm 0.02)$ mm/day, respectively (**Table 5**). The runoff correlation skill improves in the northwest U.S., but it degrades around the Great Lakes region, such as for the measurement sites in Illinois and Wisconsin (**Figure 2i**). The evaluation of TWS in the GRACE DA is shown in **Figure 2j**. It is important to note that this is not an independent evaluation (because TWS data were assimilated in the GRACE DA experiment), yet it is an internal check to confirm that the GRACE DA works as expected.

In summary, GRACE DA leads to statistically significant improvements in simulated groundwater (both in terms of R and $ubRMSD$ metrics, **Figure 3**). For runoff, GRACE DA leads to a degradation in R (albeit not statistically significant) but produces statistically significant improvements in $ubRMSD$ (**Figure 3**). The assimilation of GRACE observations leads to mixed results with respect to surface and rootzone soil moisture accuracy, with a slight (non-significant) degradation on average (**Figure 3**). These results are consistent with Girotto et al. (2016).

5.1.3. SMOS Tb Assimilation (SMOS DA)

SMOS DA leads to improvements in both surface and rootzone soil moisture estimates (**Table 5**). In fact, most of the markers in **Figure 2k** and **Figure 2l** are colored blue (improvement). SMOS assimilation has little effect on the correlation skill of groundwater or runoff (**Figure 2m** and **Figure 2n, Table 5**). Minimal changes in the

groundwater skill were expected by design, because *catdef* is not included in the state update. The SMOS DA domain average R (CI) and *ubRMSD* (CI) TWS skill values are 0.71(± 0.01) and 45.0(± 0.01) mm, respectively. The assimilation of SMOS Tb leads to degraded TWS correlation skill values in the southern part of the high plains (i.e., east Texas, **Figure 2o**). For this region, the assimilation of SMOS leads to a change in the seasonality and a different trend in simulated TWS.

In summary, SMOS DA introduces statistically significant improvements in simulated surface soil moisture, in comparison with data from the USCRN and SCAN sparse network sites (**Figure 3**). Improvements also are seen (even if not statistically significant) at the surface soil moisture Cal/Val locations (**Figure 3**). SMOS DA leads to some improvements (although not statistically significant) in the rootzone soil moisture and does not lead to significant changes in groundwater or runoff (**Figure 3**). However, a small degradation can be seen in the TWS evaluation (**Figure 2o**). These results are consistent with De Lannoy & Reichle (2016a) where the assimilation of SMOS was deemed beneficial for surface and rootzone soil moisture estimation.

5.1.4. Multi-sensor GRACE and SMOS Data Assimilation (GRACE+SMOS DA)

Results for the multi-sensor GRACE+SMOS DA experiment (**Figure 2**, column four) reflect a combination of aspects related to each of the single-sensor assimilation experiments (i.e., GRACE DA, column two and SMOS DA, column three), described above.

For the most part, the multi-sensor GRACE+SMOS DA maintains the improvements in surface soil moisture that were introduced by the SMOS DA (compare **Figure 2k** with **Figure 2p**). The average R values are 0.77(± 0.02) for the surface soil moisture at the sparse network sites and 0.86 at the Cal/Val sites (**Table 5**). For *ubRMSD* these are 0.035(± 0.001) m^3m^{-3} at the sparse network sites and 0.023 m^3m^{-3} at the Cal/Val sites (**Table 5**). These

statistics are comparable to the SMOS DA experiment. For the rootzone, the domain average R skill values are $0.69(\pm 0.03)$ for the sparse network sites and 0.76 for the Cal/Val sites (**Table 5**). The $ubRMSD$ skill values are $0.034(\pm 0.002)$ m^3m^{-3} for the sparse network sites and 0.027 m^3m^{-3} for the Cal/Val sites (**Table 5**). The rootzone soil moisture is degraded at some locations (i.e., red markers in **Figure 2q**), similar to those reported for the GRACE DA (compare **Figure 2q** with **Figure 2g**). Overall, though, the number of improved rootzone soil moisture sites is larger than the number of degraded sites. In other words, the benefits provided by the SMOS DA to simulated surface and rootzone soil moisture (Section 4.3) are maintained in the multi-sensor GRACE+SMOS DA experiment.

For groundwater, the domain average R (CI) and $ubRMSD$ (CI) values are $0.65(\pm 0.01)$ and $55.1(\pm 0.95)$ mm, respectively (**Table 5**). These statistics are similar to those reported in the GRACE DA experiment (compare **Figure 2c** to **Figure 2h**), and most sites experience improved correlation skill R with respect to the open-loop simulation (**Table 5**). Similar conclusions can be stated for the runoff validation, where most of the GRACE+SMOS DA skill differences resemble those of the GRACE DA experiment (compare **Figure 2s** to **Figure 2i**). Finally, while not an independent validation, the TWS R skill improvements are all positive in the multi-sensor GRACE+SMOS DA (**Figure 2t**). This suggests that multi-sensor GRACE+SMOS DA is able to overcome the TWS skill degradation caused by SMOS DA (**Figure 2o**), meaning that the assimilation of GRACE observations is more influential than that of the SMOS observations for estimating TWS.

< Figure 2 here please >

< Figure 3 here please >

< Table 5 here please >

5.2. Assimilation Systems Diagnostics

In this section, we analyze two diagnostics provided by the assimilation system to gain a better understanding of the impact of the observations on the entire soil profile. Specifically, we study the ensemble spread and the increments in various hydrological variables.

5.2.1. Changes in the Ensemble Spread

We define “ensemble spread” as the long-term (July 2010 – June 2016) time-average of the instantaneous ensemble standard deviation. **Figure 4** maps the ensemble spread of the open-loop experiment (column one) and the reduction in spread (i.e., the ensemble spread in the assimilation experiment minus that in the open-loop experiment) in the various assimilation experiments (columns two-to-four) for surface soil moisture, rootzone soil moisture and groundwater (that is, the model prognostic *catdef*).

For surface soil moisture, the mean ensemble spread over the CONUS region is $0.027 \text{ m}^3\text{m}^{-3}$ (**Figure 4a**). The largest ensemble spread is located in drier parts of the north-western U.S., including Montana, Wyoming, Idaho, Utah, and Nevada. The lowest ensemble spread is found for the sandy soil regions in Nebraska. The mean ensemble spread for rootzone soil moisture is $0.024 \text{ m}^3\text{m}^{-3}$ (**Figure 4b**). Large values for the rootzone soil moisture ensemble spread are co-located with regions where the surface soil moisture uncertainty values were also the largest (i.e., Montana, Wyoming, Idaho). The mean ensemble spread for *catdef* is 46 mm (**Figure 4c**). Large values for the *catdef* ensemble spread correspond to the High Plains regions. While, the rootzone soil moisture ensemble spread is small in the High Plains regions.

The reduction in ensemble spread caused by GRACE DA equals $0.004 \text{ m}^3\text{m}^{-3}$ for surface soil moisture, $0.006 \text{ m}^3\text{m}^{-3}$ for rootzone soil moisture and 13 mm for *catdef* (**Figure**

4d, **4e**, and **4f**, respectively). For all variables, the largest reductions (note that the assimilation reduces uncertainties by design) correspond to regions where the open-loop ensemble spread is the largest (compare dark blue regions in **Figure 4f** with orange areas in **Figure 4c**). GRACE DA reduces the ensemble spread in surface soil moisture (**Figure 4d**). Similarly, the SMOS DA leads to spatially averaged reductions of $0.004 \text{ m}^3\text{m}^{-3}$ for surface soil moisture, $0.002 \text{ m}^3\text{m}^{-3}$ for rootzone soil moisture and 4 mm for *catdef*. The largest reductions (expected by design) in the surface soil moisture ensemble spread are found in the central CONUS and along the Mississippi River, where the open-loop had a large ensemble spread (compare dark blue in **Figure 4g** with orange in **Figure 4a**).

Finally, the reduction in ensemble spread in the multi-sensor GRACE+SMOS DA experiment equals $0.007 \text{ m}^3\text{m}^{-3}$ for surface and rootzone soil moisture and 15 mm for *catdef*, respectively (**Figure 4** column four). Multi-sensor GRACE+SMOS DA leads to a larger reduction in ensemble spread than either single-sensor assimilation case alone (compare **Figure 4** column four with columns two and three), since the number of assimilated observations is larger and the model is therefore more strongly constrained.

< Figure 4 here please >

Figure 5 plots time-averaged ensemble standard deviation (or ensemble spread) against the actual errors (*ubRMSD*) for each in-situ site used in the validation of surface soil moisture, rootzone soil moisture and groundwater (for the sites in **Figure 2**). Only the times when in-situ measurements are available are included in the calculation of the time-average ensemble standard deviation. The probability distributions of the scatter points are also shown for each of the experiments (on the right side of **Figure 5**). In general, the estimated ensemble standard deviation underestimates the *ubRMSD* for both soil moisture and groundwater variables in all experiments (i.e., the scattered points are typically located below the 1:1 line, **Figure 5a-c**).

The assimilation experiments produce smaller average ensemble spreads than the open-loop experiment for soil moisture and groundwater (**Figure 5a-c**), as expected. The smallest ensemble spread corresponds to the multi-sensor GRACE+SMOS DA case, also as expected (i.e., the red distribution curves have the lowest mean values, right side of **Figure 5a-c**). For surface soil moisture, the shape of the distribution of the ensemble spread in the multi-sensor GRACE+SMOS DA experiment is similar to that of the SMOS DA case (i.e., the blue distribution is similar to the red distribution on the right side of the scatter plot, **Figure 5a**). Similarly, the shape of the *catdef* ensemble spread distribution for the multi-sensor GRACE+SMOS DA case is similar to that of the GRACE DA case (i.e., compare dotted grey lines with red line in **Figure 5c**). These results corroborate those derived from **Figure 4**.

< Figure 5 here please >

5.2.2 Assimilation increments

Figure 6 shows example time series of the increments for the model prognostic variables associated with soil moisture and groundwater as derived from the GRACE DA, SMOS DA and GRACE+SMOS DA experiments. This specific example corresponds to a location in Iowa (see **Figure 7** for the location). The location was chosen because it is representative of an interesting anti-correlation (shown in **Figure 7** and explained later in this Section). For this site, the catchment deficit increments are typically negative during the summer months (**Figure 6a**), meaning that GRACE DA wants to decrease the deficit in the catchment (i.e., add water). The long-term average of the increments is about zero by design because the GRACE TWS observations are rescaled to the model climatology prior to assimilation (Section 4.2).

The SMOS DA algorithm calculates increments (dots in **Figure 6b**) only during the months of the year when the ground is not frozen or snow-covered (De Lannoy & Reichle, 2016b). Increments of surface and rootzone excess vary between positive and negative values within the course of one single month, but their long-term average is about zero because of the innovations bias correction (Section 4.3). **Figure 6** also shows the 14-day running average of the *srfexc* and *rzexc* increments (dashed lines in **Figure 6b**) to help assess the overall increment dynamics. In general, increments in the *srfexc* and *rzexc* are in agreement in terms of sign.

The results for the multi-sensor GRACE+SMOS DA are shown in **Figure 6c**. The *catdef* increments (introduced by the assimilation of GRACE TWS observations) are consistent with those calculated by the single-sensor GRACE DA experiment (**Figure 6a**). Furthermore, the temporal patterns in increments for *srfexc* and *rzexc* (due to the assimilation of SMOS Tb observations) are similar to those in the single-sensor SMOS DA experiment (compare blue and green dashed lines in **Figure 6b** with **Figure 6c**). This suggests that the overall characteristics of each single-sensor assimilation case are maintained in the multi-sensor GRACE+SMOS DA. The differences in spatial and temporal error scales (Section 3.3) of the GRACE TWS and SMOS Tb observations explains why the magnitudes and signs of the DA increments generally are preserved in the multi-sensor GRACE+SMOS DA. These differences allow the multivariate DA to influence individual, but connected, vertical levels of soil moisture and groundwater.

< Figure 6 here please >

An interesting aspect of the multi-sensor GRACE+SMOS DA is that only in a few instances *catdef* increments (introduced by GRACE TWS observations) and *srfexc* and *rzexc* increments (introduced by SMOS Tb observations) both add (or both subtract) water for the same month. In other words, there is an anti-correlation between water storage increments

introduced by GRACE TWS and SMOS Tb DA. **Figure 7a** shows, for the multi-sensor GRACE+SMOS DA, the time series correlation between the *catdef* and the *srfexc* increments averaged over the first 14 days of each month. The figure demonstrates that the *srfexc* increments are indeed positively correlated with the *catdef* increments. A similar, albeit weaker, correlation is present between *catdef* and *rzexc* (**Figure 7b**). We refer to this positive correlation as “anti-correlation” because *catdef* is expressed in terms of “deficit”, whereas *srfexc* is an “excess” quantity. That is, the assimilation of GRACE removes water by increasing the *catdef* while the assimilation of SMOS adds water by increasing the *srfexc* and *rzexc*. It should be noted that this anticorrelation is somewhat weaker when the increments are averaged over the entire month before computing the correlation.

< Figure 7 here please >

Finally, **Figure 8** shows the typical magnitude of the increments obtained from each assimilation experiment as measured by the time series standard deviation of the increments. The typical magnitudes of the single-sensor SMOS DA increments for *srfexc* and *rzexc* are 0.28 mm and 0.22 mm (**Figure 8a** and **Figure 8c**). In general, larger magnitudes are reported for the central part of the CONUS region (i.e., the Great Plains) and smaller increments for the eastern and western U.S. The GRACE DA increments are shown in the **Figure 8e**, with typical *catdef* increments equal to 15.55 mm. Larger magnitudes are seen in the northwestern U.S., the Great Lakes region, and the southern U.S.

The multi-sensor GRACE+SMOS DA experiment applies increments to the water storage components in the entire soil column (*srfexc*, *rzexc*, and *catdef*). The spatial patterns in the *srfexc* and *rzexc* increments are consistent with those of the single-sensor SMOS DA experiment (compare **Figure 8a** with **Figure 8b** and **Figure 8c** with **Figure 8d**). Finally, the typical *catdef* increment for the multi-sensor GRACE+SMOS DA equals 16.76 mm (**Figure 8f**). This magnitude is slightly higher than for the single-sensor GRACE DA (**Figure 8e**).

733

< Figure 8 here please >

6. Discussion

6.1. Validation with In-situ Measurements

Consistent with the work by Giroto et al., (2016); Kumar et al., (2016); Li et al., (2012); and Tian et al., (2017), GRACE DA leads to statistically significant improvements in simulated groundwater and runoff. However, some of the groundwater and runoff sites experience degradation from the GRACE DA (e.g., red markers in **Figure 2h** and **Figure 2i**). In general, degradations seen in **Figure 2** depend upon three aspects: 1) the model, 2) the assimilated or in-situ observations; 3) the assimilation scheme. For example, the degradation in groundwater skill in New England was also reported in Giroto et al. (2016) and is likely associated with the inability of the model to represent the seasonality of the measured water table. Another example is the site in Arizona (red marker in **Figure 2h**) with poor open-loop R skill possibly caused by a negative trend in the observed groundwater versus a positive trend in the open-loop simulation. The latter was corrected by the GRACE assimilation but this also resulted in a poor representation of the groundwater seasonal cycle and a net skill degradation. Yet another example are the degraded runoff estimates in the Great Lakes, where GRACE DA reduces the peak runoff values, which were better represented in the open-loop experiment. It is possible that this loss of skill is caused by the lack of a river-routing scheme in our model. Finally, the GRACE DA does not directly compute increments for the surface and rootzone soil moisture variable, but these are updated via their model connections between the *catdef* and *swe* prognostics. Thus, the degradation seen in **Figure 2f** and **Figure 2g** could be due to limitations in our model equations that transfer storage information. This is a common problem of other land surface models (Schumacher et al., 2018; Müller Schmied et al., 2014).

The results from the SMOS DA experiment generally agree with those of previous soil moisture assimilation studies (De Lannoy & Reichle, 2016a, 2016b; Lievens et al., 2015; Reichle et al., 2007). That is, the assimilation of SMOS helps improve the surface and rootzone soil moisture estimates. For a few sites with a loss of skill (red colored markers in **Figure 2k** and **Figure 2l**), the degradation can be attributed to the erroneous quantification of vegetation in the radiative transfer model or inadequate horizontal and vertical propagation of the SMOS update information (De Lannoy & Reichle, 2016b). The SMOS DA introduces some degradation of the TWS estimates. We attribute this degradation partly to model deficiencies and partly to the assimilation scheme. In fact, there could be inefficient model connections between water storages (Schumacher et al., 2018; Müller Schmied et al., 2014). That is, changes in modeled surface soil moisture might not translate correctly into TWS changes. For example, for a location in Texas, the open loop TWS estimates indicate a positive trend (approximately starting in July 2012) in response to increased model precipitation, corresponding to the recovery period from the severe 2011 drought (Scanlon et al., 2013). The TWS trend is exacerbated in the SMOS DA case, but it is not present in the GRACE TWS observations (Figure not shown). Thus, suggesting that the TWS model may be responding quicker than reality to the increased (model) precipitation. Furthermore, we currently update surface and rootzone soil moisture every three hours using a filtering approach (provided SMOS observations are available). To avoid the jumps characteristic of filtering approaches, a better choice might be a smoother method (Evensen et al., 2000), where SMOS observations are assimilated over a temporal window that includes SMOS observations from multiple overpasses. Tian et al., (2017) indeed showed less degradation in groundwater when assimilating SMOS data using a smoother assimilation method with a temporal assimilation window of one month.

Finally, our results indicate that the performance statistics for surface and rootzone soil moisture, groundwater, runoff and TWS from the multi-sensor GRACE+SMOS DA experiment fall between those of the GRACE DA and the SMOS DA experiments. In fact, some sites still experience degraded surface soil moisture estimates. For example, a site in North Carolina has degraded correlation skill values with respect to the open-loop. A similar skill is seen in the GRACE DA case (compare **Figure 2f** to **Figure 2p**). For this site, the SMOS DA does not lead to improvements upon the open loop case (see **Figure 2f**). For this reason, SMOS DA does not have a large impact at this site. Thus, it is reasonable to expect GRACE+SMOS DA statistics are more similar to those from the GRACE DA than to those resulting from SMOS DA. Nonetheless, the multivariate GRACE+SMOS DA generally maintains the improvements seen in the single-sensor assimilation experiments individually. That is, the multi-sensor GRACE+SMOS DA maintains most of the improvements in surface and rootzone soil moisture from SMOS DA and the favorable groundwater and runoff results from GRACE DA (**Figure 3**). For surface soil moisture and groundwater, the improvements are statistically significant. This suggests that the best hydrology estimates are obtained when both SMOS and GRACE are jointly assimilated into the model. Our conclusions confirm those by Tian et al., (2017) and Khaki and Awange (2019).

6.2. Assimilation Systems Diagnostics

6.2.1. Ensemble Spread

We investigate the ensemble spread because it is a proxy for the uncertainty in the model (or assimilation) estimates. For the open loop case, our results indicate large uncertainties in *catdef* in the High Plains regions (**Figure 4c**). Here, the aquifers are exploited for irrigation purposes. Our model, like most other global hydrological models,

does not include these human driven processes, for this reason a higher model uncertainty is adequate. The model bedrock depth is the deepest (across the CONUS region) in the High Plains regions, so that the water table is likely uncoupled from rootzone soil moisture. This possibly explains why rootzone soil moisture estimates are instead characterized by small uncertainties (**Figure 4b**). Another low uncertainty region is located in Nebraska. Here, the soil texture is mostly sand, therefore water drains quickly from the surface downward, leading to little variability and uncertainty in surface soil moisture (**Figure 4a**).

One way to assess the impact of the observations across the entire soil column in each assimilation experiment is to analyze the reduction in the ensemble spread (estimated uncertainty) for soil moisture and groundwater, although keep in mind that the reduction of ensemble spread does not necessarily translate into better estimates when compared to in situ measurements. For example, some of the surface soil moisture statistics degrade after the assimilation of GRACE observations, possibly caused by errors in the model connections between TWS and surface soil moisture (Schumacher et al., 2018; Müller Schmied et al., 2014).

GRACE DA generates increments in the *catdef* (Section 4.2), and this impacts not only simulated groundwater but also surface and rootzone soil moisture. This is why GRACE DA reduces the ensemble spread in surface soil moisture (**Figure 4d**). SMOS DA generates increments in surface soil moisture (Section 4.3) but it also impacts the ensemble standard deviation of *catdef*. Since both single-sensor experiments generate increments that change the ensemble spread in surface soil moisture, rootzone soil moisture and groundwater (**Figure 4** columns two and three), it is not easy to separate the impacts of GRACE TWS and SMOS Tb observations in the multi-sensor GRACE+SMOS DA experiment. But, in general, the assimilation of SMOS Tb observations generally affects surface and rootzone soil moisture more than groundwater. The assimilation of GRACE TWS observations is responsible for

reducing most of the ensemble spread in groundwater (e.g., *catdef*, **Figure 4f** and **Figure 4I**). The latter results are consistent with Giroto et al., (2016) who demonstrated that GRACE-TWS assimilation is more valuable for groundwater because surface and rootzone soil moisture mostly vary on time scales that are shorter than the monthly resolution of the GRACE TWS observations.

Finally, when evaluating the reliability of an ensemble system, it is common practice to compare the actual errors (i.e., the *ubRMSD* vs. independent measurements, e.g., at the sites shown in **Figure 2**) to the time-averaged ensemble spread (Fortin et al., 2014). For a given in-situ location, a good ensemble spread should match the *ubRMSD* for that site, provided the in-situ observations and ensemble simulations are representative of the same quantity. Most of the *ubRMSD* values were calculated using point-scale in-situ measurements, whereas the ensemble standard deviation represents grid-cell-average estimates (EASEv2 at 36-km, Section 3.2). For this reason, the ensemble standard deviation is expected to be somewhat smaller. That said, all experiments (including the open loop) produce grid-based estimates of ensemble spread that underestimate the actual *ubRMSD* versus point-scale in-situ observations (**Figure 5**), most likely because of representativeness error (spatial scale) between model simulations and in-situ observations.

6.2.2. Assimilation Increments

Another way to understand how data assimilation impacts the soil moisture and groundwater estimates is by looking at the contribution of each type of observation (GRACE and/or SMOS) to the water storages at different vertical levels, i.e., via the assimilation increments (Section 4.1). In general, our results indicate that magnitude and patterns of the

increments are consistent with those reported by De Lannoy & Reichle (2016a) for *srfexc* and *rzexc* and by Girotto et al. (2016) for *catdef*.

As stated previously, the multi-sensor GRACE+SMOS DA generally maintains the improvements from the single-sensor experiments. However, this does not mean that all assimilated observations (SMOS or GRACE) contribute to increments in the same direction. For example, **Figure 3** indicates that, while the multi-sensor case generally improves the statistics, the surface soil moisture statistics at the sparse network sites (i.e., SCAN and USCRN) are best in the single-sensor SMOS DA case. That is, the performance of the multi-sensor assimilation maintains the improvements seen in the single sensor experiments (e.g. we still get improved surface soil moisture), but the improvements are not always as good as if only SMOS observations were assimilated. This is an indicator that in the multi-sensor assimilation, the GRACE TWS observations lead to surface soil moisture adjustments that disagree with those of the SMOS Tb assimilation. In fact, the results reported in Section 5.2.2 identify the presence of an anti-correlation between SMOS Tb and GRACE TWS analyses. This anti-correlation (**Figure 7**) could be caused by the SMOS Tb analysis trying to undo the increments from the TWS analysis that are applied at the beginning of each month of Run 2 (Section 2.5.3). In other words, it suggests that the increments from the GRACE TWS and SMOS Tb analysis components tend to battle each other. For example, when the increments introduced by GRACE TWS add water, the increments introduced by SMOS Tb act to remove it. This result is consistent with Tian et al., (2017) who similarly report instances of SMOS and GRACE analysis increments pulling in opposite directions over portions of Australia (see Figure 2 in Tian et al., 2017). Looking ahead, a revised assimilation scheme based on an ensemble smoother (Evensen et al., 2000) could potentially reduce this anti-correlation. This is because the assimilation state vector would be augmented to include increments

generated from both SMOS and GRACE observations within a given assimilation temporal window.

7. Conclusion

This work investigates to what extent the individual and joint assimilation of GRACE TWS and SMOS Tb observations can improve model estimates of soil moisture and shallow unconfined groundwater. In order to understand the relative benefits introduced by these two types of observations, we performed a series of single-sensor and multi-sensor satellite assimilation experiments using GRACE TWS and SMOS Tb data. The main findings can be summarized as follows:

1. The single-sensor GRACE TWS assimilation (GRACE DA) improves estimates of groundwater and runoff. The assimilation of SMOS Tb observations (SMOS DA) improves estimates of the surface, and marginally, the rootzone soil moisture.
2. Most of the benefits provided by each single-sensor assimilation case (GRACE DA or SMOS DA) are maintained in the multi-sensor assimilation experiment (GRACE+SMOS DA). Specifically, the multi-sensor assimilation leads to improved surface soil moisture, groundwater, and runoff estimation. These results therefore suggest that superior hydrological estimates can be achieved when both observation types (TWS from GRACE and Tb from SMOS) are assimilated.
3. GRACE DA is mostly responsible for reducing ensemble spread in groundwater, and SMOS DA is mostly responsible for ensemble spread reductions in surface and rootzone soil moisture.
4. For the multi-sensor GRACE+SMOS DA, there is an anti-correlation between the water storage increments introduced by the GRACE TWS observations and those introduced by the SMOS Tb observations.

905 The assimilation of GRACE TWS and SMOS Tb observations provides a baseline for
906 future science applications for which longer (and presumably more accurate) records of
907 TWS and Tb observations are available, e.g. from the recent GRACE Follow-On (GRACE-FO,
908 <https://gracefo.jpl.nasa.gov>) and SMAP missions. Finally, future work will include the use of
909 improved GRACE TWS solutions such as the MASCON solutions (e.g., Luthcke et al., 2013;
910 Save et al., 2016; Wiese et al., 2016) or daily time-scale solutions (Ramillien et al., 2015;
911 Sakumura et al., 2016). We anticipate that joint assimilation of improved GRACE and
912 GRACE-FO solutions with SMOS and SMAP observations will produce even more accurate
913 hydrological estimates.

914

915

916 **Acknowledgment**

917 The GRACE observations used here are from <http://grace.jpl.nasa.gov>, which is
918 supported by the NASA MEaSURES Program. Computational resources were provided by the
919 NASA High-End Computing Program through the NASA Center for Climate Simulation at the
920 Goddard Space Flight Center. This study was supported by the NASA Terrestrial Hydrology
921 and GRACE & GRACE Follow-On Science Team programs. We thank the five anonymous
922 reviewers and the RSE editorial board.

923

924

References

- Bolten, J. D., Crow, W. T., Zhan, X., Reynolds, C. A., & Jackson, T. J. (2009). Assimilation of a satellite-based soilmoisture product into a two-layer water balance model for a global crop production decision support system. In *Data assimilation for atmospheric, oceanic and hydrologic applications* (pp. 449–463). Springer.
- Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., & Hasenauer, S. (2010). Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrology and Earth System Sciences*, 14(10), 1881–1893.
- Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., & Savoie, M. H. (2012). EASE-Grid 2.0: Incremental but significant improvements for Earth-gridded data sets. *ISPRS International Journal of Geo-Information*, 1(1), 32–45.
- Burgers, G., Jan van Leeuwen, P., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126(6), 1719–1724.
- Cosh, M. H., Jackson, T. J., Moran, S., & Bindlish, R. (2008). Temporal persistence and stability of surface soil moisture in a semi-arid watershed. *Remote Sensing of Environment*, 112(2), 304–313.
- Crow, W. T., Chen, F., Reichle, R. H., & Liu, Q. (2017). L band microwave remote sensing and land data assimilation improve the representation of prestorm soil moisture conditions for hydrologic forecasting. *Geophysical Research Letters*, 44(11), 5495–5503.
- Crow, W. T., Chen, F., Reichle, R. H., Xia, Y., & Liu, Q. (2018). Exploiting Soil Moisture, Precipitation, and Streamflow Observations to Evaluate Soil Moisture/Runoff Coupling in Land Surface Models. *Geophysical Research Letters*, 45(10), 4869–4878.

- 948 De Lannoy, G. J., & Reichle, R. H. (2016a). Assimilation of SMOS brightness temperatures or
949 soil moisture retrievals into a land surface model. *Hydrology and Earth System*
950 *Sciences*, 20(12), 4895.
- 951 De Lannoy, G. J., & Reichle, R. H. (2016b). Global assimilation of multiangle and
952 multipolarization SMOS brightness temperature observations into the GEOS-5
953 catchment land surface model for soil moisture estimation. *Journal of*
954 *Hydrometeorology*, 17(2), 669–691.
- 955 De Lannoy, G. J., Reichle, R. H., Houser, P. R., Arsenault, K. R., Verhoest, N. E., & Pauwels, V. R.
956 (2010). Satellite-scale snow water equivalent assimilation into a high-resolution
957 land surface model. *Journal of Hydrometeorology*, 11(2), 352–369.
- 958 De Lannoy, G. J., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E., &
959 Pauwels, V. (2012). Multiscale assimilation of Advanced Microwave Scanning
960 Radiometer–EOS snow water equivalent and Moderate Resolution Imaging
961 Spectroradiometer snow cover fraction observations in northern Colorado. *Water*
962 *Resources Research*, 48(1).
- 963 De Lannoy, G. J., Reichle, R. H., & Pauwels, V. R. (2013). Global calibration of the GEOS-5 L-
964 band microwave radiative transfer model over nonfrozen land using SMOS
965 observations. *Journal of Hydrometeorology*, 14(3), 765–785.
- 966 De Lannoy, G. J., Reichle, R. H., & Vrugt, J. A. (2014). Uncertainty quantification of GEOS-5 L-
967 band radiative transfer model parameters using Bayesian inference and SMOS
968 observations. *Remote Sensing of Environment*, 148, 146–157.
- 969 De Lannoy, G. J., Reichle, R. H., Peng, J., Kerr, Y., Castro, R., Kim, E. J., & Liu, Q. (2015).
970 Converting between SMOS and SMAP level-1 brightness temperature observations
971 over nonfrozen land. *IEEE Geoscience and Remote Sensing Letters*, 12(9), 1908–1912.

- 972 Diamond, H. J., Karl, T. R., Palecki, M. A., Baker, C. B., Bell, J. E., Leeper, R. D., ... Helfert, M. R.
 973 (2013). US Climate Reference Network after one decade of operations: Status and
 974 assessment. *Bulletin of the American Meteorological Society*, 94(4), 485–498.
- 975 Draper, C. S., Mahfouf, J.-F., & Walker, J. P. (2011). Root zone soil moisture from the
 976 assimilation of screen-level variables and remotely sensed soil moisture. *Journal of*
 977 *Geophysical Research: Atmospheres*, 116(D2).
- 978 Ducharne, A., Koster, R. D., Suarez, M. J., Stieglitz, M., & Kumar, P. (2000). A catchment-based
 979 approach to modeling land surface processes in a general circulation model: 2.
 980 Parameter estimation and model demonstration. *Journal of Geophysical Research:*
 981 *Atmospheres*, 105(D20), 24823–24838.
- 982 Dziubanski, D. J., & Franz, K. J. (2016). Assimilation of AMSR-E snow water equivalent data
 983 in a spatially-lumped snow model. *Journal of Hydrology*, 540, 26–39.
- 984 Eicker, A., Schumacher, M., Kusche, J., Döll, P., & Schmied, H. M. (2014). Calibration/data
 985 assimilation approach for integrating GRACE data into the WaterGAP Global
 986 Hydrology Model (WGHM) using an ensemble Kalman filter: First results. *Surveys in*
 987 *Geophysics*, 35(6), 1285–1309.
- 988 Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., ...
 989 Johnson, J. (2010a). The soil moisture active passive (SMAP) mission. *Proceedings of*
 990 *the IEEE*, 98(5), 704–716.
- 991 Entekhabi, D., Reichle, R. H., Koster, R. D., & Crow, W. T. (2010b). Performance metrics for
 992 soil moisture retrievals and application requirements. *Journal of*
 993 *Hydrometeorology*, 11(3), 832–840.
- 994 Entekhabi, D., Yueh, S., O'Neill, P. E., Kellogg, K. H., Allen, A., Bindlish, R., ... Crow, W. T.
 995 (2014). SMAP Handbook–Soil Moisture Active Passive: Mapping Soil Moisture and
 996 Freeze/Thaw from Space.

- 997 Evensen, G., & Van Leeuwen, P. J. (2000). An ensemble Kalman smoother for nonlinear
 998 dynamics. *Monthly Weather Review*, 128(6), 1852-1867.
- 999 Felfelani, F., Wada, Y., Longuevergne, L., & Pokhrel, Y. N. (2017). Natural and human-induced
 1000 terrestrial water storage change: A global analysis using hydrological models and
 1001 GRACE. *Journal of Hydrology*, 553, 105–118.
- 1002 Forman, B. A., Reichle, R. H., & Rodell, M. (2012). Assimilation of terrestrial water storage
 1003 from GRACE in a snow-dominated basin. *Water Resources Research*, 48(1).
- 1004 Fortin, V., Abaza, M., Anctil, F., & Turcotte, R. (2014). Why should ensemble spread match
 1005 the RMSE of the ensemble mean? *Journal of Hydrometeorology*, 15(4), 1708–1713.
- 1006 Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., ... Reichle, R. (2017).
 1007 The modern-era retrospective analysis for research and applications, version 2
 1008 (MERRA-2). *Journal of Climate*, 30(14), 5419–5454.
- 1009 Giroto, M., Margulis, S. A., & Durand, M. (2014). Probabilistic SWE reanalysis as a
 1010 generalization of deterministic SWE reconstruction techniques. *Hydrological*
 1011 *Processes*, 28(12), 3875–3895.
- 1012 Giroto, M., De Lannoy, G. J., Reichle, R. H., & Rodell, M. (2016). Assimilation of gridded
 1013 terrestrial water storage observations from GRACE into a land surface model. *Water*
 1014 *Resources Research*, 52(5), 4164–4183.
- 1015 Giroto, M., De Lannoy, G. J., Reichle, R. H., Rodell, M., Draper, C., Bhanja, S. N., & Mukherjee,
 1016 A. (2017). Benefits and pitfalls of GRACE data assimilation: A case study of
 1017 terrestrial water storage depletion in India. *Geophysical Research Letters*, 44(9),
 1018 4107–4115.
- 1019 Houborg, R., Rodell, M., Li, B., Reichle, R., & Zaitchik, B. F. (2012). Drought indicators based
 1020 on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial
 1021 water storage observations. *Water Resources Research*, 48(7).

- 1022 Jackson, T. J., Cosh, M. H., Bindlish, R., Starks, P. J., Bosch, D. D., Seyfried, M., ... Du, J. (2010).
1023 Validation of advanced microwave scanning radiometer soil moisture products.
1024 *IEEE Transactions on Geoscience and Remote Sensing*, 48(12), 4256–4272.
- 1025 Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J., Font, J., & Berger, M. (2001). Soil
1026 moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission.
1027 *IEEE Transactions on Geoscience and Remote Sensing*, 39(8), 1729–1735.
- 1028 Khaki, M., Schumacher, M., Forootan, E., Kuhn, M., Awange, J. L., & van Dijk, A. (2017).
1029 Accounting for spatial correlation errors in the assimilation of GRACE into
1030 hydrological models through localization. *Advances in Water Resources*, 108, 99–112.
- 1031 Khaki, M., & Awange, J. (2019). The application of multi-mission satellite data assimilation
1032 for studying water storage changes over South America. *Science of the Total*
1033 *Environment*, 647, 1557-1572.
- 1034 Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., & Kumar, P. (2000). A catchment-based
1035 approach to modeling land surface processes in a general circulation model: 1.
1036 Model structure. *Journal of Geophysical Research: Atmospheres*, 105(D20), 24809–
1037 24822.
- 1038 Koster, R. D., Brocca, L., Crow, W. T., Burgin, M. S., & De Lannoy, G. J. (2016). Precipitation
1039 estimation using L-band and C-band soil moisture retrievals. *Water Resources*
1040 *Research*, 52(9), 7213–7225.
- 1041 Koster, R. D., Liu, Q., Mahanama, S. P., & Reichle, R. H. (2018). Improved Hydrological
1042 Simulation Using SMAP Data: Relative Impacts of Model Calibration and Data
1043 Assimilation. *Journal of Hydrometeorology*, 19(4), 727–741.
- 1044 Kumar, S. V., Peters-Lidard, C. D., Mocko, D., Reichle, R., Liu, Y., Arsenault, K. R., ... Livneh, B.
1045 (2014). Assimilation of remotely sensed soil moisture and snow depth retrievals for
1046 drought estimation. *Journal of Hydrometeorology*, 15(6), 2446–2469.

- 1047 Kumar, S. V., Zaitchik, B. F., Peters-Lidard, C. D., Rodell, M., Reichle, R., Li, B., ... De Lannoy, G.
 1048 (2016). Assimilation of gridded GRACE terrestrial water storage estimates in the
 1049 North American Land Data Assimilation System. *Journal of Hydrometeorology*, 17(7),
 1050 1951–1972.
- 1051 Kumar, S. V., Jasinski, M., Mocko, D., Rodell, M., Borak, J., LI, B., ... Peters-Lidard, C. D. (2018).
 1052 NCA-LDAS land analysis: Development and performance of a multisensor,
 1053 multivariate land data assimilation system for the National Climate Assessment.
 1054 *Journal of Hydrometeorology*.
- 1055 Landerer, F. W., & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage
 1056 estimates. *Water resources research*, 48(4).
- 1057 Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015). Inroads
 1058 of remote sensing into hydrologic science during the WRR era. *Water Resources*
 1059 *Research*, 51(9), 7309-7342.
- 1060 Li, B., Rodell, M., Zaitchik, B. F., Reichle, R. H., Koster, R. D., & van Dam, T. M. (2012).
 1061 Assimilation of GRACE terrestrial water storage into a land surface model:
 1062 Evaluation and potential value for drought monitoring in western and central
 1063 Europe. *Journal of Hydrology*, 446, 103–115.
- 1064 Lievens, H., Tomer, S. K., Al Bitar, A., De Lannoy, G. J. M., Drusch, M., Dumedah, G., ... Pan, M.
 1065 (2015). SMOS soil moisture assimilation for improved hydrologic simulation in the
 1066 Murray Darling Basin, Australia. *Remote Sensing of Environment*, 168, 146–162.
- 1067 Liu, Q., Reichle, R. H., Bindlish, R., Cosh, M. H., Crow, W. T., De Jeu, R., ... Jackson, T. J. (2011).
 1068 The contributions of precipitation and soil moisture observations to the skill of soil
 1069 moisture estimates in a land data assimilation system. *Journal of Hydrometeorology*,
 1070 12(5), 750–765.

- 1071 Luthcke, S.B., T.J. Sabaka, B.D. Loomis, et al. (2013), Antarctica, Greenland and Gulf of Alaska
1072 land ice evolution from an iterated GRACE global mascon solution, *J. Glac.*; 59(216),
1073 613-631, doi:10.3189/2013JoG12J147
- 1074 McCabe, M. F., Rodell, M., Alsdorf, D. E., Gonzalez Miralles, D., Uijlenhoet, R., Wagner, W., ... &
1075 Shi, J. (2017). The future of Earth observation in hydrology. *Hydrology and Earth*
1076 *System Sciences*, 21(7), 3879-3914.
- 1077 Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., & Döll,
1078 P. (2014). Sensitivity of simulated global-scale freshwater fluxes and storages to
1079 input data, hydrological model structure, human water use and
1080 calibration. *Hydrology and Earth System Sciences*, 18(9), 3511-3538.
- 1081 Pan, M., Sahoo, A. K., Troy, T. J., Vinukollu, R. K., Sheffield, J., & Wood, E. F. (2012).
1082 Multisource estimation of long-term terrestrial water budget for major global river
1083 basins. *Journal of Climate*, 25(9), 3191–3206.
- 1084 Ramillien, G. L., Seoane, L., Frappart, F., Biancale, R., Gratton, S., Vasseur, X., & Bourgogne, S.
1085 (2012). Constrained regional recovery of continental water mass time-variations
1086 from GRACE-based geopotential anomalies over South America. *Surveys in*
1087 *geophysics*, 33(5), 887-905.
- 1088 Reager, J. T., Thomas, A. C., Sproles, E. A., Rodell, M., Beaudoin, H. K., Li, B., & Famiglietti, J. S.
1089 (2015). Assimilation of GRACE terrestrial water storage observations into a land
1090 surface model for the assessment of regional flood potential. *Remote Sensing*, 7(11),
1091 14663–14679.
- 1092 Reichle, R. H., & Koster, R. D. (2003). Assessing the impact of horizontal error correlations in
1093 background fields on soil moisture estimation. *Journal of Hydrometeorology*, 4(6),
1094 1229–1242.

- 1095 Reichle, R. H., & Koster, R. D. (2005). Global assimilation of satellite surface soil moisture
1096 retrievals into the NASA Catchment land surface model. *Geophysical Research*
1097 *Letters*, 32(2).
- 1098 Reichle, R. H., Koster, R. D., Liu, P., Mahanama, S. P., Njoku, E. G., & Owe, M. (2007).
1099 Comparison and assimilation of global soil moisture retrievals from the Advanced
1100 Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the
1101 Scanning Multichannel Microwave Radiometer (SMMR). *Journal of Geophysical*
1102 *Research: Atmospheres*, 112(D9).
- 1103 Reichle, R. H., De Lannoy, G. J., Liu, Q., Ardizzone, J. V., Colliander, A., Conaty, A., ... Kimball, J.
1104 S. (2017a). Assessment of the SMAP Level-4 Surface and Root-Zone Soil Moisture
1105 Product Using In Situ Measurements. *Journal of Hydrometeorology*, 18(10), 2621–
1106 2645.
- 1107 Reichle, R. H., De Lannoy, G. J., Liu, Q., Koster, R. D., Kimball, J. S., Crow, W. T., ... Conaty, A. L.
1108 (2017b). Global Assessment of the SMAP Level-4 Surface and Root-Zone Soil
1109 Moisture Product Using Assimilation Diagnostics. *Journal of Hydrometeorology*,
1110 18(12), 3217–3237.
- 1111 Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P., & Partyka, G. S. (2017c).
1112 Land surface precipitation in MERRA-2. *Journal of Climate*, 30(5), 1643–1664.
- 1113 Renzullo, L. J., Barrett, D. J., Marks, A. S., Hill, M. J., Guerschman, J. P., Mu, Q., & Running, S. W.
1114 (2008). Multi-sensor model-data fusion for estimation of hydrologic and energy flux
1115 parameters. *Remote Sensing of Environment*, 112(4), 1306-1319.
- 1116 Renzullo, L. J., Van Dijk, A., Perraud, J.-M., Collins, D., Henderson, B., Jin, H., ... McJannet, D. L.
1117 (2014). Continental satellite soil moisture data assimilation improves root-zone
1118 moisture analysis for water resources assessment. *Journal of Hydrology*, 519, 2747–
1119 2762.

- 1120 Rodell, M., H.K. Beaudoin, T. L'Ecuyer, W. Olson, J.S. Famiglietti, P.R. Houser, R. Adler, M.
 1121 Bosilovich, C.A. Clayson, D. Chambers, E. Clark, E. Fetzner, X. Gao, G. Gu, K. Hilburn, G.
 1122 Huffman, D.P. Lettenmaier, W.T. Liu, F.R. Robertson, C.A. Schlosser, J. Sheffield, and
 1123 E.F. Wood, The observed state of the water cycle in the early 21st century, *J. Climate*,
 1124 28 (21), 8289-8318, doi:10.1175/JCLI-D-14-00555.1, 2015.
- 1125 Sakumura, C., Bettadpur, S., Save, H., & McCullough, C. (2016). High-frequency terrestrial
 1126 water storage signal capture via a regularized sliding window mascon product from
 1127 GRACE. *Journal of Geophysical Research: Solid Earth*, 121(5), 4014-4030.
- 1128 Save, H., Bettadpur, S., & Tapley, B. D. (2016). High-resolution CSR GRACE RL05
 1129 mascons. *Journal of Geophysical Research: Solid Earth*, 121(10), 7547-7569.
- 1130 Scanlon, B. R., Duncan, I., & Reedy, R. C. (2013). Drought and the water-energy nexus in
 1131 Texas. *Environmental Research Letters*, 8(4), 045033.
- 1132 Schaefer, G. L., Cosh, M. H., & Jackson, T. J. (2007). The USDA natural resources conservation
 1133 service soil climate analysis network (SCAN). *Journal of Atmospheric and Oceanic*
 1134 *Technology*, 24(12), 2073-2077.
- 1135 Schumacher, M., Forootan, E., van Dijk, A., Schmied, H. M., Crosbie, R. S., Kusche, J., & Döll, P.
 1136 (2018). Improving drought simulations within the Murray-Darling Basin by
 1137 combined calibration/assimilation of GRACE data into the WaterGAP Global
 1138 Hydrology Model. *Remote Sensing of Environment*, 204, 212-228.
- 1139 Su, H., Yang, Z.-L., Dickinson, R. E., Wilson, C. R., & Niu, G.-Y. (2010). Multisensor snow data
 1140 assimilation at the continental scale: The value of Gravity Recovery and Climate
 1141 Experiment terrestrial water storage information. *Journal of Geophysical Research:*
 1142 *Atmospheres*, 115(D10).
- 1143 Swenson, S., & Wahr, J. (2006). Post-processing removal of correlated errors in GRACE
 1144 data. *Geophysical Research Letters*, 33(8).

- 1145 Tangdamrongsub, N., Steele-Dunne, S. C., Gunter, B. C., Ditmar, P. G., & Weerts, A. H. (2015).
 1146 Data assimilation of GRACE terrestrial water storage estimates into a regional
 1147 hydrological model of the Rhine River basin. *Hydrology and Earth System Sciences*,
 1148 19(4), 2079.
- 1149 Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The gravity recovery and
 1150 climate experiment: Mission overview and early results. *Geophysical Research*
 1151 *Letters*, 31(9).
- 1152 Tian, S., Tregoning, P., Renzullo, L. J., van Dijk, A. I., Walker, J. P., Pauwels, V., & Allgeyer, S.
 1153 (2017). Improved water balance component estimates through joint assimilation of
 1154 GRACE water storage and SMOS soil moisture retrievals. *Water Resources Research*,
 1155 53(3), 1820–1840.
- 1156 van Dijk, A. I., Renzullo, L. J., Wada, Y., & Tregoning, P. (2014). A global water cycle
 1157 reanalysis (2003–2012) merging satellite gravimetry and altimetry observations
 1158 with a hydrological multi-model ensemble. *Hydrology and Earth System Sciences*,
 1159 18(8), 2955.
- 1160 Walker, J. P., Willgoose, G. R., & Kalma, J. D. (2001). One-dimensional soil moisture profile
 1161 retrieval by assimilation of near-surface observations: a comparison of retrieval
 1162 algorithms. *Advances in Water Resources*, 24(6), 631–650.
- 1163 Wahr, J., Swenson, S., & Velicogna, I. (2006). Accuracy of GRACE mass estimates. *Geophysical*
 1164 *Research Letters*, 33(6).
- 1165 Wiese, D. N., Landerer, F. W., & Watkins, M. M. (2016). Quantifying and reducing leakage
 1166 errors in the JPL RL05M GRACE mascon solution. *Water Resources Research*, 52(9),
 1167 7490–7502.

- 1168 Zaitchik, B. F., & Rodell, M. (2009). Forward-looking assimilation of MODIS-derived snow-
1169 covered area into a land surface model. *Journal of Hydrometeorology*, 10(1), 130-
1170 148.
- 1171 Zaitchik, B. F., Rodell, M., & Reichle, R. H. (2008). Assimilation of GRACE terrestrial water
1172 storage data into a land surface model: Results for the Mississippi River basin.
1173 *Journal of Hydrometeorology*, 9(3), 535-548.
- 1174 Zhao, L., & Yang, Z. L. (2018). Multi-sensor land data assimilation: Toward a robust global
1175 soil moisture and snow estimation. *Remote sensing of environment*, 216, 13-27.
1176
1177
1178

1179 **Tables and Figures**

1180

1181

List of Figure Captions

Figure 1. Simplified flowchart of the multi-sensor assimilation of GRACE TWS and SMOS Tb observations (GRACE+SMOS DA). Run 1: Conduct one month forecast ensemble integration with SMOS Tb assimilation (SMOS DA Run 1, De Lannoy et al., 2016a). GRACE DA: Calculate model TWS observation prediction through spatial aggregation (model to observation resolution) and temporal aggregation (daily to monthly). Calculate the increments via 3D EnKF analysis (Section 4.1). Rewind the model to the beginning of the month and apply the GRACE increments (Giroto et al., 2016). Note that during Run 1, SMOS DA is performed only to make TWS observation predictions. Run 2: Integrate the model from the 1st - to the last day and perform SMOS DA assimilation (SMOS DA Run 2, De Lannoy et al., 2016a). Repeat for the following month.

Figure 2. Column one) skill (R), and (columns two-to-four) difference in skill (ΔR) between the data assimilation (DA) and open-loop (i.e., no assimilation) estimates for surface soil moisture (sfmc), rootzone soil moisture (rzmc), groundwater (GW), runoff, and terrestrial water storage (TWS). Skill is measured as the correlation coefficient (R) versus in-situ and GRACE (for TWS) measurements. Blue colors indicate skill improvement, that is, DA is more skillful than OL, and red colors indicate skill degradation. Note that the TWS skill metrics for GRACE DA and GRACE+SMOS DA in panels (j) and (t) do not provide an independent validation (see text for details).

Figure 3. Open-loop (OL) and assimilation experiments (DA) skill differences (i.e., skill of DA minus skill of the OL). The assimilation experiments are listed in Table 4. Skill is measured in terms of correlation coefficient (R) and unbiased root-mean-squared difference (ubRMSD) when compared to independent in-situ measurements of groundwater (GW),

rootzone soil moisture (rzmc), surface soil moisture (sfmc), and runoff. Metrics for terrestrial water storage (TWS) are computed against the satellite-based GRACE TWS observations. The mean values across the sites and the 95% confidence intervals are obtained after clustering of the sites. Soil moisture metrics for the sparse network sites are computed from the available sites in the Soil Climate Analysis Network (SCAN) and the U.S. Climate Reference Network (USCRN).

Figure 4. 1 July 2010 – 1 July 2016 average of the ensemble standard deviation of (column 1) open-loop (i.e., OL, no assimilation), and (columns 2-4) reduction in ensemble standard deviation ($\text{stdv}_{\text{DA}} - \text{stdv}_{\text{OL}}$) due to column 2) single-sensor GRACE data assimilation (GRACE DA); column 3) single-sensor SMOS DA; column 4) multi-sensor GRACE+SMOS DA experiments for surface soil moisture (sfmc), rootzone soil moisture (rzmc), and catchment deficit (catdef). The domain average value is reported in each panel.

Figure 5. Actual errors vs. estimate uncertainties. The actual errors in the (ensemble mean) estimates are represented in terms of in-situ ubRMSD. The (ensemble-based) estimate uncertainties are shown in terms of ensemble spread average over the times when in-situ measurements are available. Each scatter plots contains results for the open-loop case (grey stars), the single-sensor GRACE DA (black crosses), the single-sensor SMOS DA (blue circles), and the multi-sensor GRACE+SMOS DA (red triangles). The ubRMSD are calculated for a) sparse network surface soil moisture; b) sparse network rootzone soil moisture; c) catdef vs. groundwater observations. The distributions of the scatter points are also shown on the y-axis.

Figure 6. Example time series of increments calculated for catchment deficit (catdef), rootzone excess (rzexc), surface excess (srfexc) (Section 3.2) from the a) single-sensor GRACE DA; b) single-sensor SMOS DA; c) multi-sensor GRACE+SMOS DA. The increments brought by SMOS Tb observations (in SMOS DA and GRACE+SMOS DA) are shown as green and blue dots for the srfexc and rzexc, respectively. The dashed lines show the 14-days running mean of srfexc and rzexc. The specific location is shown in Figure 7 and corresponds to a 36 km grid in Iowa (lat=41.81, lon=-91.31). Note that there are different y-axes for the catdef and soil moisture excess increments because an increase in catdef is a decrease in water storage, and because the catdef increments tend to be much larger.

Figure 7. Temporal correlation between increments of catchment deficit (catdef) and (a) increments of surface excess (srfexc) and (b) rootzone excess (rzexc) in the multi-sensor GRACE+SMOS DA. Note that positive correlation indicates that the addition of water in the “catdef” (deficit) reservoir implies a subtraction of water from the “rzexc” or “srfexc” (excess) reservoirs. In the text, this positive correlation is therefore referred to as “anti-correlation. The yellow star indicates the location of the time series shown in Figure 6. Grey areas have non-significant correlation values.

Figure 8. Average absolute increments (July 2010-July 2016) obtained from a,c) the single sensor SMOS DA; e) the single sensor GRACE DA and b,d,f) the multi-sensor GRACE+SMOS DA. The shown increments are in terms of the CLSM variables a,b) surface excess (srfexc); c,d) rootzone excess (rzexc); and e,f) catchment deficit (catdef).

Table 1. Summary of relevant literature on the assimilation of GRACE TWS data in multi-sensor data assimilation systems.

Study	Region of Focus	Period of Study	Hydrologic Focus	Model Used	Data Assimilation Approach	Assimilated Observations	Key Findings
Su et al., 2010	North America	Jan. 2002 - Jun. 2007	Snow	Community Land Model (CLM, Bonan et al., 2002)	1D Ensemble Kalman Smoother and Filter	GRACE TWS Spherical Harmonics MODIS Snow Cover Fraction MOD10C1	- The multisensor approach can provide significant improvements over a MODIS-only approach
Tian et al., 2017	Australia	Jan. 2010 - Dec. 2013	Surface and Subsurface Hydrology	World-Wide Water (W3, van Dijk et al., 2013)	1D Ensemble Kalman Smoother for the joint DA and Filter for the single sensor DA	GRACE Mascons JPL SMOS L3 Soil Moisture Retrievals	- The joint assimilation can improve water balance component estimates, especially in SM profile and groundwater - The joint assimilation performs better than assimilation of GRACE or SMOS - The SMOS assimilation degrades groundwater estimates
Zhao and Yang 2018	Global	Jan. 2003 - Dec. 2009	Soil Moisture and Snow	Community Land Model Version 4 (CLM4, Oleson et al., 2010)	1D Deterministic Ensemble Adjustment Kalman Filter	GRACE daily solutions (Sakumura et al., 2016) AMSR-E Brightness Temperature (Tb) MODIS Snow Cover Fraction MOD10C2 GLASS LAI	- The assimilation of MODIS snow cover fraction improves snow estimation in mid- and high-latitudes - The lower and higher frequencies of AMSR-E brightness temperatures improve global soil moisture and snow - GRACE TWS degrades soil moisture but improves snow in high-latitude regions
Khaki and Awange 2019	South America	Jan. 2002 - Jan. 2013	Surface and Subsurface Hydrology	World-Wide Water Resources Assessment (W3RA, van Dijk, 2010)	3D Ensemble Square Root Filter	GRACE TWS Spherical Harmonics SMOS and AMSR-E L3 soil moisture	- Data assimilation of soil moisture and TWS observations leads to more accurate estimates of groundwater and soil moisture variations
This work	Continental US	Jun. 2010 - Jun. 2016	Surface and Subsurface Hydrology	Catchment Land Surface Model (CLSM, Koster et al., 2000)	3D Ensemble Kalman Smoother and Filter	GRACE TWS Spherical Harmonics SMOS L1 brightness temperatures	- The assimilation of GRACE TWS improves groundwater estimates, while SMOS DA improves surface soil moisture estimates - The multi-sensor assimilation of GRACE and SMOS maintains the single-sensor assimilation benefits - Water storage increments introduced by GRACE and SMOS DA are anti-correlated

Table 2. Model error perturbation parameters. The perturbation types are either additive (A) or multiplicative (M), with the standard deviation given for a normal (if additive) or lognormal (if multiplicative) distribution. Temporal correlations are implemented using a first-order autoregressive model. Spatial correlations are considered isotropic. Perturbations to prognostic variables are uncorrelated with each other while cross-correlations are considered across forcing perturbations: precipitation errors are correlated at -0.8 and 0.5 with shortwave and longwave incoming radiation errors, respectively, while shortwave and longwave incoming radiation errors are correlated at -0.5.

	Type	Std. dev.		Spatial Correlation [°]	Temporal Correlation [hours]
Precipitation	M	0.5	-	2	72
Shortwave Radiation	M	0.3	-	2	72
Longwave Radiation	A	20	W m ⁻²	2	72
catdef	A	0.15	kg m ⁻² hr ⁻¹	2	24
srfexc	A	0.16	kg m ⁻² hr ⁻¹	0.5	3
swe	M	0.0012	-	2	24

Table 3. Parameters of the GRACE terrestrial water storage (TWS) and SMOS brightness temperature (Tb) analyses. The multi-sensor assimilation (GRACE+SMOS DA) uses the same observation types and parameters as the single-sensor assimilation case.

Analysis type	TWS (GRACE DA)	Tb (SMOS DA)
State Vector ($x_{i,t}^j$)	[<i>catdef</i> , <i>swe</i>]'	[<i>srfexc</i> , <i>rzexc</i> , <i>tc1</i> , <i>tc2</i> , <i>tc4</i> , <i>ght1</i> ']
Observations ($Z_{i,T}^j$)	level 3 TWS anomalies based on RL05 spherical harmonics	fitted (40° incidence angle) Tb
Spatial Resolution	3°	EASEv2 36 km
Temporal Resolution (T)	monthly	every 3 hours
Obs. Spatial Correlation	2°	0.25°
Localization Radius (r_l)	6°	1°
Observation Error Std. Dev.	15 mm	5 K

Table 4. Overview of model and assimilation experiments.

	Observation Type	Main Reference
1. Ensemble Open-Loop	none	Section 5.1.1
2. GRACE DA	TWS	Section 4.2 and Giroto et al., (2016)
3. SMOS DA	Tb	Section 4.3 and De Lannoy and Reichle, (2016a)
4. GRACE+SMOS DA	TWS+Tb	Section 4.4 (This work)

Table 5. Mean of the correlation coefficient (R) and the unbiased root-mean-square difference ($ubRMSD$) across all validation locations for estimates from the open-loop (OL) and the GRACE data assimilation (GRACE DA), SMOS data assimilation (SMOS DA), and multi-sensor GRACE and SMOS assimilation (GRACE+SMOS DA). Mean statistics and 95% confidence intervals (CI) are obtained from the clustering of the sites (Section 3.4). The bold text indicates statistically significant improvements relative to the OL case.

			R		$ubRMSD$
N. Sites			Mean ($\pm$ CI)	Mean ($\pm$ CI)	
Surface Soil Moisture	Cal/Val	4	OL	0.82 (-)	0.027 (-) m ³ m ⁻³
			GRACE DA	0.84 (-)	0.024 (-) m ³ m ⁻³
			SMOS DA	0.85 (-)	0.023 (-) m ³ m ⁻³
			GRACE+SMOS DA	0.86 (-)	0.023 (-) m ³ m ⁻³
	USCRN and SCAN	93	OL	0.73 ($\pm$ 0.03)	0.038 ($\pm$ 0.002) m ³ m ⁻³
			GRACE DA	0.73 ($\pm$ 0.03)	0.038 ($\pm$ 0.002) m ³ m ⁻³
			SMOS DA	0.78 ($\pm$0.02)	0.034 ($\pm$0.001) m ³ m ⁻³
			GRACE+SMOS DA	0.77 ($\pm$0.02)	0.035 ($\pm$0.001) m ³ m ⁻³
Rootzone Soil Moisture	Cal/Val	2	OL	0.77 (-)	0.027 (-) m ³ m ⁻³
			GRACE DA	0.74 (-)	0.028 (-) m ³ m ⁻³
			SMOS DA	0.81 (-)	0.025 (-) m ³ m ⁻³
			GRACE+SMOS DA	0.76 (-)	0.027 (-) m ³ m ⁻³
	USCRN and SCAN	84	OL	0.68 ($\pm$ 0.03)	0.035 ($\pm$ 0.002) m ³ m ⁻³
			GRACE DA	0.65 ($\pm$ 0.03)	0.035 ($\pm$ 0.002) m ³ m ⁻³
			SMOS DA	0.71 ($\pm$ 0.03)	0.034 ($\pm$ 0.002) m ³ m ⁻³
			GRACE+SMOS DA	0.69 ($\pm$ 0.03)	0.034 ($\pm$ 0.002) m ³ m ⁻³
Groundwater		157	OL	0.63 ($\pm$ 0.01)	57.8 ($\pm$ 1.00) mm
			GRACE DA	0.65 ($\pm$0.01)	55.5 ($\pm$0.96) mm
			SMOS DA	0.63 ($\pm$ 0.01)	57.6 ($\pm$ 1.00) mm
			GRACE+SMOS DA	0.65 ($\pm$0.01)	55.1 ($\pm$0.95) mm
Runoff		238	OL	0.62 ($\pm$ 0.02)	0.82 ($\pm$ 0.02) mm/d
			GRACE DA	0.61 ($\pm$ 0.02)	0.76 ($\pm$0.02) mm/d
			SMOS DA	0.63 ($\pm$ 0.02)	0.82 ($\pm$ 0.02) mm/d
			GRACE+SMOS DA	0.62 ($\pm$ 0.02)	0.76 ($\pm$0.02) mm/d
TWS	CONUS		OL	0.72 ($<\pm$ 0.01)	46.8 ($<\pm$ 0.01) mm
			GRACE DA*	0.96 ($<\pm$ 0.01)	24.6 ($<\pm$ 0.01) mm
			SMOS DA	0.71 ($<\pm$ 0.01)	45.0 ($<\pm$ 0.01) mm
			GRACE+SMOS DA*	0.95 ($<\pm$ 0.01)	25.8 ($<\pm$ 0.01) mm

*** This evaluation does not represent an independent validation**

Figure 1. Simplified flowchart of the multi-sensor assimilation of GRACE TWS and SMOS Tb observations (GRACE+SMOS DA). Run 1: Conduct one month forecast ensemble integration with SMOS Tb assimilation (SMOS DA Run 1, De Lannoy et al., 2016a). GRACE DA: Calculate model TWS observation prediction through spatial aggregation (model to observation resolution) and temporal aggregation (daily to monthly). Calculate the increments via 3D EnKF analysis (Section 4.1). Rewind the model to the beginning of the month and apply the GRACE increments (Giroto et al., 2016). Note that during Run 1, SMOS DA is performed only to make TWS observation predictions. Run 2: Integrate the model from the 1st - to the last day and perform SMOS DA assimilation (SMOS DA Run 2, De Lannoy et al., 2016a). Repeat for the following month.

Figure 2. Column one) skill (R), and (columns two-to-four) difference in skill (ΔR) between the data assimilation (DA) and open-loop (i.e., no assimilation) estimates for surface soil moisture (sfmc), rootzone soil moisture (rzmc), groundwater (GW), runoff, and terrestrial water storage (TWS). Skill is measured as the correlation coefficient (R) versus in-situ and GRACE (for TWS) measurements. Blue colors indicate skill improvement, that is, DA is more skillful than OL, and red colors indicate skill degradation. Note that the TWS skill metrics for GRACE DA and GRACE+SMOS DA in panels (j) and (t) do not provide an independent validation (see text for details).

1310

Figure 3. Open-loop (OL) and assimilation experiments (DA) skill differences (i.e., skill of DA minus skill of the OL). The assimilation experiments are listed in Table 4. Skill is measured in terms of correlation coefficient (R) and unbiased root-mean-squared difference (ubRMSD) when compared to independent in-situ measurements of groundwater (GW), rootzone soil moisture (rzmc), surface soil moisture (sfmc), and runoff. Metrics for terrestrial water storage (TWS) are computed against the satellite-based GRACE TWS observations. The mean values across the sites and the 95% confidence intervals are obtained after clustering of the sites. Soil moisture metrics for the sparse network sites are computed from the available sites in the Soil Climate Analysis Network (SCAN) and the U.S. Climate Reference Network (USCRN).

1324

Figure 4. 1 July 2010 – 1 July 2016 average of the ensemble standard deviation of (column 1) open-loop (i.e., OL, no assimilation), and (columns 2-4) reduction in ensemble standard deviation ($\text{stdv}_{\text{DA}} - \text{stdv}_{\text{OL}}$) due to column 2) single-sensor GRACE data assimilation (GRACE DA); column 3) single-sensor SMOS DA; column 4) multi-sensor GRACE+SMOS DA experiments for surface soil moisture (sfmc), rootzone soil moisture (rzmc), and catchment deficit (catdef). The domain average value is reported in each panel.

Figure 5. Actual errors vs. estimate uncertainties. The actual errors in the (ensemble mean) estimates are represented in terms of in-situ ubRMSD. The (ensemble-based) estimate uncertainties are shown in terms of ensemble spread average over the times when in-situ measurements are available. Each scatter plots contains results for the open-loop case (grey stars), the single-sensor GRACE DA (black crosses), the single-sensor SMOS DA (blue circles), and the multi-sensor GRACE+SMOS DA (red triangles). The ubRMSD are calculated for a) sparse network surface soil moisture; b) sparse network rootzone soil moisture; c) catdef vs. groundwater observations. The distributions of the scatter points are also shown on the y-axis.

1344

Figure 6. Example time series of increments calculated for catchment deficit (catdef), rootzone excess (rzexc), surface excess (srfexc) (Section 3.2) from the a) single-sensor GRACE DA; b) single-sensor SMOS DA; c) multi-sensor GRACE+SMOS DA. The increments brought by SMOS Tb observations (in SMOS DA and GRACE+SMOS DA) are shown as green and blue dots for the srfexc and rzexc, respectively. The dashed lines show the 14-days running mean of srfexc and rzexc. The specific location is shown in Figure 7 and corresponds to a 36 km grid in Iowa (lat=41.81, lon=-91.31). Note that there are different y-axes for the catdef and soil moisture excess increments because an increase in catdef is a decrease in water storage, and because the catdef increments tend to be much larger.

1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355

(a) Correlation between Increments of *srfexc* and *catdef*

(b) Correlation between Increments of *rzexc* and *catdef*

Figure 7. Temporal correlation between increments of catchment deficit (*catdef*) and (a) increments of surface excess (*srfexc*) and (b) rootzone excess (*rzexc*) in the multi-sensor GRACE+SMOS DA. Note that positive correlation indicates that the addition of water in the “*catdef*” (deficit) reservoir implies a subtraction of water from the “*rzexc*” or “*srfexc*” (excess) reservoirs. In the text, this positive correlation is therefore referred to as “anti-correlation. The yellow star indicates the location of the time series shown in Figure 6. Grey areas have non-significant correlation values.

Figure 8. Average absolute increments (July 2010–July 2016) obtained from a,c) the single sensor SMOS DA; e) the single sensor GRACE DA and b,d,f) the multi-sensor GRACE+SMOS DA. The shown increments are in terms of the CLSM variables a,b) surface excess (srfexc); c,d) rootzone excess (rzexc); and e,f) catchment deficit (catdef).