The Impact of Rainfall Error Characterization on the Estimation of Soil Moisture Fields in a Land Data Assimilation System

VIVIANA MAGGIONI

Department of Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut

ROLF H. REICHLE

Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

EMMANOUIL N. ANAGNOSTOU

Department of Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut

(Manuscript received 9 September 2011, in final form 2 February 2012)

ABSTRACT

This study presents a numerical experiment to assess the impact of satellite rainfall error structure on the efficiency of assimilating near-surface soil moisture observations. Specifically, the study contrasts a multidimensional satellite rainfall error model (SREM2D) to a simpler rainfall error model (CTRL) currently used to generate rainfall ensembles as part of the ensemble-based land data assimilation system developed at the NASA Global Modeling and Assimilation Office. The study is conducted in the Oklahoma region using rainfall data from a NOAA multisatellite global rainfall product [the Climate Prediction Center (CPC) morphing technique (CMORPH)] and the National Weather Service rain gauge–calibrated radar rainfall product [Weather Surveillance Radar-1988 Doppler (WSR-88D)] representing the "uncertain" and "reference" model rainfall forcing, respectively. Soil moisture simulations using the Catchment land surface model (CLSM), obtained by forcing the model with reference rainfall, are randomly perturbed to represent satellite retrieval uncertainty, and assimilated into CLSM as synthetic near-surface soil moisture observations. The assimilation estimates show improved performance metrics, exhibiting higher anomaly correlation coefficients (e.g., ~0.79 and ~0.90 in the SREM2D nonassimilation and assimilation experiments for root zone soil moisture, respectively) and lower root-mean-square errors (e.g., ~0.034 $m^3 m^{-3}$ and ~0.024 $m^3 m^{-3}$ in the SREM2D nonassimilation and assimilation experiments for root zone soil moisture, respectively). The more elaborate rainfall error model in the assimilation system leads to slightly improved assimilation estimates. In particular, the relative enhancement due to SREM2D over CTRL is larger for root zone soil moisture and in wetter rainfall conditions.

1. Introduction

Soil moisture is a key variable of the water and energy cycles, playing a role in many research fields, such as hydrology, agriculture, and ecology. It also controls decomposition in terms of the biogeochemical cycling (e.g., carbon and nitrogen cycles). Being a storage component for precipitation and radiation anomalies, soil water content might control cloud coverage, precipitation, and hydrological parameters, such as runoff and evapotranspiration (Betts and Ball 1998). Moreover, soil moisture is involved in several feedbacks at local, regional, and global scales. In particular, soil moisture–temperature and soil moisture–precipitation feedbacks might have a significant impact on climate change projections (Seneviratne et al. 2010). Therefore, a realistic characterization of soil moisture and its uncertainty is important to improve weather, climate, and hydrologic predictions.

Soil moisture can be estimated through different approaches: (i) direct ground measurements (e.g., Walker et al. 2004), (ii) retrievals from low-frequency active and passive microwave data (e.g., Schmugge et al. 2002), (iii) integration of a land surface model forced with meteorological data derived from observations (e.g., Peters-Lidard et al. 2007), and (iv) through land data

Corresponding author address: Viviana Maggioni, Civil and Environmental Engineering, University of Connecticut, 261 Glenbrook Road, Unit 2037, Storrs, CT 06269.
E-mail: viviana@engr.uconn.edu

DOI: 10.1175/JHM-D-11-0115.1

© 2012 American Meteorological Society

assimilation, combining the complementary information from measurements and models of the land surface into a superior estimate of soil moisture (e.g., Reichle and Koster 2005; Li et al. 2010).

In situ observations continue to be scarce as only few ground-based networks are available worldwide (Robock et al. 2000; Robinson et al. 2008; Dorigo et al. 2011). Satellite remote sensing observations represent an alternative. However, satellite retrievals of soil moisture are affected by significant errors due to sensor limitations (e.g., sampling, resolution, and land cover heterogeneity effects), uncertainty in the parameterization of the relationship between brightness temperature and soil moisture, and difficulty in obtaining a global distribution of parameters for the retrieval algorithm. Furthermore, satellite retrievals only represent near-surface soil moisture fields. Therefore, a common approach to estimate continuous and spatially distributed soil moisture fields is to force a land surface model with meteorological observations. The main uncertainties affecting model predictions of soil moisture include errors in the meteorological forcing variables, erroneous estimates of the model parameters, and deficient model formulations.

Data assimilation systems can provide a superior product by merging the satellite retrieval information with the spatially and temporally complete information given by a land surface model (Parajka et al. 2006; Reichle et al. 2008; Drusch et al. 2009; among others). Reichle et al. (2007) demonstrated that by assimilating satellite retrievals of near-surface soil moisture retrievals the resulting estimates of surface and root zone soil moisture are superior to either satellite data or model data alone. This is achieved by correcting the model-generated values of soil moisture toward the observational estimates depending on the level of error associated with each product. However, the improvement from retrieved soil moisture assimilation strongly depends on the quality of meteorological forcing observations. A recent study by Liu et al. (2011) showed that (i) assimilating surface soil moisture retrievals and (ii) improving the precipitation forcing through gauge-based corrections contribute similar and largely independent amounts of information to the skill of surface and root zone soil moisture estimates.

A key issue in land data assimilation for soil moisture is that the model and the observational uncertainties are poorly known, while data assimilation using a poor characterization of uncertainty is likely to produce poor estimates of land surface variables (Crow and Van Loon 2006; Reichle et al. 2008). Therefore, the quality of the assimilation estimates depends critically on the realism of the error estimates for the model and the observations. Arguably, the way in which model errors are handled in standard land data assimilation systems is still very simplistic. This is particularly true for rainfall errors, whose multidimensional character at fine space and time scales is typically ignored in current land assimilation systems (see below for details). Thus, improved error modeling strategies may be needed to characterize the uncertainty in the simulation of soil moisture fields from a land surface model in order to enhance the efficiency of the data assimilation system.

The land data assimilation system GMAO-LDAS developed at the National Aeronautics and Space Administration Global Modeling and Assimilation Office (NASA GMAO) is based on the ensemble Kalman filter (EnKF) approach that dynamically updates model error covariance information by producing an ensemble of model predictions, which are individual model realizations perturbed by an assumed model error. The current rainfall error modeling approach in GMAO-LDAS simply scales the input precipitation forcing with a multiplicative perturbation (Reichle et al. 2007). This implies, for example, that all ensemble members have zero precipitation whenever the input precipitation is zero. The approach is numerically convenient but does not fully describe the error characteristics of remote sensing rainfall retrievals, particularly with respect to rain detection and false alarms.

Hossain and Anagnostou (2006a) have developed a more sophisticated satellite rainfall error model (SREM2D) for generating ensembles of satellite rain fields on the basis of high-accuracy "reference" rain fields. SREM2D is capable of conserving the satellite retrieval error structure across scales, unlike simpler error modeling approaches that revealed significant scale-dependent biases (Hossain and Anagnostou 2006b). Furthermore, Hossain and Anagnostou (2005) and Maggioni et al. (2011) demonstrated that soil moisture ensembles from land surface models driven with SREM2D-generated rainfall forcing better capture soil moisture error characteristics.

These considerations merit further investigation as to how rainfall error modeling can affect the efficiency of a soil moisture data assimilation system, and, in particular, how a more elaborate rainfall error model would impact the performance of the GMAO-LDAS in terms of soil moisture predictions. The present paper is a follow-up study to Maggioni et al. (2011) with the specific objective of comparing SREM2D to a simpler multiplicative rainfall error model in terms of the efficiency in assimilating soil moisture data using the GMAO-LDAS. The present paper uses the study area, study period, and experiment setup of Maggioni et al. (2011) and adds the assimilation of synthetic observations of soil moisture.

The study area, period, and data employed are briefly described in section 2. In section 3 we describe the GMAO-LDAS and the rainfall error modeling

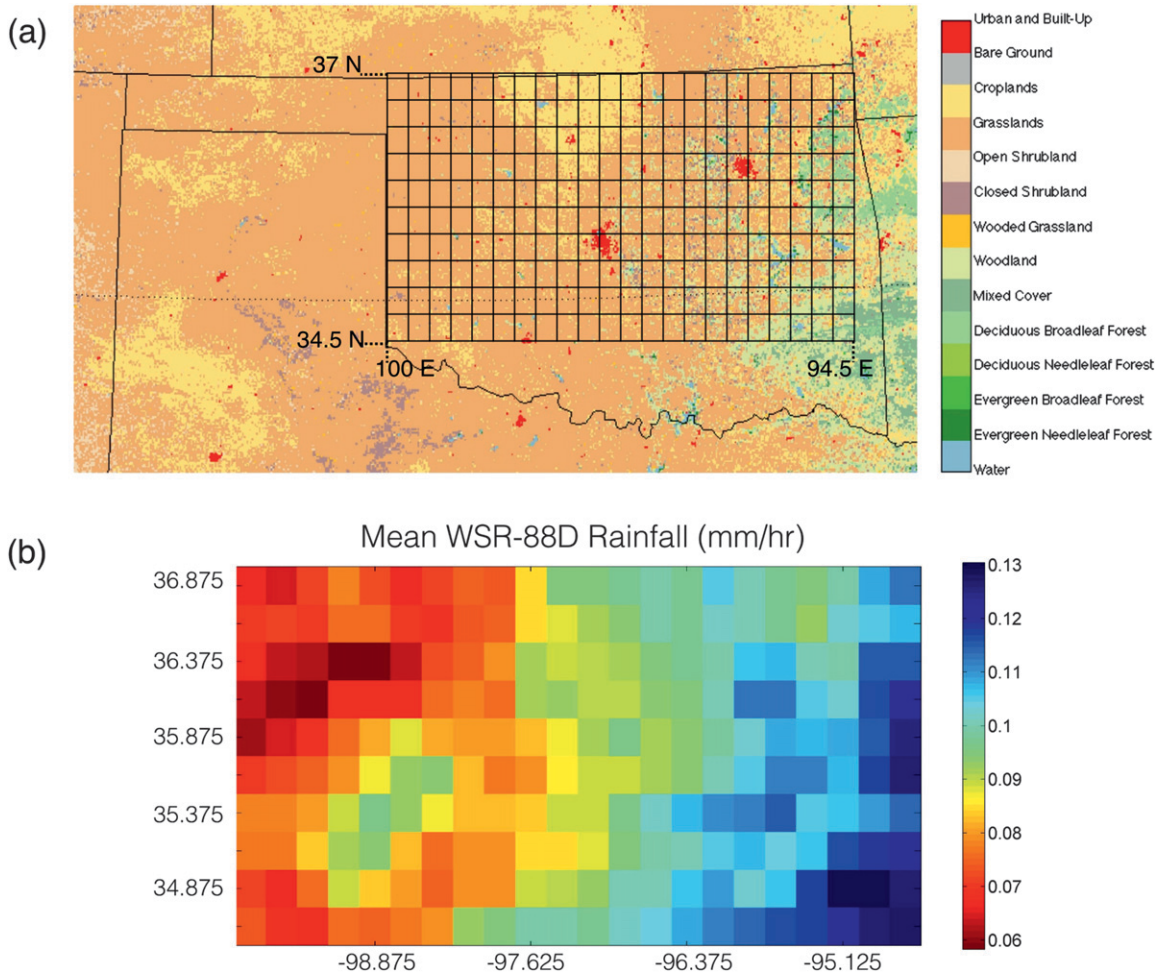

FIG. 1. (a) MODIS land cover product [based on the University of Maryland (UMD) classification] overlaid by the 25-km spatial interpolation grid of the study domain, and (b) 3-yr (2004–06) mean WSR-88D rainfall (mm h^{-1}) over the study domain. The figure shown here corrects a mistake in Fig. 1b of Maggioni et al. (2011).

schemes and provide an overview of the experiment setup. In section 4 we present and discuss the results. In section 5 we provide conclusions on the major findings of this study and discuss future research directions.

2. Study region and data

We refer the reader to Maggioni et al. (2011) for a detailed description of study area and datasets employed in this study; we provide here only the most important information. The study is conducted for a 3-yr period (2004–06) and for a domain in Oklahoma in the United States. A 25×25-km^2 resolution Cartesian modeling grid (34.5°–37°N latitude and 100°–94.5°W longitude) was chosen as the domain for this study (Fig. 1). The Moderate Resolution Imaging Spectroradiometer (MODIS) land cover product map illustrates that the study region is mainly covered by croplands and grasslands. As shown by the 3-yr-average rainfall map in Fig. 1, the western half of the study region is characterized by drier conditions than the wetter eastern half, which displays some mixed cover and broadleaf forests.

The rainfall forcing for the land surface model is from the National Weather Service (NWS) rain gauge–calibrated radar rainfall and the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) morphing (CMORPH) multisatellite rainfall product, representing the reference and ''uncertainty'' rainfall, respectively. Radar rainfall fields are derived from the stage-IV NWS Weather Surveillance Radar-1988 Doppler (WSR-88D) precipitation estimation algorithm at 4-km/1-h spatiotemporal resolution with real-time adjustment based on mean-field radar–rain gauge hourly accumulation comparisons (Fulton 1998; Lin et al. 2005). The radar precipitation dataset was regridded to the 25×25-km^2 resolution modeling grid and aggregated to a 3-hourly

time step. The CMORPH product is based on a combination of passive microwave retrievals from low earth orbiting satellites and geostationary satellite window channel infrared data (Joyce et al. 2004). The infrared data are used to propagate the relatively high-quality precipitation estimates derived from passive microwave data. The CMORPH product is available at the same space (25 km) and time (3 hourly) resolution as the modeling grid domain.

For this study, CMORPH rainfall estimates (uncertain rainfall fields) were adjusted to the mean climatology of the radar rainfall (reference rainfall fields) to meet the assimilation system assumption of unbiased rainfall forcing fields. The bias adjustment factor was determined as the ratio of WSR-88D to CMORPH 3-yr time series domain-average rainfall estimates, and it was found to be 0.66. The remaining surface meteorological forcing data (e.g., air temperature and humidity, radiation, and wind speed) are from the Global Land Data Assimilation System (GLDAS) project (Rodell et al. 2004; http://ldas.gsfc.nasa.gov) based on output from the global atmospheric data assimilation system at the NASA GMAO (Bloom et al. 2005).

3. Methodology

a. The land data assimilation system

The GMAO-LDAS, which constitutes the modeling and assimilation framework of this study, utilizes the NASA Catchment land surface model (CLSM; Koster et al. 2000) to simulate soil moisture and other land surface parameters from meteorological forcing. CLSM includes an explicit treatment of subgrid soil moisture variability and its effect on runoff and evaporation. Within each irregularly shaped computational element, the variability of soil moisture is related to three bulk soil moisture variables: one representing equilibrium conditions associated with water table distribution, and the other two representing nonequilibrium conditions near the surface. The Catchment model and its soil and vegetation parameters are components of the atmospheric general circulation model of the NASA Goddard Earth Observing System version 5 (Rienecker et al. 2008). In this study, land-only model integrations were initialized from a spinup simulation conducted with the WSR-88D radar precipitation by looping three times through the 3 years of forcing data. Previous studies have demonstrated the Catchment model's viability for soil moisture modeling (Bowling et al. 2003; Nijssen et al. 2003; Boone et al. 2004; among others). For further details about the model and its performance in the Oklahoma study region, we refer the reader to Maggioni et al. (2011).

In a land data assimilation system, the model-generated soil moisture is corrected toward the observational estimate. In GMAO-LDAS, soil moisture assimilation is based on the EnKF technique, which is a Monte Carlo variant of the Kalman filter (Evensen 1994) and based on the idea that a small ensemble of model trajectories captures the relevant parts of the error structure. Because of its flexibility with respect to the type of model error, the ensemble approach is widely used in hydrologic data assimilation and is appropriate for the nonlinear character of land surface processes (Reichle et al. 2002; Crow and Wood 2003). The algorithm steps recursively through time, alternating between a model propagation step and a data assimilation update step when observations are available. During the model propagation step, the ensemble members are perturbed by assumed model and forcing errors (Reichle et al. 2007). At the update step, the model forecast is adjusted toward the observational estimate based on the relative uncertainties of the observations and the model forecast and based on cross correlations of the observed variables and the updated model states. In this study, the land data assimilation system propagates the surface soil moisture information into deeper soil layers to produce a superior product (i.e., root zone soil moisture). In this way the land assimilation system can add significant value to surface soil moisture estimates from satellite missions.

Soil moisture estimates from a land assimilation system are sensitive to model and observation error covariances and may even be worse than open-loop model estimates if input error parameters are poor (Crow and Van Loon 2006). As these error parameters are difficult to determine, adaptive filtering methods have been developed to estimate model and observation error parameters. However, for soil moisture, a nonadaptive EnKF is shown to perform well, even when obviously wrong input error parameters are employed, and produces better soil moisture estimates with respect to the open-loop experiment (Reichle et al. 2008).

b. The rainfall error models

The key objective of the study is to contrast the standard rainfall error model used in the GMAO-LDAS (named CRTL) to the more elaborate SREM2D rainfall error model in terms of their ability to characterize model-predicted uncertainty in the context of soil moisture assimilation. The CTRL approach assumes a perfect delineation of rainy and nonrainy areas and scales the precipitation forcing based on an ensemble of multiplicative perturbation fields that are correlated in space and in time (different scaling factor for each time, location, and ensemble member). All ensemble members will differ only in terms of rainfall amount, but they will agree in

terms of rain occurrence. In this study, the temporal correlation was set to zero for compatibility with the SREM2D implementation. The rainfall error structure modeled by the CTRL approach is straightforward in its implementation and parsimonious in its input parameter requirements; however, it can only approximate the variability in the rain-rate errors, since it does not take into account missed rain detection and false alarm—uncertainties that are characteristic of satellite retrievals (Hossain and Anagnostou 2006b).

By contrast, SREM2D characterizes rainfall errors more completely (Hossain and Anagnostou 2006a). Like the CTRL rainfall error model, SREM2D employs stochastic space–time formulations to characterize the multidimensional error structure of satellite retrievals. However, SREM2D allows more complexity in the error modeling structure of rainfall. Whereas both models describe the spatial variability of rain-rate estimation error, the major characteristic of the error structure in satellite rainfall estimation, which is modeled by SREM2D and not by CTRL, is the joint probability of successful delineation of rainy and nonrainy areas accounting for a spatial structure. Unlike CTRL, SREM2D may, in fact, introduce rain in areas where the satellite does not detect rain (missed rain detection). Moreover, SREM2D may assign zero rain where the satellite measures rain (false alarms).

The input parameters for the SREM2D and CTRL error models are identical to those used in Maggioni et al. (2011). The mean value of the lognormal multiplicative perturbation was set to unity in order to obtain (nearly) unbiased replicates in both models. The remaining parameters were carefully calibrated in order to obtain replicates of CMORPH precipitation that reproduce the overall standard deviation of the CMORPH versus radar rainfall errors. Maggioni et al. (2011) also observed that, after calibration, the two error schemes could capture the magnitude of the rainfall error and adequately describe the satellite error variability across scales.

c. *The assimilation experiment setup*

Figure 2 describes the setup of the assimilation experiments. The left side of the schematic provides the reference soil moisture fields derived from the Catchment land surface model forced with gauge-calibrated radar rainfall fields. This single unperturbed hydrologic model realization generates soil moisture fields that represent the reference against that which the model and assimilation estimates are evaluated. These reference soil moisture fields are also used to generate synthetic observations of near-surface soil moisture by perturbing the corresponding simulations with a random field generator representing satellite retrieval error of near-surface soil moisture. Perturbations are set to have a zero mean and a standard deviation of 0.04 $m^3\ m^{-3}$, which is consistent with the target accuracy for satellite soil moisture retrievals (Entekhabi et al. 2010a). In this study we used a one-dimensional EnKF and no spatial correlation structure was imposed to soil moisture perturbations.

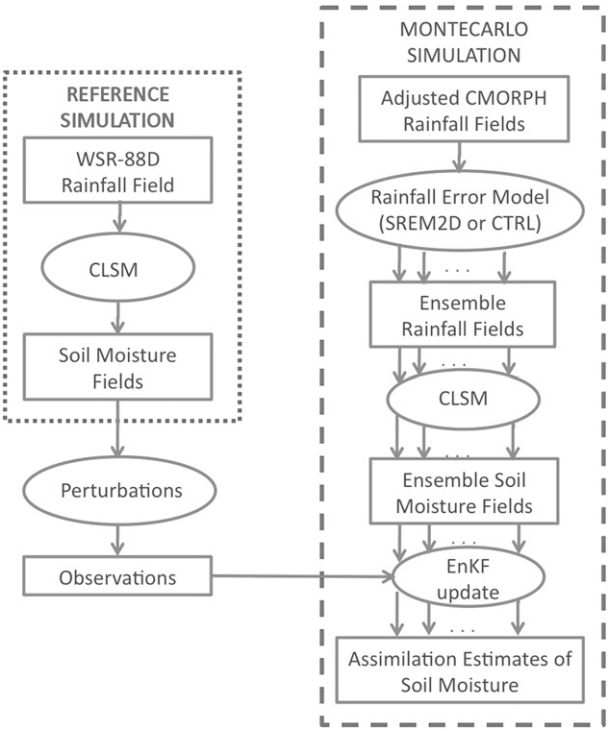

FIG. 2. Experimental setup showing (left) the generation of reference and synthetic soil moisture observations and (right) the ensemble model simulations based on the two rainfall error models and the assimilation activation (on/off).

The right box of the schematic in Fig. 2 describes a Monte Carlo simulation where CMORPH satellite rainfall is perturbed in two different and separate ways: first, using the (standard) CTRL rainfall error model, and second, using the more complex SREM2D rainfall error model. Each of these processes produces an ensemble of satellite rain realizations that are used to force the Catchment model and generate an ensemble of soil moisture fields. Four experiments are carried out: each of the two satellite rainfall error model ensembles is used in GMAO-LDAS without assimilation (open-loop runs) and with the assimilation of the synthetic near-surface soil moisture observations. The output from each of these four experiments is then compared against the reference soil moisture fields to derive performance metrics. Results are described in the next section.

The performance metrics are calculated for both surface and root zone soil moisture based on the 25 × 25-km^2 resolution data, which is a resolution typically used by current global satellite rainfall and soil moisture

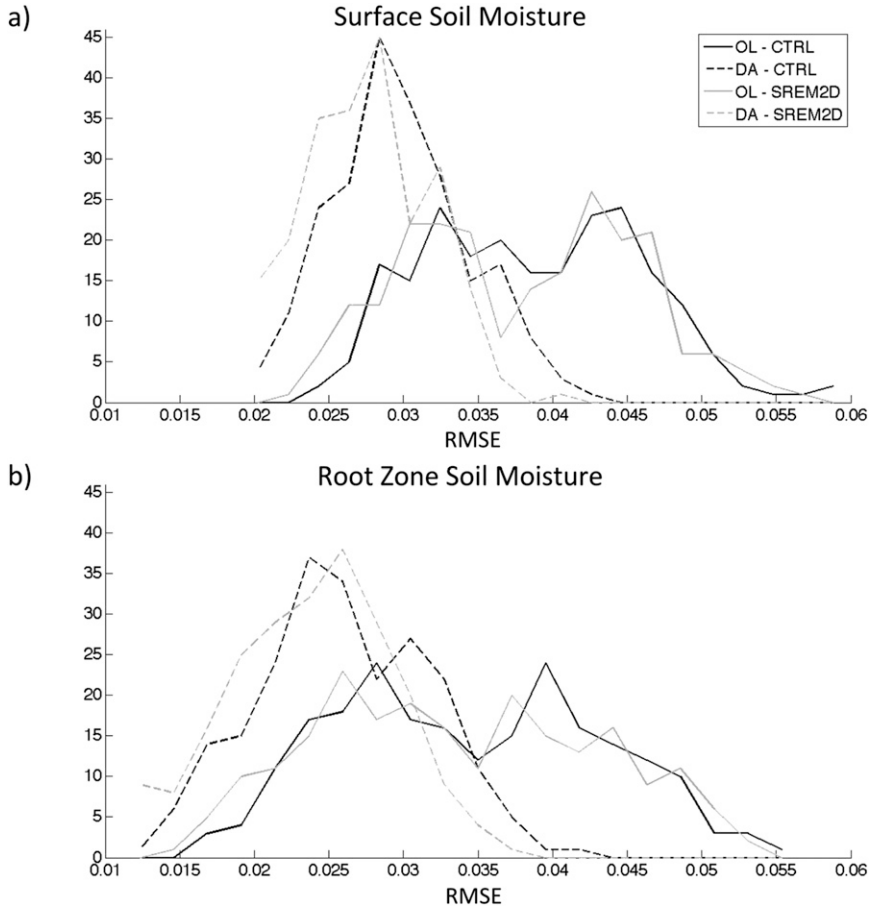

FIG. 3. Distribution of RMSE computed for the 3-yr time series at each grid cell for (a) surface and (b) root zone soil moisture.

products. ''Surface'' soil moisture refers to the (0–2) cm surface soil moisture output from the Catchment model. ''Root zone'' soil moisture is defined here as the (0–100) cm soil moisture output from the Catchment model. Two performance metrics are analyzed for the four experiments: the root-mean-square error (RMSE) and the anomaly correlation coefficient (ACC).

For each (open loop or assimilation) experiment, the RMSE is computed separately for each grid cell from the differences between the ensemble mean of simulated soil moisture and the corresponding reference soil moisture. The ACC is also determined separately for each grid cell based on soil moisture anomaly time series. Anomalies of surface or root zone soil moisture are defined as differences between the actual values and the monthly climatological average values of the 3-yr time series. The ACC metric captures the correspondence in phase between model estimates and the reference, regardless of potential long-term or seasonal mean biases or differences in dynamic range (Entekhabi et al. 2010b).

4. Results and discussion

a. Aggregate performance metrics

We first demonstrate the well-known fact that the assimilation of soil moisture observations improves over the model (open loop) estimates. Figure 3 shows the distributions of RMSE values for the open-loop and data assimilation experiments using the CTRL rainfall error model. Data assimilation shifts the distribution to lower RMSE values. The RMSE distributions also show a reduction in the upper tail of the distributions when assimilation is included—namely, values do not exceed 0.045 $m^3\ m^{-3}$. Specifically, the RMSE of open-loop simulated soil moisture shows a bimodal distribution, with values that spread between 0.020 (0.010) and 0.060 (0.055) $m^3\ m^{-3}$ for surface soil moisture (root zone soil moisture). When assimilation is turned on, the shape of the distributions changes; the values distribute closer to the mean and shift to lower RMSE values. The mean RMSE for the open-loop (OL) case, shown in Table 1, is 0.039 (0.034) $m^3\ m^{-3}$, which decreases to 0.030 (0.026)

TABLE 1. RMSEs and 95% confidence intervals for (second column) OL and (third column) DA experiments, and (fourth column) the corresponding relative RMSE reductions due to DA. Numbers in regular (italic) font are for GMAO-LDAS integrations using the CTRL (SREM2D) rainfall error model, and ΔP in the fourth column indicates the rainfall regime (section 4b).

			Relative RMSE reduction 100 × (DA − OL) / OL		
CTRL *SREM2D*	OL	DA	All	$\Delta P < 0$	$\Delta P > 0$
Surface soil moisture	0.039 ± 0.001	0.030 ± 0.001	23.1%	17.8%	27.1%
	0.038 ± 0.001	*0.028 ± 0.001*	*26.3%*	*20.1%*	*31.4%*
Root zone soil moisture	0.034 ± 0.001	0.026 ± 0.001	23.5%	19.2%	27.7%
	0.034 ± 0.001	*0.024 ± 0.001*	*29.4%*	*22.5%*	*32.1%*

$m^3\ m^{-3}$ in the assimilation integration for surface soil moisture (root zone soil moisture). This corresponds to a relative reduction of 23.1% for surface soil moisture and 23.5% for root zone soil moisture due to data assimilation (DA). As indicated by 95% confidence intervals (Table 1), the reduction in the RMSE introduced by assimilation is statistically significant.

The ACC distributions show a similar picture (Fig. 4): data assimilation shifts the distribution toward higher values. The lowest ACC values increase from ~0.65 in the open-loop simulations to 0.75, while the maximum values exceed 0.95 in both surface and root zone soil moisture when assimilation is turned on. Table 2 shows that the mean ACC for the open-loop simulation using the CTRL rainfall error model is 0.81 (0.79), which increases to 0.88 (0.89) in the data assimilation experiment for surface soil moisture (root zone soil moisture). This translates into a relative increase of 8.6% for surface soil moisture and 12.7% for the deeper soil moisture when assimilation is turned on. Confidence intervals indicate that the increase in ACC obtained through assimilation is significant at the 5% confidence level.

The improvements from data assimilation are slightly better when the SREM2D rainfall error model is used in the assimilation system. Figure 3 also shows the corresponding results for the open-loop and assimilation integrations using SREM2D. No significant difference can be noticed in the performance metrics of the two error models in the open-loop cases, whereas in the data assimilation integrations the RSME distributions are more skewed toward lower values. This is also shown in Table 1. With SREM2D, a relative RMSE reduction of 26.3% is obtained for surface soil moisture when assimilation is on, compared to the 23.1% improvement in the CTRL case. Similarly, a 29.4% RMSE reduction is observed for root zone soil moisture, which compares to the 23.5% reduction obtained with the CTRL error model.

Analogously, Fig. 4 shows the corresponding open-loop and assimilation ACC distributions for the SREM2D rainfall error model. In the SREM2D data assimilation case, the distributions are more shifted toward higher ACC values for both surface and root zone soil moisture (compared to assimilation with CTRL). As shown in Table 2, the mean relative ACC increase due to assimilation using SREM2D is 11.3% for surface soil moisture, which compares to the 8.6% increase in the CTRL experiment. For root zone soil moisture, the relative increase of ACC due to assimilation is 13.9%, which compares to the 12.7% increase in the CTRL error model. In section 4c, the difference in the performances of the two rainfall error models will be investigated further with respect to the different rainfall climatological conditions.

b. *Innovations statistics*

Innovations are defined as the difference between the synthetic observations and the corresponding ensemble mean model values prior to the assimilation update. If the filter operates in accordance with its underlying assumptions and model and observation error parameters are appropriately chosen, the mean of the innovations should be statistically indistinguishable from zero, and the normalized innovations (defined as innovations divided by their expected standard deviation) should approximately obey a standard normal distribution with zero mean and standard deviation equal to 1 (Reichle et al. 2007).

The innovation mean values of the assimilation experiment runs with the CTRL and SREM2D rainfall

TABLE 2. As in Table 1, but for ACC.

			Relative ACC increase 100 × (DA − OL)/OL		
CTRL *SREM2D*	OL	DA	All	$\Delta P < 0$	$\Delta P > 0$
Surface soil moisture	0.81 ± 0.01	0.88 ± 0.01	8.6%	8.0%	10.3%
	0.80 ± 0.01	*0.89 ± 0.01*	*11.3%*	*9.6%*	*12.8%*
Root zone soil moisture	0.79 ± 0.01	0.89 ± 0.01	12.7%	11.3%	13.6%
	0.79 ± 0.01	*0.90 ± 0.01*	*13.9%*	*13.8%*	*16.2%*

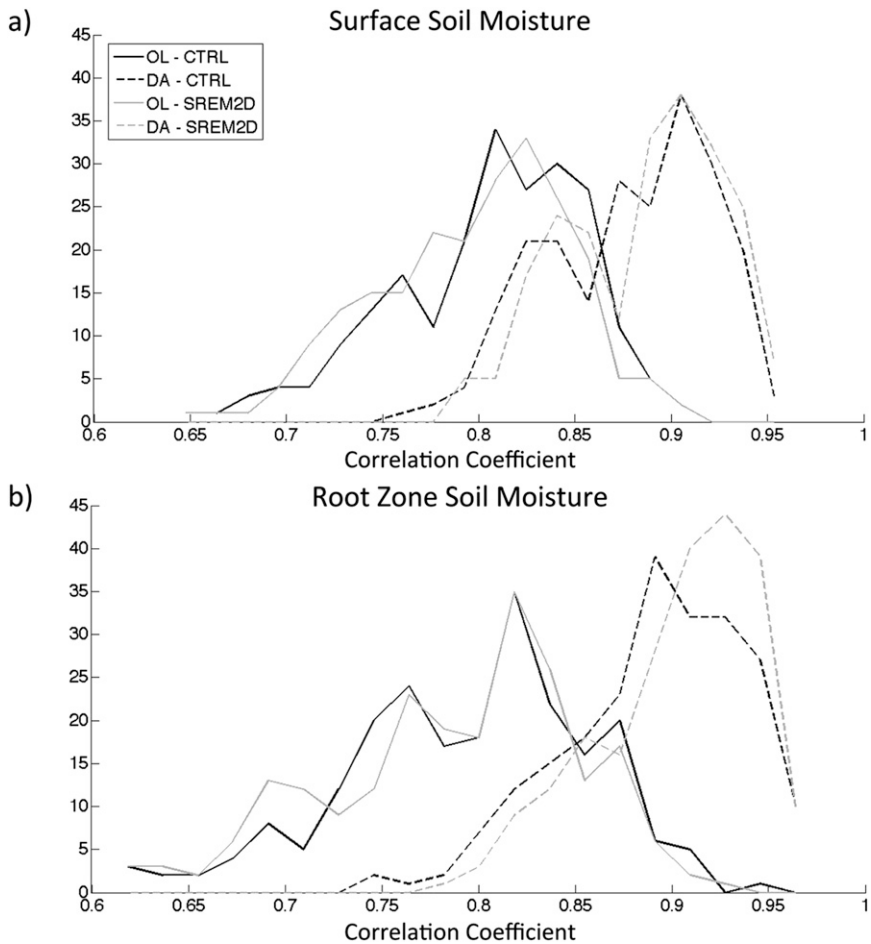

FIG. 4. As in Fig. 3, but for ACC.

error models are 0.006 and 0.005 $m^3\ m^{-3}$, respectively (not shown), which is considered approximately zero. Next, the normalized ensemble mean innovations are computed for the two assimilation experiments. Figure 5 shows maps of standard deviations of the normalized ensemble mean innovations (dimensionless) for the 3-yr time series and for both assimilation experiments. The figure shows that the standard deviations of normalized ensemble innovations are higher than 1 across the entire study region, with an average of 1.18 (1.15) for the CTRL (SREM2D) simulation. This is consistent with results presented by Reichle et al. (2007) at a global scale, who showed that the variance of normalized innovations exceeds 1 in central North America when assimilating two different sets of global soil moisture retrievals. This suggests either an underestimation of the observation error variance or a need to account for larger modeling error. Slight improvements are observed overall when SREM2D is employed, exhibiting lower standard deviation values that are closer to 1 than in the case of the CTRL simulation.

c. *Impact of climatological rainfall regime*

For further analysis we introduce a rainfall climatology parameter ΔP, defined as

$$\Delta P = \frac{P_i - P_{\text{mean}}}{P_{\text{mean}}}, \tag{1}$$

where P_i is the reference WSR-88D rainfall for the ith grid cell averaged over the entire 3-yr period and P_{mean} is the mean radar rainfall value for the entire 10×22 grid area and the 3-yr period. Here ΔP can be interpreted as a climatological wetness indicator of the area covered by each grid cell with respect to the domain average; positive (negative) values of ΔP would indicate areas that are generally wet (dry) with respect to the climatology of the entire domain—defined as the 3-yr-average rainfall value. Four different classes of rainfall climatology are considered: (i) ΔP smaller than -0.2, (ii) ΔP falls between -0.2 and 0, (iii) ΔP falls between 0 and 0.2, and (vi) ΔP is

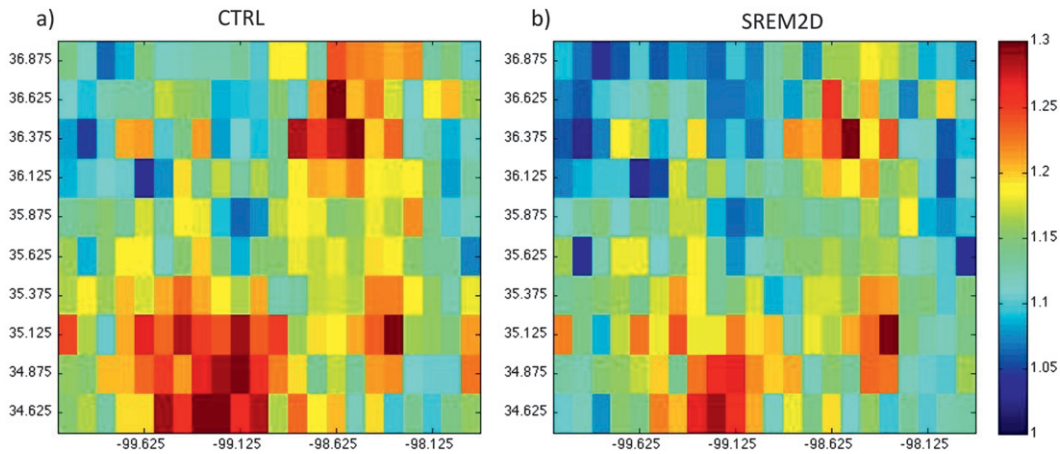

FIG. 5. Maps of standard deviation of normalized ensemble mean innovations evaluated over the 3-yr time series for the (a) CTRL and (b) SREM2D assimilation experiments (dimensionless). The average standard deviation is 1.18 for CTRL and 1.15 for SREM2D.

larger than 0.2. The two negative classes (i) and (ii) represent drier conditions, while the two positive classes (iii) and (iv) correspond to wetter-than-climatology conditions.

Figure 6 shows the relative reduction in the RMSE of (near surface and root zone) soil moisture between the assimilation and open-loop cases, normalized by the RMSE of the open-loop simulation, as a function of the rainfall climatology parameter ΔP. This normalized RMSE difference is always positive, showing a reduction of the random errors due to the assimilation of soil moisture observations. The relative RMSE reduction is higher in wetter conditions for surface and root zone soil moisture. Moreover, the (spatial) variability in the RMSE reduction amplifies with increasing ΔP as indicated by the one-standard-deviation bars in the plots. Analogously, Fig. 7 illustrates the relative increase in soil moisture ACC values between the assimilation and open-loop cases. Unlike for the RMSE reduction, the maximum relative ACC increase is near neutral conditions, which corresponds to nonextreme rainfall regimes. This behavior is apparent for both soil moisture depths.

Figure 6 also indicates that the relative RMSE reduction due to assimilation is higher for the SREM2D error model than for the CTRL error model. Likewise, the relative ACC increases are higher with SREM2D than with the CTRL error model. A hypothesis test was carried out to evaluate the significance of the difference in the performance metric improvements obtained for the two rainfall error models. Specifically, a two-sample t test for the difference in the mean was applied to evaluate the significance in the difference between the mean relative RMSE reductions (or ACC increases) obtained

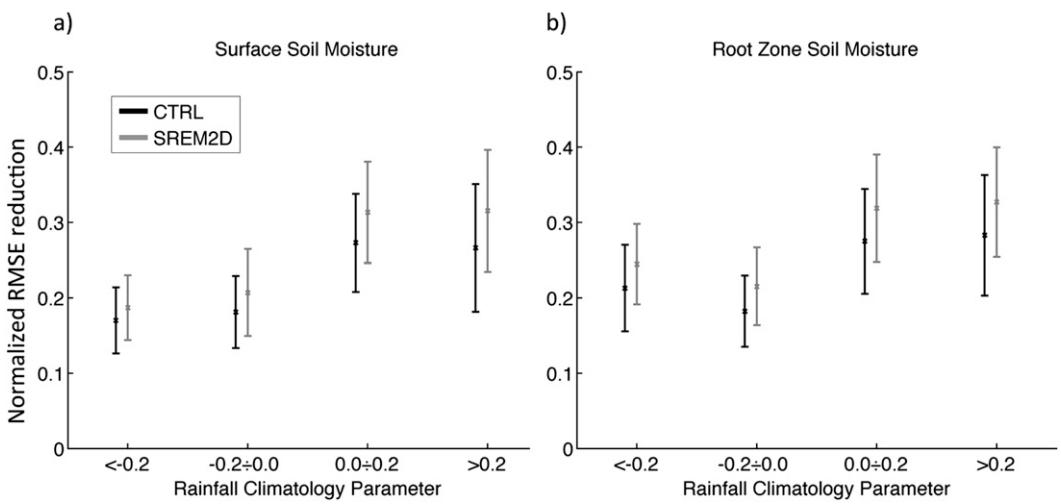

FIG. 6. Normalized RMSE reduction vs rainfall climatology parameter ΔP for (a) surface and (b) root zone soil moisture.

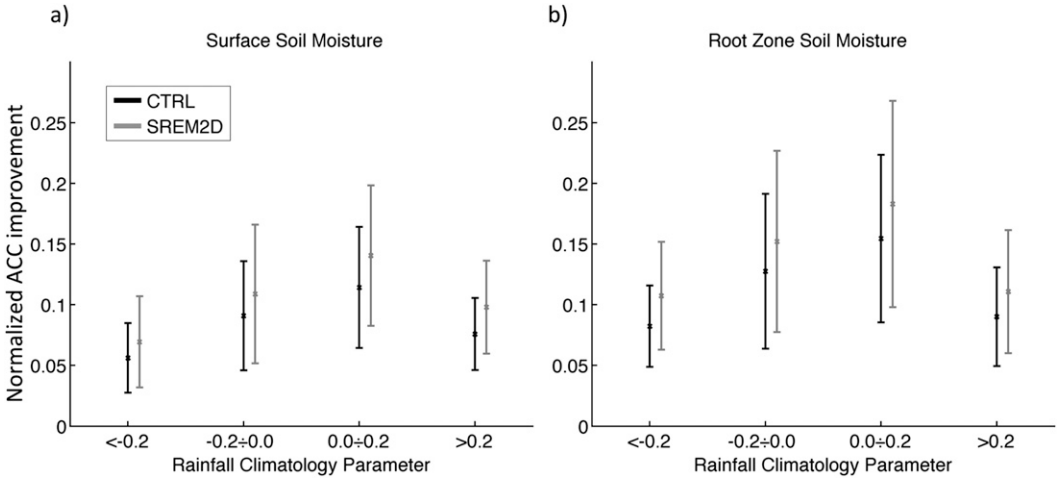

FIG. 7. As in Fig. 6, but for ACC.

by SREM2D versus CTRL. The test is performed using independent samples of the performance metrics from the two experiments. We devise a null hypothesis that the difference in the mean of the two relative RMSE reductions (or ACC increases) is equal to zero, while the alternative hypothesis is that the difference in the mean is different from zero. The significance test is presented for two main classes of rainfall climatology: negative ΔP (drier conditions) and positive ΔP (wetter conditions).

Table 1 summarizes the mean relative RMSE reduction for SREM2D and CTRL for the two rainfall regimes ($\Delta P < 0$ and $\Delta P > 0$) and for both soil moisture depths, and Table 3 lists the corresponding results for the confidence levels at which the null hypothesis is rejected. For surface soil moisture, the RMSE reductions for CTRL and SREM2D differ by 2.3% and are statistically different from zero at the 20% confidence level in dry climatological conditions. In wet conditions, the surface soil moisture RMSE reductions for CTRL and SREM2D differ by 4.3% and are statistically different at the 15% confidence level. In the case of root zone soil moisture, the differences in RMSE reduction are 3.3% in dry conditions and 4.4% in wet conditions. These differences are statistically different from zero at the 5% significance level. In other words, Table 3 shows that the root zone soil moisture RMSE values obtained with SREM2D are better than those obtained with CTRL at the 5% significance level.

TABLE 3. Significance levels at which the null hypothesis of equal mean for relative RMSE reductions (values shown in Table 1) is rejected.

	$\Delta P < 0$	$\Delta P > 0$
Surface soil moisture	20%	15%
Root zone soil moisture	5%	5%

Tables 2 and 4 present the analogous analysis for ACC. For near-surface soil moisture the difference in the mean ACC increase between SREM2D and CTRL is 1.6% in dry rainfall regimes—a difference that is statistically different from zero at the 25% significance level. In moist conditions the difference in the mean of ACC increase between the two rainfall error models is 2.5% and the null hypothesis of equal means is rejected at 25% confidence level. Similar results are obtained for root zone soil moisture, where the difference between the ACC increase for SREM2D and CTRL is significant at the 25% level in both dry and wet rainfall regimes.

The improvement from using SREM2D in the assimilation system is higher in wet rainfall regimes than in dry rainfall regimes for both surface and root zone soil moisture (i.e., larger RMSE reductions and ACC increases). This is expected as soil water content is highly dependent on rainfall variability and consequently varies considerably in response to rainfall forcing. In summary, our numerical experiment shows that the use of the more elaborate SREM2D rainfall error model provides slightly higher mean relative RMSE reductions and ACC increases than the CTRL rainfall error model when used in the assimilation of synthetic soil moisture observations within GMAO-LDAS, especially in the wet rainfall regime. The greatest impact from using the SREM2D error model in data assimilation is in the RMSE reduction of root zone soil moisture in both dry and wet rainfall regimes (significant at 5% confidence level).

TABLE 4. As in Table 3, but for ACC increases.

	$\Delta P < 0$	$\Delta P > 0$
Surface soil moisture	25%	25%
Root zone soil moisture	25%	25%

5. Conclusions

This study investigated the effect of two satellite rainfall error models of different complexity on the efficiency of assimilating (synthetic) surface soil moisture observations. Specifically, four numerical experiments have been run by perturbing input model precipitation (separately) with two rainfall error models of different complexity, without (open-loop runs) and with the assimilation of the synthetic near-surface soil moisture (assimilation runs). Surface and root zone soil moisture outputs from each experiment were compared against the ''reference'' soil moisture to derive performance metrics.

This comparison showed that assimilation provides an improvement in terms of lower RMSE values and higher ACC values. Merging synthetic surface soil moisture into the model through data assimilation not only leads to better estimates of surface soil moisture but also to an increase in skill for the root zone soil moisture. Examining the average metrics, slightly better improvements due to assimilation were obtained with the more elaborate SREM2D rainfall error model compared to the CTRL error model, while performances of the two error models are very similar in the open-loop simulations.

An analysis of the innovation statistics showed that the innovation means are close to zero in both CTRL and SREM2D assimilation experiments. Standard deviations of normalized ensemble mean innovations are shown to be close to the expected value of 1 (dimensionless), with slightly better results in the case of SREM2D. This demonstrates that the filter largely operates in accordance with its assumptions.

A further investigation based on different rainfall climatological regimes showed that the relative improvement due to the assimilation of soil moisture observations with respect to the open-loop case is higher when SREM2D is employed (i.e., larger relative RMSE reductions and ACC increases). The relative enhancement due to the use of SREM2D is shown to be larger for root zone soil moisture, which carries the memory of previous precipitation events, and in wetter rainfall conditions, which is ascribed to the high dependence of soil moisture on precipitation variability. In summary, using a more complex rainfall error model, the relative improvement provided by the data assimilation slightly enhances surface and root zone soil moisture estimates, exhibiting lower RMSE and higher ACC values. Regardless of the results of the assimilation, the use of a more sophisticated error model, which includes a more descriptive characterization of uncertainties, is encouraged for rainfall and other model forcing variables and parameters. In addition, a more inclusive description of rainfall uncertainties could be obtained by taking into account for temporal correlation within the error models. A supplemental analysis (not shown) suggests that the temporal error correlation coefficient for CMORPH precipitation (versus WSR-88D estimates) is less than 0.3 for the periods and storms examined in this study.

Further studies should investigate the benefit of the more sophisticated rainfall error model at the event scale. Specifically, accounting for uncertainties in the presence (or absence) of rainfall might be more significant during summer months with large but spatially limited convective storms than during winter months, when stratiform rainfall dominates. Moreover, the work presented in this study is a controlled experiment that makes use of synthetic soil moisture observations, and that is based on the assumptions of unbiased rainfall forcing and zero mean perturbations of soil moisture. A challenging step for future research is to demonstrate improvements based on actual satellite soil moisture retrievals, where these assumptions might not hold.

The broader impacts of this study include a direct contribution to the development of the NASA Global Modeling and Assimilation Office (GMAO) land data assimilation system. In particular, the study provides valuable insights about the use of satellite rainfall data and associated error structure for modeling hydrologic processes and useful feedback to future satellite hydrologic missions, such as the Global Precipitation Measurement mission and NASA's Soil Moisture Active Passive mission.

The presented work was done with a view toward a global-scale land data assimilation, for which the GMAO-LDAS has been applied successfully (Reichle et al. 2007). But our results are strictly valid only for a specific assimilation system—the GMAO-LDAS—which is built on the Catchment land surface model and the ensemble Kalman filter. Likewise, we concentrated only on the Oklahoma region for the experiments. Future studies should investigate the sensitivity of these results to different combinations of land models and assimilation methods, and to other hydroclimatic regimes. However, we are confident that our general conclusions can be extended to different land data assimilation approaches and can be transferred to regions of the world with climatology and terrain that are not too different from that of the Oklahoma region.

Acknowledgments. V. Maggioni was supported by a NASA Earth System Science Graduate Fellowship. R. Reichle was supported by the NASA research program ''The Science of Terra and Aqua'' and the SMAP Science Definition Team. E. Anagnostou was supported by NASA Precipitation Science Team Grant NNX07AE31G. Computing was supported by the NASA High End Computing Program.

REFERENCES

Betts, A. K., and J. H. Ball, 1998: FIFE surface climate and site-average dataset 1987–89. *J. Atmos. Sci.,* **55,** 1091–1108.

Bloom, S., and Coauthors, 2005: Documentation and validation of the Goddard Earth Observing System (GEOS) Data Assimilation System: Version 4. Tech. Rep. Series on Global Modeling and Data Assimilation NASA/TM-2005-104606, Vol. 26, 187 pp.

Boone, A., and Coauthors, 2004: The Rhône-Aggregation Land Surface Scheme intercomparison project: An overview. *J. Climate,* **17,** 187–208.

Bowling, L. C., and Coauthors, 2003: Simulation of high latitude hydrological processes in the Torne–Kalix basin: PILPS Phase 2(e): 1: Experiment description and summary intercomparisons. *Global Planet. Change,* **38,** 1–30.

Crow, W. T., and E. F. Wood, 2003: The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. *Adv. Water Resour.,* **26,** 137–149.

——, and E. Van Loon, 2006: Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture. *J. Hydrometeor.,* **7,** 421–432.

Dorigo, W. A., P. Van Oevelen, W. Wagner, M. Drusch, S. Mecklenburg, A. Robock, and T. Jackson, 2011: A new international network for in situ soil moisture data. *Eos, Trans. Amer. Geophys. Union,* **92,** 141–142.

Drusch, M., K. Scipal, P. de Rosnay, G. Balsamo, E. Andersson, P. Bougeault, and P. Viterbo, 2009: Towards a Kalman Filter based soil moisture analysis system for the operational ECMWF Integrated Forecast System. *Geophys. Res. Lett.,* **36,** L10401, doi:10.1029/2009GL037716.

Entekhabi, D., and Coauthors, 2010a: The Soil Moisture Active Passive (SMAP) mission. *Proc. IEEE,* **98,** 704–716.

——, R. H. Reichle, R. D. Koster, and W. T. Crow, 2010b: Performance metrics for soil moisture retrievals and application requirements. *J. Hydrometeor.,* **11,** 832–840.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5),10 143–10 162.

Fulton, R. A., 1998: WSR-88D polar-to-HRAP mapping. Hydrologic Research Laboratory, Office of Hydrology, National Weather Service Tech. Memo., 34 pp.

Hossain, F., and E. N. Anagnostou, 2005: Using a multidimensional satellite rainfall error model to characterize uncertainty in soil moisture fields simulated by an offline land surface model. *Geophys. Res. Lett.,* **32,** L15402, doi:10.1029/2005GL023122.

——, and ——, 2006a: A two-dimensional satellite rainfall error model. *IEEE Trans. Geosci. Remote Sens.,* **44,** 1511–1522.

——, and ——, 2006b: Assessment of a multidimensional satellite rainfall error model for ensemble generation of satellite rainfall data. *Geosci. Remote Sens. Lett.,* **3,** 419–423.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeor.,* **5,** 487–503.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model 1. Model structure. *J. Geophys. Res.,* **105** (D20), 24 809–24 822.

Li, F., W. T. Crow, and W. P. Kustas, 2010: Towards the estimation root-zone soil moisture via the simultaneous assimilation of thermal and microwave soil moisture retrievals. *Adv. Water Resour.,* **33,** 201–214.

Lin, C., S. Vasić, A. Kilambi, B. Turner, and I. Zawadzki, 2005: Precipitation forecast skill of numerical weather prediction models and radar nowcasts. *Geophys. Res. Lett.,* **32,** L14801, doi:10.1029/2005GL023451.

Liu, Q., and Coauthors, 2011: The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system. *J. Hydrometeor.,* **12,** 750–765.

Maggioni, V., R. H. Reichle, and E. N. Anagnostou, 2011: The effect of satellite rainfall error modeling on soil moisture prediction uncertainty. *J. Hydrometeor.,* **12,** 413–428.

Nijssen, B., and Coauthors, 2003: Simulation of high latitude hydrological processes in the Torne-Kalix basin: PILPS Phase 2(e): 2: Comparison of model results with observations. *Global Planet. Change,* **38,** 31–53.

Parajka, J., V. Naeimi, G. Blöschl, W. Wagner, R. Merz, and K. Scipal, 2006: Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale. *Hydrol. Earth Syst. Sci.,* **10,** 353–368, doi:10.5194/hess-10-353-2006.

Peters-Lidard, C. D., and Coauthors, 2007: High-performance Earth system modeling with NASA/GSFC's Land Information System. *Innovations Syst. Software Eng.,* **3** (3), 157–165.

Reichle, R. H., and R. D. Koster, 2005: Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model. *Geophys. Res. Lett.,* **32,** L02404, doi:10.1029/2004GL021700.

——, D. McLaughlin, and D. Entekhabi, 2002: Hydrological data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.,* **130,** 103–114.

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

——, W. T. Crow, and C. L. Keppenne, 2008: An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resour. Res.,* **44,** W03423, doi:10.1029/2007WR006357.

Rienecker, M. M., and Coauthors, 2008: The GEOS-5 Data Assimilation System—Documentation of versions 5.0.1, 5.1.0, and 5.2.0. Tech. Rep. Series on Global Modeling and Data Assimilation NASA/TM-2008-104606, Vol. 27, 101 pp. [Available online at http://gmao.gsfc.nasa.gov/pubs/docs/Rienecker369.pdf.]

Robinson, D. A., and Coauthors, 2008: Soil moisture measurements for ecological and hydrological watershed scale observatories: A review. *Vadose Zone J.,* **7,** 358–389, doi:10.2136/vzj2007.0143.

Robock, A., K. Ya Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai, 2000: The global soil moisture data bank. *Bull. Amer. Meteor. Soc.,* **81,** 1281–1299.

Schmugge, T. J., W. P. Kustas, J. C. Ritchie, T. J. Jackson, and A. Rango, 2002: Remote sensing in hydrology. *Adv. Water Resour.,* **25,** 1367–1385.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, 2010: Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Sci. Rev.,* **99,** 125–161, doi:10.1016/j.earscirev.2010.02.004.

Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.,* **85,** 381–394.

Walker, J. P., G. R. Willgoose, and J. D. Kalma, 2004: In situ measurement of soil moisture: A comparison of techniques. *J. Hydrol.,* **293,** 85–99.