

Assimilation and downscaling of satellite observed soil moisture over the Little River Experimental Watershed in Georgia, USA

Alok Kumar Sahoo^{a,*}, Gabriëlle J.M. De Lannoy^{b,c}, Rolf H. Reichle^c, Paul R. Houser^d

^a Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, USA

^b Laboratory of Hydrology and Water Management, Ghent University, Coupure links 653, B-9000 Ghent, Belgium

^c Global Modeling and Assimilation Office (Code 610.1), NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

^d College of Sciences, George Mason University, Fairfax, VA, USA

ARTICLE INFO

Article history:

Received 27 May 2011

Received in revised form 21 August 2012

Accepted 23 August 2012

Available online 7 September 2012

Keywords:

Data assimilation

Kalman filter

Soil moisture

Multi-scale

Little River Experimental Watershed

Satellite observations

ABSTRACT

A three dimensional Ensemble Kalman filter (3-D EnKF) and a one dimensional EnKF (1-D EnKF) are used in this study to assimilate Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) coarse grid (25 km) soil moisture retrievals into the Noah land surface model for fine-scale (1 km) surface soil moisture estimation over the Little River Experimental Watershed (LREW), Georgia, USA. For the 1-D EnKF integration, the satellite observations are a priori partitioned to the model fine scale resolution, whereas in the 3-D EnKF integration, the original coarse grid satellite observations are directly used and downscaling is accomplished within the 3-D EnKF update step. In both cases, a first order a priori forecast bias correction is applied. Validation against in situ observations shows that both EnKF algorithms improve the soil moisture estimates, but the 3-D EnKF algorithm better preserves the spatial coherence. It is illustrated how surface soil moisture assimilation affects the deeper layer soil moisture and other water budget variables. Through sensitivity experiments, it is shown that data assimilation accelerates the moisture redistribution compared to the model integrations without assimilation, as surface soil moisture updates are effectively propagated over the entire profile. In the absence of data assimilation, the atmospheric conditions (especially the ratio of evapotranspiration to precipitation) control the model state balancing.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Soil moisture and sea surface temperature are two important initialization parameters for seasonal to interannual climate prediction. Soil moisture variations exert control over land-atmosphere energy flux, and therefore may have a large influence on climate change over land [58,57]. Soil moisture also influences the hydrologic cycle from local to global scales by redistributing rainfall into infiltration, surface runoff and evaporation at the earth surface [17,62,63]; supplying water to the vegetation above the ground [62]; affecting the surface energy exchange [46] and helping determine the Bowen ratio [18]. Therefore, an accurate estimation of soil moisture has recently been a priority for many hydrologic applications.

Yet quantifying soil moisture and capturing its mesoscale variability in space, time and depth is very difficult when using current satellite observations, state-of-the-art land surface models and ground based data separately; each having their own limitations [50,53]. One approach that addresses the limitations of quantifying

soil moisture from any single source is to merge the available observations with land surface models. The current study explores data assimilation techniques using satellite observations and a land surface model for soil moisture estimates to attenuate forecast errors and improve the model's predictive skills. The usefulness of the Earth observing satellites for operational hydrologic applications have not been widely explored until recently [64].

A number of practical issues arise when assimilating a specific observation type into a particular model. One such major issue is the spatial scale mismatch between the satellite observations and land surface model simulations. Passive microwave satellite soil moisture products, such as from the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E), the Soil Moisture and Ocean Salinity (SMOS) mission and the future Soil Moisture Active/Passive (SMAP) mission, are typically available at a spatial resolution of 10–40 km whereas the model forecasts are becoming feasible at a spatial resolution of 1 km over large domains. Many previous studies have applied algorithms to partition the coarse scale spatial observations to the fine scale model grid cells prior to data assimilation. The spatial scale discrepancy can also be addressed dynamically within the data assimilation framework. There is also a vertical scale mismatch between the model surface layer (0.1 m) and the satellite sensing depth (~0.02 m)

* Corresponding author.

E-mail address: sahoo@princeton.edu (A.K. Sahoo).

which may add to the bias between the model and satellite-observed soil moisture values. Another major issue is the mismatch in the climatology or in the dynamic range of the physical variable between satellite observations and land surface model forecasts. This bias can be addressed by implementing techniques like cumulative distribution function (CDF) matching or dynamic bias correction in the data assimilation framework (e.g. [11,50,49,12,13,35,33]).

Over the past few years, sequential data assimilation methods have been applied to soil moisture (e.g. [65,66,9,43,33]), soil temperature (e.g. [25,52]), snow (e.g. [61,1,8,16]) and other hydrologic variables. Most applications have been limited to 1-D (one-dimensional, vertical dimension) filtering, where the observations are assimilated into the land surface model for each model grid cell independently. Only a few studies (e.g. [47,48,44,68,43]) used 3-D (three-dimensional, both horizontal and vertical dimension) data assimilation techniques.

Our study focuses on analyzing the performance of a 3-D EnKF algorithm and makes the following new contributions: (a) use of satellite observations (as opposed to synthetic or in situ observations) in a 3-D EnKF data assimilation framework for soil moisture analysis, (b) application of dynamic downscaling in the context of 3-D data assimilation and (c) assessment of the contributions of land model initialization in the context of data assimilation. Specifically, we answer the following questions: (a) How does the 3-D EnKF algorithm perform for AMSR-E soil moisture assimilation with dynamic downscaling? (b) What is the impact of the updated surface soil moisture on the deeper layer moisture variables and other water cycle variables? (c) What is the effect of assimilation on the moisture vertical redistribution and how sensitive is the 3-D EnKF algorithm to model initialization conditions?

To our knowledge, the dynamic downscaling approach has not yet been used for satellite soil moisture retrievals (as opposed to synthetic observations). Similarly, the interaction between model initialization and assimilation with a 3-D EnKF algorithm has not been tested for soil moisture estimation. De Lannoy et al. [15] have used the same 3-D EnKF filter and downscaling method for snow data assimilation. However, our study differs from De Lannoy et al. [15,16] in that (1) here we investigate soil moisture assimilation rather than snow assimilation and (2) soil moisture is modeled with four soil layers as opposed to a single snow layer and hence the information from the assimilation of surface layer observations propagates downward. Our study also differs from the 3-D EnKF study performed in De Lannoy et al. [14], where in situ observations were assimilated with an adaptive spatial filter.

The layout of this paper is as follows: The study area, the assimilated satellite observations, the in situ observations for validation and bias estimation, the land surface model, and the data assimilation (DA) module are described in Section 2. The experiment setup and a brief summary of the EnKF algorithm are summarized in Section 3. The simulation results are analyzed and discussed in Section 4. Finally, conclusions are provided in Section 5.

2. Description of the study area, observations and land surface model

2.1. Study area

The Little River Experimental Watershed (LREW) located near Tifton, Georgia (Fig. 1) is one of four designated watersheds selected to calibrate and validate the AMSR-E soil moisture observations and was a part of a soil moisture field campaign conducted in June and July, 2003 (SMEX03; <http://www.hydrolab.arsusda.gov/smex03/>). This watershed encompasses an area of 334 km². The main watershed includes seven gauged sub-watersheds ranging

in size from 3 to 115 km². This watershed is in the headwaters of the Suwannee River Basin that begins in Georgia and empties into the Gulf of Mexico. The Little River is a tributary of the Withlacoochee River; one of the two main tributaries of the Suwannee River. The LREW has a flat topography with broad flood plains and is poorly defined by stream channels [60]. A detailed description of the soil properties, vegetation, land use information, physical characteristics, and precipitation patterns over the LREW can be found in earlier studies [67,26,59,45,2,3]. We define our experiment domain as a rectangular 100 km × 75 km area (~1 km² fine scale resolution) as shown in Fig. 1 (left bottom corner: 31°N, 84°W; upper right corner: 32°N, 83.25°W).

2.2. Satellite observed soil moisture data used for assimilation

The assimilated satellite observations are retrieved from the AMSR-E sensor, aboard the Aqua platform, using the Land Surface Microwave Emission Model (LSMEM). LSMEM is an iterative forward model which is based on the radiative transfer theory described in Kerr and Njoku [29]. The soil moisture is retrieved from the AMSR-E 10.65 GHz frequency, horizontal polarization (10.65H) brightness temperature data. A detailed description of the LSMEM model has been provided in Gao et al. [22,23] and Sahoo et al. [54]. The LSMEM soil moisture retrievals were validated at SCAN sites, Oklahoma Mesonet sites and Illinois sites in Continental USA and at OzNet sites in Australia and were found to be well matched with the in situ measurements [55]. We also validated the same product at this LREW study location [54] along with the official AMSR-E soil moisture product where the LSMEM product outperformed the official AMSR-E product [42].

Our study domain contains 4 × 3 coarse 25 × 25 km² satellite observations. For the 1-D EnKF, we perform a priori observation partitioning into 1 km grid cells (assigning the same coarse-scale value to all the 1 km grid cells contained within each 25 km AMSR-E grid cell) whereas we avoid the a priori observation partitioning for the 3-D EnKF algorithm. The Aqua equatorial overpass is twice a day at 1:30 am (descending pass) and 1:30 pm (ascending pass) with a revisit time of 2–3 days. Therefore, the study domain is often only partially observed. Non-observed areas are updated through the spatial 3-D EnKF filter, but are excluded from the update in the 1-D EnKF.

2.3. LREW in situ observation data used for validation and bias estimation

Fig. 1 locates the in situ observations that are used to validate the data assimilation results. There is a Soil Climate Analysis Network (SCAN) site at the southern end of the watershed. As a part of the AMSR-E calibration and validation project, a network of instruments [3] have been installed at seventeen rain-gauge sites since 2001 to monitor soil water and temperature every half hour at 0.05 m, 0.2 m and 0.3 m depths located within the LREW (Fig. 1) [5].

Field observation data were obtained from the United States Department of Agriculture-Agricultural Research Service (USDA-ARS) located in Beltsville, Maryland [28,27] and USDA-ARS Southeast watershed research laboratory located in Tifton, Georgia [4]. The data used in this study include all available instantaneous soil moisture and precipitation measurements for the period 2002–2007 at 16 out of the 17 sites. These 16 sites are 12, 16, 22, 26, 31, 32, 34, 39, 40, 43, 50, 62, 63, 66, 67 and 75 (Fig. 1). A detailed description of these in situ data can be found in Jackson et al. [28] and Sahoo et al. [53]. Again, unlike the satellite observations, the in situ soil moisture data in this study are primarily used for validation and bias estimation purposes.

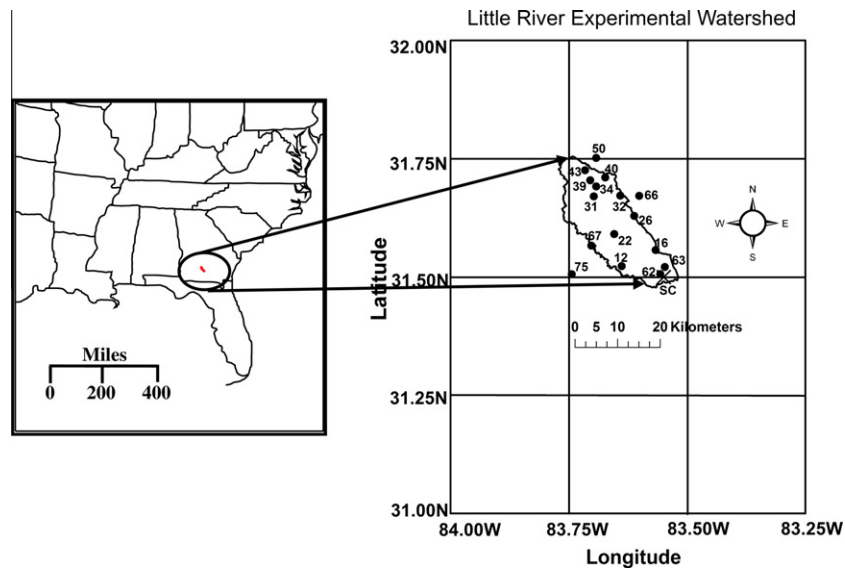


Fig. 1. Little River Experimental Watershed (LREW), Tifton, Georgia, USA [from [5]]. The left figure shows the location of the LREW in the state of Georgia. The simulation domain with satellite coarse grid cells, LREW boundary and 16 in situ soil moisture stations used in this study is shown for reference (right). The location of a single SCAN site is indicated by a cross ('x') mark (right). The simulation domain includes 12 (4×3) coarse-scale satellite grid cells and 7500 (100×75) fine-scale model grid cells. The 16 in situ stations are: 12, 16, 22, 26, 31, 32, 34, 39, 40, 43, 50, 62, 63, 66, 67 and 75.

2.4. The Noah land surface model (LSM)

The Noah LSM (version 2.7.1) gets its lineage from the Oregon State University (OSU) LSM [34]. This model vertical soil profile consists of 4 soil layers with lower boundaries at 10 cm, 40 cm, 100 cm, and 200 cm below the surface. The rooting depth of the model is fixed at 100 cm. It has one snow layer and one canopy layer. The physics of vertical water mass movement between the soil layers is governed by the mass conservation law and the diffusive form of the Richard's law whereas the infiltration is governed by a conceptual parameterization for the sub-grid treatment of precipitation and soil moisture [56]. A detailed description of the Noah model is provided in Chen et al. [6], Chen and Dudhia [7] and Ek et al. [19]. The Noah model is chosen, because its soil moisture forecasts over the LREW for 2003 outperformed those of the HySSiB (Hydrologic improvement of Simplified Simple Biosphere model; [40]) and the CLM (Community Land Model, version 2.0; [10]) land surface models [53].

2.5. The EnKF Module in the Land Information System (LIS)

The Noah model and the assimilation simulations are performed within NASA-GSFC's LIS framework [30]. For a comprehensive discussion of the LIS EnKF module, please refer to Kumar et al. [31]. The three-dimensional EnKF (3-D EnKF; used in this study) algorithm was developed by the Global Modeling and Assimilation Office (GMAO/NASA, [51]) and incorporated within LIS5.0 (after modifying LIS5.0 to enable the GMAO EnKF capability for horizontal error correlations, covariance localization, and distributed filtering; [15]).

3. Experiment setup

Our study time period is from July 1, 2002 through December 31, 2007. Two sets of experiments are designed in this study pertaining to the 1-D and 3-D EnKF algorithms. The basic structure of the two sets of experiments is as follows: The observations from the 16 in situ sites are used as the "truth" in this study and validation of the results is performed with respect to the in situ observa-

tions. The "observed" (and assimilated) soil moisture retrievals are derived from the AMSR-E sensor using the LSMEM system. The Noah simulations are conducted using the North American Land Data Assimilation System (NLDAS) atmospheric forcing data [38]. It should be pointed out here that the NLDAS precipitation is not downscaled to the in situ validation scale, but is interpolated from the $1/8^\circ$ to 1 km^2 for the simulations performed in this study. For land cover, we use the University of Maryland's (UMD) 1-km global land cover product [24]. The sand, silt and clay fraction and soil color data are generated at 1 km^2 resolution from the original Pennsylvania State University-USDA State Soil Geographic Database (STATSGO) $1/120^\circ$ resolution soil maps [37]. The "ensemble open loop" runs are performed by forwarding the ensemble integration without assimilation. Lastly, the "assimilation" integrations are conducted by assimilating the AMSR-E "observations" into the ensemble simulations. Both the "ensemble open loop" and "assimilation" integrations are initialized from the restart files generated by "ensemble spin-up" simulations carried out from October 1, 1997 to June 30, 2002.

3.1. Ensemble generation

The EnKF algorithm approximates the model and forcing error covariances through the propagation of an ensemble of model trajectories. Each ensemble member represents a different realization of model and forcing errors. Based on previous studies [31,43,15], we choose 20 ensemble members for this study.

The ensemble perturbations are applied to both the forcing and state variables. Normally distributed, zero mean, spatially correlated and temporally uncorrelated additive random perturbations are applied hourly to the longwave radiation (LW, standard deviation, $\text{stdv} = 30 \text{ W/m}^2$) and the near-surface air temperature (T_a , $\text{stdv} = 2 \text{ K}$) forcings, as well as to the forecasted soil moisture (sm) and soil layer temperature (st) (see Table 1). Log-normally distributed multiplicative perturbations with mean 1 are applied hourly to the precipitation (P, $\text{stdv} = 2$) and the shortwave radiation (SW, $\text{stdv} = 1.5$). Cross-correlation between the forcing perturbations is also included (LW-SW: -0.3 , SW-P: -0.1 , SW- T_a : 0.3 , LW-P: 0.5 , LW- T_a : 0.6 , P- T_a : -0.1 , assuring both positive definite-

Table 1
Description of the additive state perturbation parameters used in this study.

Variable	Std dev (m ³ /m ³ or K)	Cross correlations							
		<i>sm</i> ₁	<i>sm</i> ₂	<i>sm</i> ₃	<i>sm</i> ₄	<i>st</i> ₁	<i>st</i> ₂	<i>st</i> ₃	<i>st</i> ₄
<i>sm</i> ₁	6.00E−3	1	0.6	0.4	0.2	0	0	0	0
<i>sm</i> ₂	1.10E−4	0.6	1	0.6	0.4	0	0	0	0
<i>sm</i> ₃	6.00E−5	0.4	0.6	1	0.6	0	0	0	0
<i>sm</i> ₄	4.00E−5	0.2	0.4	0.6	1	0	0	0	0
<i>st</i> ₁	1.00E−1	0	0	0	0	1	0	0	0
<i>st</i> ₂	5.00E−2	0	0	0	0	0	1	0	0
<i>st</i> ₃	5.00E−2	0	0	0	0	0	0	1	0
<i>st</i> ₄	5.00E−2	0	0	0	0	0	0	0	1

ness and some balancing of the forcing fields). The cross-correlations between the eight state variable perturbations are shown in Table 1 and adapted from Kumar et al. [32] (see Table 1 in their paper). A spatial correlation length (*L*_{corr}) of 30 km is included in the perturbation fields for all variables and a localization radius of 75 km is applied at the 3-D assimilation update steps. Given the size of the study domain, the choice of a 75 km localization radius effectively implies that information from any observation can propagate to any pixel within the study domain. We discuss in Section 4.1.4 how acceptable perturbations were assured through a trial and error analysis of the innovation distributions.

3.2. 3-D EnKF algorithm

The EnKF is a Monte-Carlo application of the Kalman filter [21] and has been used frequently for land surface data assimilation (e.g. [36,47,41,13,16,33]). We only highlight some aspects that are relevant for our study. The 1-D EnKF application in this study assimilates a priori partitioned observations at the fine scale model grid cells (1D-F1 case of [15]). The 3-D EnKF algorithm downscales the coarse observations within the assimilation scheme and uses multiple coarse observation grid cells (3D-Cm case of [15]), as shown in Fig. 2. The schematic diagram explains the update step of the 3-D EnKF algorithm at any particular time step '*i*' (index dropped for simplicity) and for a particular ensemble member '*j*'. Each ensemble member model forecast (a.k.a. prediction: \hat{x}_j^-) at the fine scale (model space) is mapped to the coarse satellite observations (observation space) to calculate the ensemble member observation prediction ($\hat{y}_j^- = H\hat{x}_j^-$) using an observation/upscaling

operator (*H*). The innovation ($y_j - H\hat{x}_j^-$) is calculated from the perturbed coarse satellite observations (y_j) and the observation predictions ($H\hat{x}_j^-$) at the coarse grid scale (observation space). The innovation is then mapped back to the fine model scale (model space) and applied to the individual fine model grid cells to update each ensemble forecast member (a.k.a. analysis: \hat{x}_j^+) at the fine scale via the diagnosed error cross correlations between the fine scale state variables and the coarse scale observation variables, as expressed in the Kalman gain. Here, y_j vector contains multiple coarse scale satellite observations.

We define a localization radius of 75 km, which is equal to 2.5 times the spatial error correlation length *L*_{corr} = 30 km, which in turn covers at least the dimension of one coarse observation grid cell [15]. Based on the localization radius, we find the multiple coarse satellite observations that affect any single fine scale model forecast and we use the innovations from all those respective coarse grid cells to update the model forecast of that model fine grid cell (downward dotted arrows on the right hand side of the Fig. 2). Conversely, the innovation calculated for a given coarse grid cell contributes to the increments for a number of fine scale model grid cells whose centers are within 75 km from the observed coarse grid cell (shown by the multiple downward solid arrows in the right hand side of the figure).

The update equation for a single ensemble member '*j*' at a single fine scale location '*k*' at one time step can be written as:

$$\hat{x}_j^{k+} = \hat{x}_j^{k-} + K^k[y_j - \hat{y}_j^-] \quad (1)$$

where the Kalman gain is defined as

$$K^k = Cov[\hat{x}_j^{k-}, \hat{y}_j^-][Cov[\hat{y}_j^-, \hat{y}_j^-] + R]^{-1} \quad (2)$$

In Eq. (1), \hat{x}_j^{k-} is the a priori state estimate (or forecast) whereas \hat{x}_j^{k+} is the a posteriori state estimate (or analysis). The term \hat{x}_j^{k-} is a part of the total state vector \hat{x}_j^- and contains the state variables at a single grid cell '*k*'. The term \hat{x}_j^{k-} in Eq. (2) refers to the ensemble of forecasts \hat{x}_j^{k-} (crf. first term on the right hand side of Eq. (1)). The term '*R*' denotes the observation error covariance. In our case, the state vector '*x*' contains eight state variables (4 soil moisture variables (*sm*₁, *sm*₂, *sm*₃ and *sm*₄) and 4 soil temperature variables (*st*₁, *st*₂, *st*₃ and *st*₄) for the 4 soil layers). Given the small standard deviation values for the temperature perturbations (Table 1) and given the small error correlations between the soil moisture and soil temperature variables (not shown), the updates to the temperature variables are negligible. In practice, the state vector therefore consists essentially of the 4 soil moisture variables.

The model state variables are updated using Eq. (1) at the time step when the observations (y_j) are available, after contrasting them to the observation prediction (\hat{y}_j^-). The latter are calculated as the spatial mean of all the 1 km fine scale soil moisture values within any single 25 km coarse grid cell. The first term on the right hand side of Eq. (2) includes the error correlation between the ensemble of forecasted *sm*₁ over different coarse scale areas (\hat{y}_j^- , size depends on the localization radius) and the ensemble of the eight fine scale state variables at a single fine scale model grid cell (\hat{x}_j^{k-}). This correlation information is the basis for updating the eight state variables in response to coarse surface layer soil moisture observations (y_j).

3.3. Forecast bias correction

The Kalman filter is only optimal in the presence of zero mean, temporally uncorrelated forecast and observation errors. As will be shown below, the model forecasted soil moisture regime is different from both the in situ and satellite observations, while the latter two show a very similar regime despite the scale discrepancy. Because of the good agreement between the satellite observations

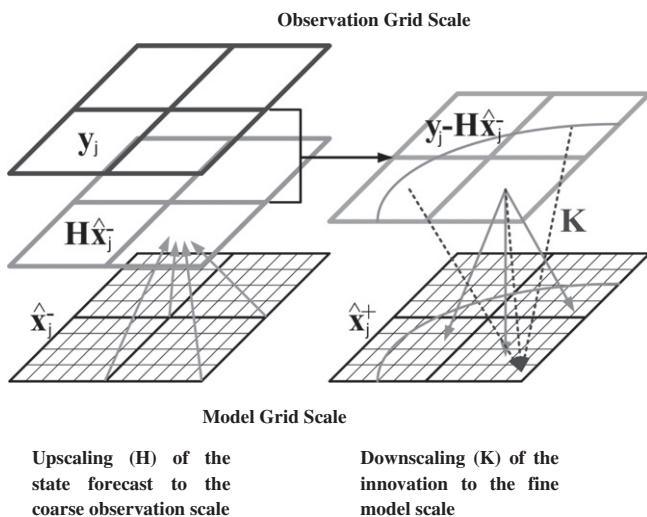


Fig. 2. A schematic diagram of the 3-D EnKF approach illustrated for four coarse-scale pixels, each containing 4×6 fine-scale pixels.

and in situ measurements, we assume no observation bias and attribute the time mean innovation (difference between satellite observation and forecast) value to forecast bias. As a crude estimate of the forecast bias ‘ b_1 ’ in the surface layer soil moisture, the spatial and temporal mean difference between the model forecasts and all available in situ observations during the period July 1, 2002 through December 31, 2007 (prior to assimilation) is computed ($\hat{b} = [b_1, 0, 0, \dots, 0]^T$) and removed from each ensemble member innovation. The ensemble member analysis at a single fine scale grid cell ‘ k ’ is \hat{x}_j^{k+} and fed back into the model as an internal state:

$$\hat{x}_j^{k+} = \hat{x}_j^{k-} + \tilde{K}[y_j - H(\hat{x}_j^{k-} + \hat{b})] \quad (3)$$

After assimilation, this analysis is optionally bias-corrected (\hat{x}_j^{k+} , not fed back into the model) to bring the output to the true (in situ) climatology by adding the constant bias term to the top layer soil moisture as follows:

$$\hat{x}_j^{k+} = \hat{x}_j^{k+} + \hat{b} \quad (4)$$

The analysis output (not fed back into the model) is the ensemble mean of the ensemble members ‘ \hat{x}_j^{k+} ’ in Eq. (4). Tildes refer to quantities which contain biased information. Note that we will also remove this constant bias (single value) from the ensemble open loop to obtain a “bias-corrected ensemble open loop” in the discussion of the results.

The approach is similar to approaches where the observations are rescaled to the model climatology (e.g. CDF matching; [49]), but differs in the fact that here the bias is assigned to the forecast and only the first order (mean) bias is corrected based on the spatial average across the in situ observations. Our method is also very similar to EnBKF_2 in De Lannoy et al. [13], but here (i) a static bias correction is applied and (ii) the bias vector b only contains a non-zero element for the observed state variable of surface soil moisture. The latter implies that here non-observed variables are updated with anomaly increments, and the analysis for the non-observed variables (e.g. deep layer soil moisture) is kept at the model climatology without bias correction.

4. Results and discussion

The results in this study are analyzed in three parts: (1) the performance of the EnKF algorithms for surface soil moisture; (2) the impact of surface soil moisture assimilation on the deeper layer soil moisture and other water cycle variables and (3) the effect of data assimilation on soil moisture state balancing.

4.1. Performance of the EnKF algorithms

The surface soil moisture assimilation results are compared to the in situ soil moisture observations over the 16 in situ locations from 2002 to 2007 (Fig. 1). Both the daily ascending and descending soil moisture data at 1:30 pm and 1:30 am are assimilated whenever the satellite data are available, but we restrict our validation to the assimilation estimates at 1:30 pm, as in our previous studies [53,54]. We emphasize that the in situ data from the entire time period at three-hourly temporal resolution (8 time points per day) are used to calculate the overall coarse-scale forecast bias ($=0.13 \text{ m}^3/\text{m}^3$) prior to the data assimilation.

4.1.1. Temporal evolution of the spatially averaged top layer soil moisture

Fig. 3 shows the spatially averaged (over 16 in situ locations) time series of the top layer soil moisture from the ensemble open loop, 1-D and 3-D EnKF simulations along with the observed LREW in situ soil moisture data for the period 2002–2007. Fig. 3 also

includes the assimilated satellite observations. The satellite observations include values only from a single coarse 25 km^2 satellite grid cell which contains all the 16 in situ locations (Fig. 1). The corresponding in situ total precipitation/irrigation data are shown in Fig. 3 as well. A 10-day moving average is applied to all the time-series data prior to plotting; however all the validation statistics below are calculated using the original daily data at 1:30 pm. All the data products exhibit the dominant seasonal cycle very well and show a similar dynamic response to precipitation events. This time series plot clearly shows the large persistent bias over time for the surface soil moisture in the ensemble open loop case with respect to the in situ data. The 1-D and 3-D EnKF results with bias correction are within close range of the in situ observations. The satellite observations are able to capture the high-frequency soil moisture variations that are observed in the in situ data during the very dry spells; however the satellite estimates differ from the in situ observed values.

Fig. 4 shows the corresponding scatter plots of the daily 1:30 pm soil moisture estimates (satellite, station-averaged ensemble open loop and assimilation results) against the in situ observations. In Fig. 4a, the number of data points is limited (1346 days versus 2010 days in other datasets) by the satellite revisit time (once every 2–3 days). The satellite observations show a small negative bias ($-0.02 \text{ m}^3/\text{m}^3$) over this watershed. The grey highlighted box in Fig. 4a shows that the satellite retrievals overestimate soil moisture for dry conditions and hence supports the findings in the time series plot in Fig. 3. The ensemble open loop integration has a bias of $0.12 \text{ m}^3/\text{m}^3$ and a RMSE of $0.13 \text{ m}^3/\text{m}^3$ with respect to the in situ data. This bias is very systematic except when the soil moisture values are low and the ensemble open loop forecast results become uncorrelated with the in situ observations as shown in the grey box in Fig. 4b. Both the 1-D (Fig. 4c) and 3-D (Fig. 4d) EnKF assimilation analyses show a fine-scale positive bias, which is negligibly higher for the 1-D EnKF case than those of the 3-D EnKF case. Both the EnKF simulations improve the correlation values over those from either the satellite observations or the ensemble open loop, indicating that the assimilation has the capability to improve the soil moisture estimates over the model results or observations alone. Since the precipitation in the NLDAS forcing data has not been downscaled to the in situ observations and since precipitation directly controls the model forecast soil moisture estimates, it is not possible to achieve near perfect time series correlation through any data assimilation filter alone. Nevertheless, the correlation values are high (0.68–0.83) for all the soil moisture datasets with respect to the in situ observations.

Fig. 5 shows the temporal evolution of the spatial RMSE for the ensemble open loop and assimilation results, calculated over the 16 in situ stations. The spatial RMSE values range between 0.08–0.17 m^3/m^3 (open loop without bias correction) and 0.01–0.09 m^3/m^3 (open loop with bias correction and assimilation results), respectively. The RMSEs for the assimilation analyses generally indicate enhanced spatial soil moisture fields over the bias-corrected open loop and most so for the 3-D EnKF results. The RMSE time series for both the EnKF cases show a consistent negative linear trend over the entire study period which is not noticed in the case of the ensemble open loop. Furthermore, the EnKF RMSE time series follows the original soil moisture time series discussed earlier in Fig. 3. The peaks in all the RMSE time series roughly follow the precipitation peaks. This suggests that the finer-scale precipitation details are not included in the simulations and wetter conditions thus produce higher spatial RMSE values.

Fig. 6 shows the temporally averaged spatial maps of the analysis error standard deviation (estimated from the ensemble spread) for the 1-D and 3-D EnKF cases, calculated from daily instantaneous data at 1:30 pm over the entire period of July 1, 2002 to December 31, 2007. The range of the analysis error

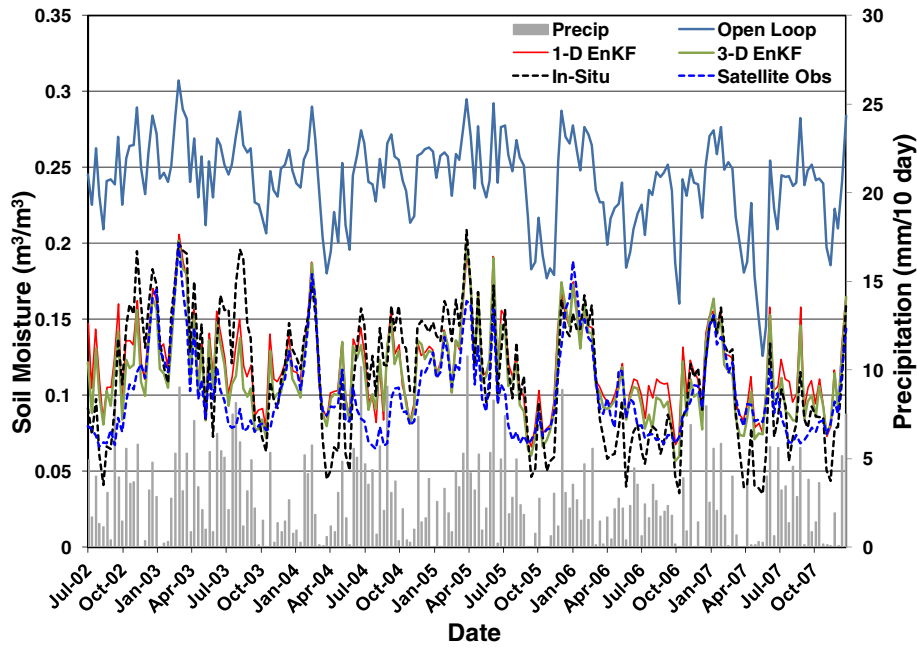


Fig. 3. 10-Day average surface layer soil moisture evolution from Noah (0.1 m layer) ensemble open loop, 1-D and 3-D EnKF simulations; in situ measurements (0.05 m layer) and satellite retrievals (~0.02 m) (left vertical axis). 10-Day average in situ precipitation is shown on the right vertical axis. The model and in situ results are spatially averaged over 16 in situ locations in the Little River Watershed.

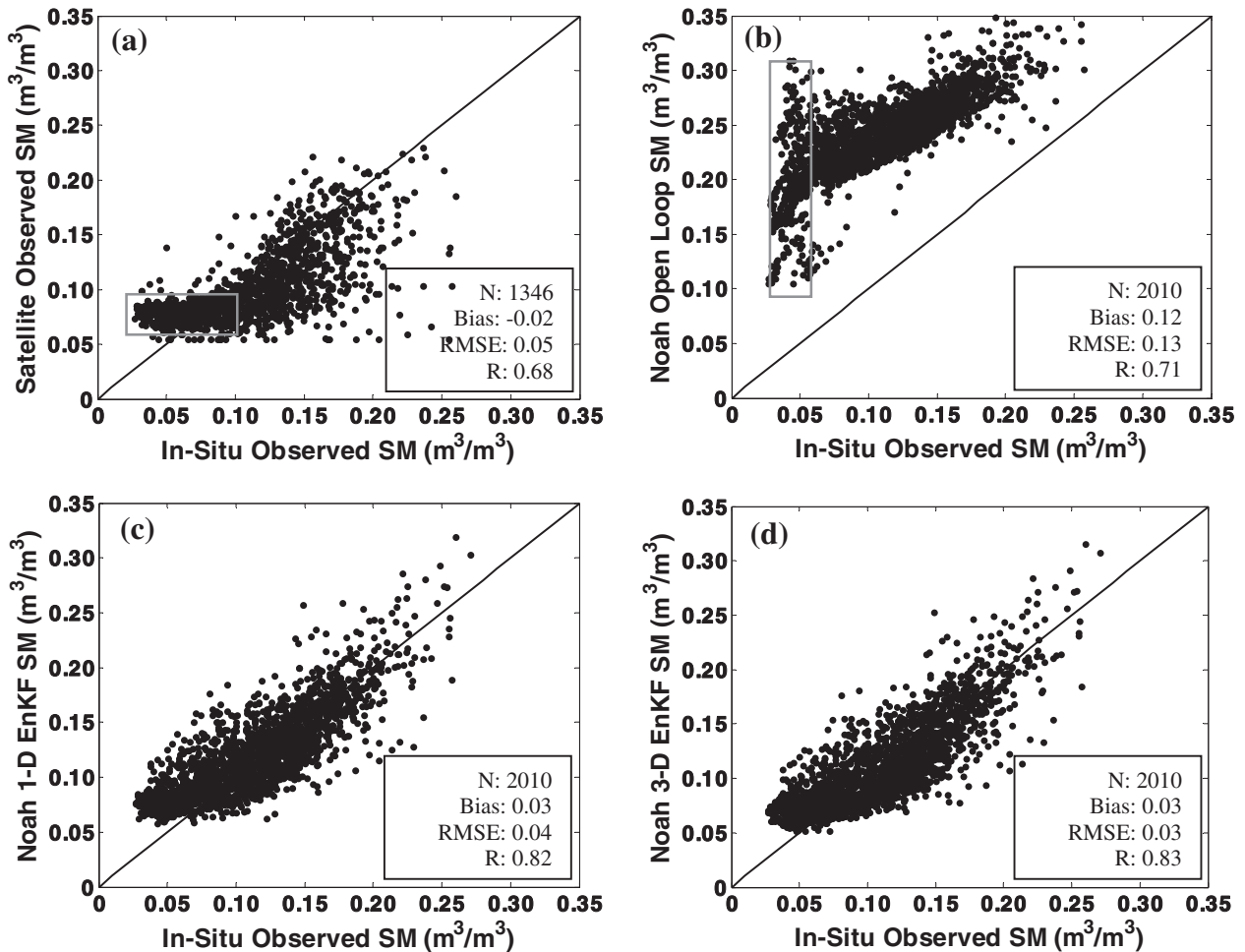


Fig. 4. Daily soil moisture estimates versus average in situ measurements from 16 stations in the Little River Watershed. (a) Satellite retrievals, (b) ensemble open loop, (c) 1-D EnKF and (d) 3-D EnKF.

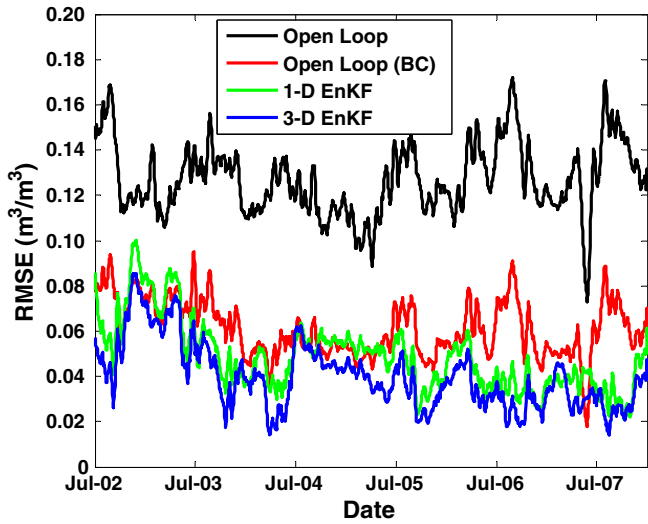


Fig. 5. Time series of the spatial RMSEs (Root Mean Square Errors) in surface soil moisture for the ensemble open loop, bias-corrected open loop (BC), 1-D and 3-D EnKF simulations over 16 in situ stations.

standard deviation is very similar for both the EnKF cases. For the 1-D EnKF the spread does not show any specific spatial pattern whereas for the 3-D EnKF the spread is generally higher in the center of each coarse satellite pixel area and lower in the periphery which can be attributed to the spatial correlation structure and the observations at coarse scale (25 km²) spatial resolution (periphery is being affected by multiple satellite observations from neighboring grids). Additionally, we also notice that the lower spread follows the dendritic structure seen in the map which comes from the soil texture.

4.1.2. Analysis at individual stations

Next, we quantify the time series behavior at all the 16 individual in situ stations and the nearby SCAN (SC) site. Fig. 7 shows a comparison of the skill scores (bias, ubRMSE \equiv sqrt(RMSE² -

- bias²), and correlation coefficient) of the ensemble open loop (both non-bias-corrected and bias-corrected), 1-D and 3-D EnKF soil moisture analyses. The skill scores are calculated from the daily instantaneous (1:30 pm) soil moisture estimates for the period 2002–2007. The station averaged (AVG) values are included in the figures as well. The x-axis in each plot represents the station identification number (see Fig. 1 for station location). Fig. 7a shows large positive bias values in the non-bias-corrected ensemble open loop soil moisture results for all stations. The bias reaches as large as 0.16 m³/m³ for the Station 32. The bias-corrected ensemble open loop results still contain positive bias (around 0.03–0.13 m³/m³). As discussed above we are here applying a first order bias correction (a constant value) is applied by determining the spatially averaged bias at 16 in situ points and temporally averaging over the entire study period. Moreover, the constant bias is calculated based on a comparison of three-hourly model forecast and the in situ observations for the entire period whereas the results are shown from averaging the soil moisture data only at 1:30 pm time instants for the entire period. Therefore this first order constant bias correction does not completely remove the bias. This presence of bias is also supported by the innovation histograms which will be discussed in Section 4.1.4. Both the EnKF results also show positive, but reduced biases over the 16 stations and the SCAN site. The average bias for the ensemble open loop, bias-corrected ensemble open loop, 1-D and 3-D EnKF are 0.13, 0.07, 0.04 and 0.03 m³/m³, respectively.

Fig. 7b shows the time averaged ubRMSE values for all the simulation results for the same 16 stations and the SCAN site. This ‘unbiased’ metric is particularly interesting to measure the assimilation skill, because the success of a Kalman filter is measured by its ability to improve the time series dynamics, rather than the time mean climatology level. The ubRMSE values increase when the validating in situ data and the assimilation analyses (i) differ more in time series variance magnitudes and/or (ii) have a lower time series correlation [20]. The EnKF results produce the lowest ubRMSE values for nearly all stations. The magnitudes of the ubRMSE values for the 3-D EnKF case are comparable to those of the 1-D EnKF case, but always slightly better. The time series correlation is plotted in Fig. 7c and high correlation values are seen for

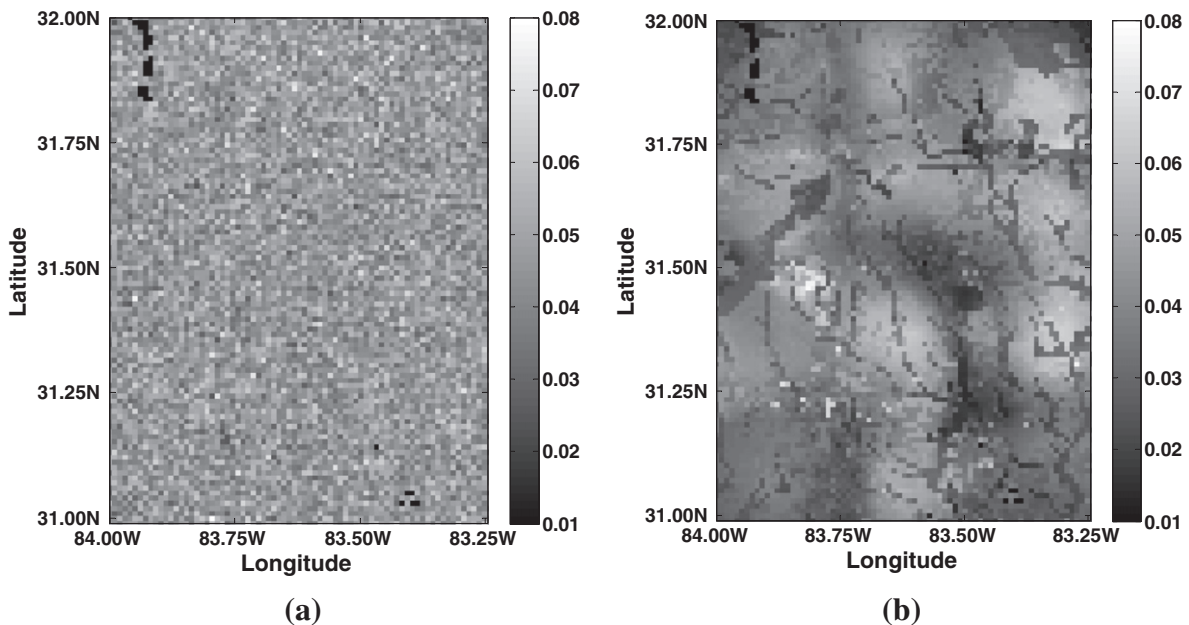


Fig. 6. Temporally averaged spatial maps of the analysis error (ensemble spread) for the (a) 1-D EnKF and (b) 3-D EnKF cases calculated at daily instant AMSR-E ascending time (i.e. 1:30 pm) over the entire period of July 1, 2002 to December 31, 2007. The unit is in m³/m³.

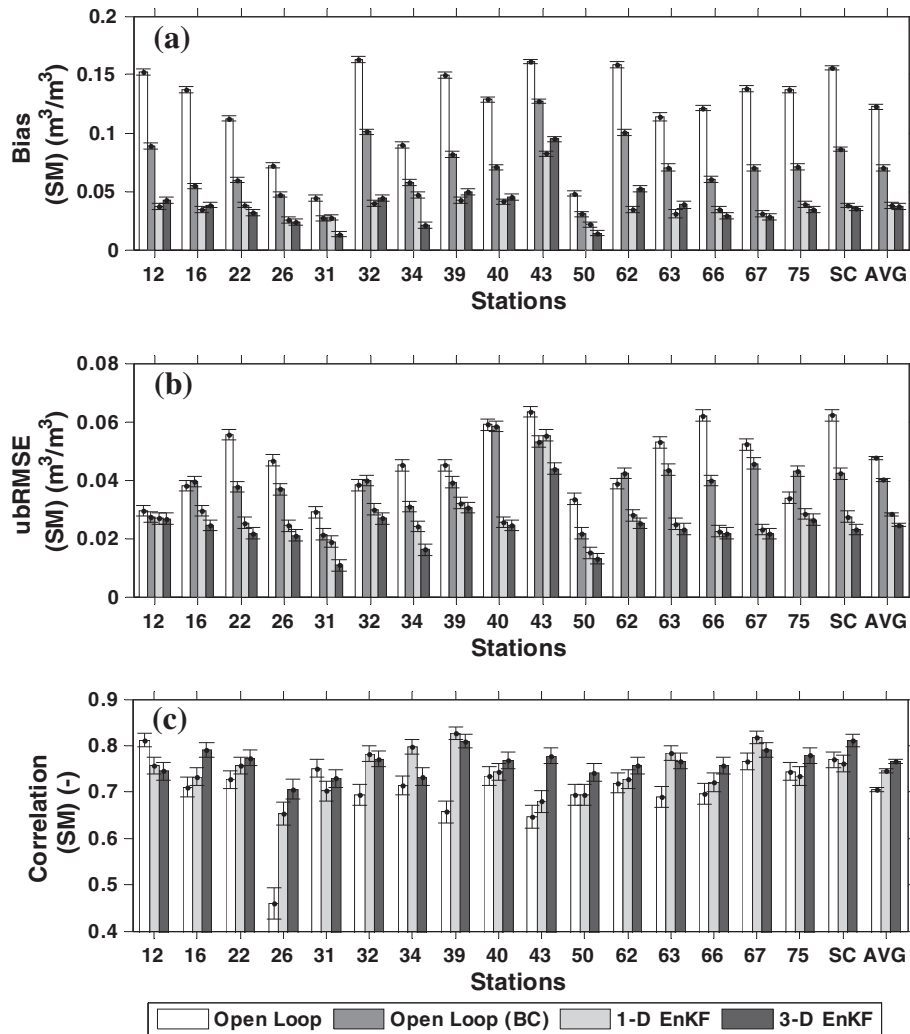


Fig. 7. (a) Mean bias, (b) ubRMSE = $\sqrt{\text{RMSE}^2 - \text{bias}^2}$, and (c) correlation for the ensemble open loop, bias-corrected ensemble open loop (not shown for correlation since the values are same as those of the ensemble open loop), 1-D and 3-D EnKF simulated soil moisture for the 16 LREW in situ stations (identified by their station number) and the SCAN (SC) site. Also shown are the station average metrics (AVG) and 95% confidence intervals.

the ensemble open loop and EnKF soil moisture estimates (between 0.6 and 0.8) when compared to the in situ measured data at all the stations except at Station 26. The station averaged correlation increases from 0.7 (open loop) to 0.74 (1-D) and 0.77 (3-D). The 95% confidence intervals around the estimated bias, ubRMSE and correlation coefficient values indicate that these improvements are statistically significant for the station-average metrics and for many individual stations.

4.1.3. Precipitation-soil moisture spatial coherence maps

The 3-D EnKF filter updates soil moisture based on both vertical and horizontal error correlations, as opposed to the 1-D EnKF filter which considers only vertical error correlations (independently at each fine scale grid cell); thereby the 3-D EnKF filter maintains the spatial coherence pattern of the soil moisture distribution.

To analyze the spatial distribution of soil moisture and precipitation-soil moisture coherence patterns, we choose a time period of wetting and drying dynamics (February 26 to March 1, 2003) over our simulation domain. Fig. 8 shows daily NLDAS precipitation forcing (column 1) and the corresponding daily soil moisture difference images for the 1-D EnKF (column 2) and 3-D EnKF (column 3) results at 1-km spatial resolution. The light blue to dark blue color scale for precipitation shows lower to higher precipitation. The red color in the soil moisture difference maps represents

negative values (dry-out) whereas the blue color represents the positive values (wetting). There is no precipitation on February 27; heavy precipitation of more than 16 mm/day in the central and southern parts and some precipitation of less than 4 mm/day in other parts of the study region on February 28. The dendritic patterns (also in Fig. 6) in all the soil moisture difference maps come from the soil texture maps. Generally, the soil moisture change maps correspond very well to the precipitation events in space and time. The 3-D EnKF maintains a realistic pattern, whereas the 1-D EnKF results fail to preserve the spatial coherence of the atmospheric signature (precipitation) in the soil moisture maps. Also, the coarse grid cell (25 km) boundaries are quite often imprinted in the 1-D EnKF soil moisture maps; especially when the soil moisture values are very low (dry period) and when the soil moisture values are significantly different across the satellite grid cells (not shown). The spatial correlation (calculated over all the 16 sites, not shown) yields similar high values for both the ensemble open loop and the 3-D EnKF results and lower values for the 1-D EnKF results.

4.1.4. Filter diagnostics

The performance of the EnKF strongly depends on the choice of the assimilation parameters [51]. For a consistently operating filter, the normalized innovations sequence should be white, i.e. it

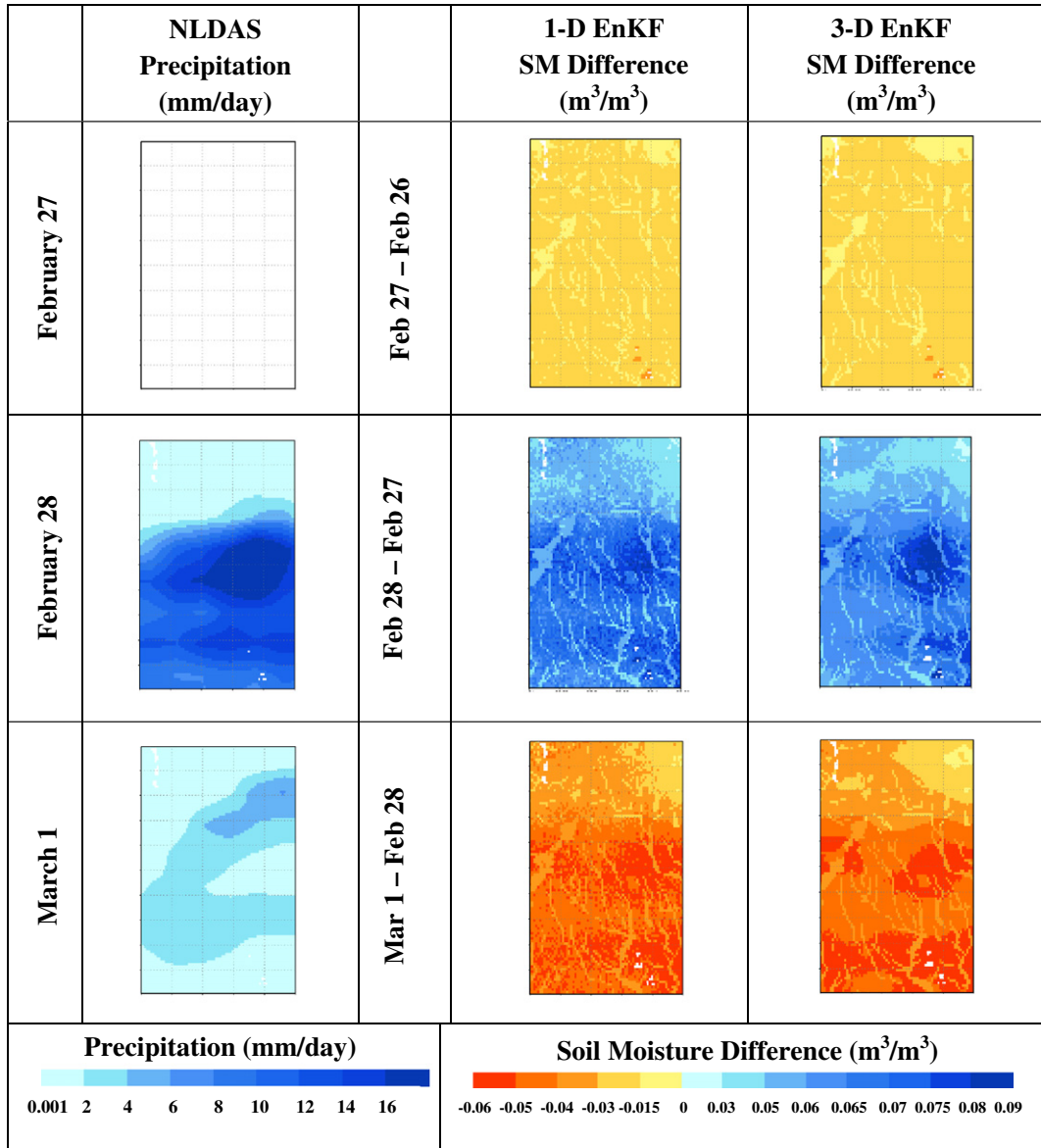


Fig. 8. Spatial precipitation and soil moisture difference maps during a wet event for the period February 26–March 1, 2003.

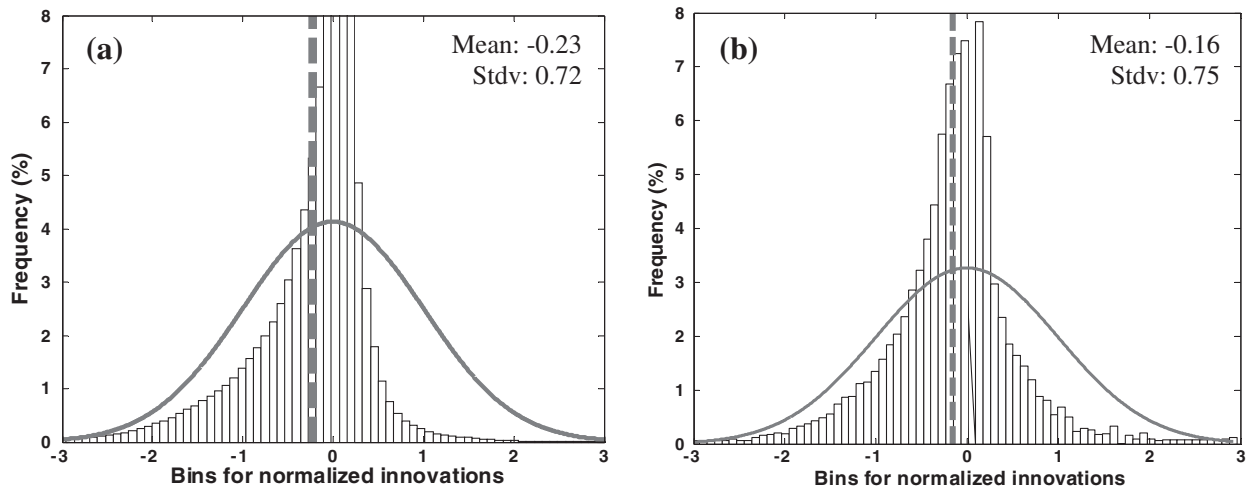


Fig. 9. Histograms of the normalized innovations (dimensionless) for (a) 1-D EnKF and (b) 3-D EnKF.

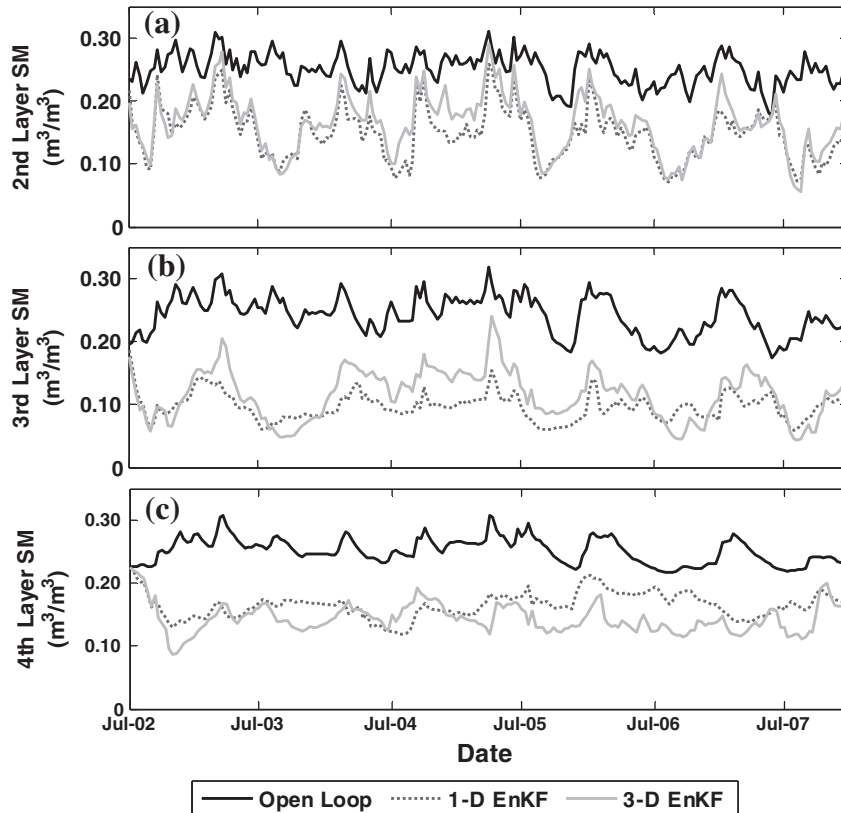


Fig. 10. Daily time series of the station averaged moisture from (a) Layer 2 (0.1–0.4 m), (b) Layer 3 (0.4–1 m) and (c) Layer 4 (1–2 m) for all the ensemble open loop and EnKF simulations.

should exhibit no correlation in time, and it should follow the standard Gaussian (Normal) distribution with zero mean and unit variance ($N(0,1)$). The normalized ensemble mean innovations are calculated by normalizing each innovation value corresponding to each individual coarse scale observation (m) by the square root of its filter estimated uncertainty ($\sqrt{\{Cov[\hat{y}^m - \hat{y}^m] + [R]_{m,m}\}}$). We manually calibrated the state forecast error and observation error covariance matrices (P and R , respectively) through trial and error with several combinations of standard deviations for the state/forcing perturbations and observation error. We use $0.05 \text{ m}^3/\text{m}^3$ as the observation error standard deviation in this study. Fig. 9 shows the distribution of the normalized ensemble mean innovation values for both the 1-D and 3-D EnKF algorithms with the final set of “ P ” and “ R ” that have been used in this study. Note that the 1-D EnKF innovations in Fig. 9a are at the fine scale (1 km grid; total 7500 grid cells over the simulation domain) whereas the 3-D EnKF innovations in Fig. 9b are at the coarse scale (25 km grid; total 12 grid cells over the simulation domain). The dotted vertical line in Fig. 9a and b show the mean of the innovations. The standard normal distribution is also shown for reference.

There is a negative bias in the innovations (satellite observations minus ensemble forecast) averaged over the entire domain, even if they are first-order bias-corrected. This is because the first order bias correction (a constant value) is determined based on 16 in situ points (site averaged) and not sufficient to remove the bias entirely for the whole area. It is possible to remove the negative innovations bias by increasing “ P ” and forcing the model towards the observation climatology, but the histogram spread then suggests the sub-optimality of the filter for a reasonable range of “ R ”. Overall, the innovation histogram distributions for both the 1-D and 3-D EnKF cases produce similar statistics; though the

mean of the distribution is closer to zero for the 3-D EnKF case. The standard deviation values of both the innovation histogram distributions are reasonably close to 1, indicating that the filter parameters are acceptable.

4.2. Impact of surface layer soil moisture assimilation on other variables

In this section, we verify the impact of surface layer soil moisture assimilation on the deeper layer soil moisture and other water cycle variables by comparing EnKF simulation results with those of the ensemble open loop. We will focus on a qualitative analysis of the model simulated results here, except for the deeper layer soil moisture, for which in situ observations allow a quantitative evaluation.

4.2.1. Deep layer soil moisture

Fig. 10 shows the time series for the moisture in the model 2nd layer (Fig. 10a), 3rd layer (Fig. 10b) and 4th layer (Fig. 10c) for the ensemble open loop, 1-D and 3-D EnKF simulations averaged over the 16 stations. A 10-day moving average is applied to the time series for visual clarity prior to plotting the time series. Bias is likely present in the model predicted deeper layer soil moisture values, but we do not know the bias error correlations between different layers and hence we do not bias-correct the deeper layers. However, the deeper layer soil moisture variables respond to the surface layer soil moisture anomaly assimilation through the state error cross-correlation matrix and eventually through the model propagation. Fig. 10 shows that the deeper layer soil moisture variables immediately respond to the surface soil moisture assimilation and the deeper layer soil moisture values decrease within a few assimilation steps for the 1-D and 3-D EnKF cases, introducing

Table 2

Correlation between the model simulated 2nd layer (0.1–0.4 m) soil moisture and the in situ observed soil moisture averaged from 0.2 and 0.3 m measurement depths at 16 in situ locations. Also shown are the station average (AVG) correlation values.

Sites	Ensemble open loop	1-D EnKF	3-D EnKF
RG12	0.76	0.79	0.77
RG16	0.63	0.69	0.70
RG22	0.69	0.70	0.79
RG26	0.54	0.61	0.63
RG31	0.79	0.71	0.72
RG32	0.65	0.68	0.69
RG34	0.65	0.63	0.69
RG39	0.67	0.63	0.66
RG40	0.71	0.65	0.65
RG43	0.66	0.61	0.70
RG50	0.70	0.66	0.64
RG62	0.60	0.61	0.70
RG63	0.59	0.62	0.60
RG66	0.53	0.53	0.53
RG67	0.68	0.64	0.68
RG75	0.66	0.70	0.66
AVG	0.66	0.65	0.68

a difference between the ensemble open loop and 3-D EnKF deep layer soil moisture values as high as $0.10 \text{ m}^3/\text{m}^3$ (2nd layer) to $0.15 \text{ m}^3/\text{m}^3$ (4th layer). It can be noticed that there is a shift in climatology in deep layer soil moisture between ensemble open loop and EnKF cases. The seasonality is dampened in the deeper layers, because the deeper layers are less likely to respond quickly to the atmospheric signals. Additionally, the thicker subsurface layers hold more water.

The in situ soil moisture observation depths (0.05 m, 0.2 m and 0.3 m) differ from the Noah model layer depths (1st layer: 0–0.1 m, 2nd: 0.1–0.4 m, 3rd: 0.4–1 m and 4th: 1–2 m). Information from the nearby SCAN site reveals that the soil texture changes considerably with depth (sand percent from 88% to 70% and clay percent from 2% to 20% from the surface to 2 m soil depth). Yet, in our simulations, we use only a single soil texture map for all the soil depths. It is difficult to match the in situ soil moisture data to the depths equivalent to the Noah model soil layer since the water retention curve is different for different soil texture types. Therefore, we exclude possible observation–forecast bias from our validation by focusing on the time series correlation between the model simulated 2nd layer (0.1–0.4 m) soil moisture and the in situ observed soil moisture averaged from 0.2 and 0.3 m measurement depths. Table 2 shows the time series correlation results for the 2nd layer SM at individual sites. The station-average (AVG) time series correlation values are 0.66, 0.65 and 0.68 for the ensemble open loop, 1-D and 3-D EnKF, respectively. The correlation is quite good at all the in situ stations even though the absolute soil moisture values are not directly comparable.

4.2.2. Water cycle variables

In this section, we discuss the response of the surface and subsurface runoff and evaporation to the surface soil moisture assimilation. These variables are diagnostic variables in the assimilation and we let these variables respond to the updated soil moisture and temperature through model physics and dynamics.

The percentage change in water cycle variables introduced as a result of EnKF assimilation averaged over the entire study period is presented in Table 3. Out of all the water cycle variables, the subsurface runoff shows the largest relative changes due to lower soil moisture climatology in the deeper layers and the surface runoff shows the smallest relative changes. The surface runoff in Noah model is least affected, because the precipitation water is removed through surface runoff before the rest of the precipitation water enters the soil. It can be also noticed that there is a considerable difference in the changes in subsurface runoff results between

Table 3

Percentage change in water cycle variables introduced by surface soil moisture assimilation averaged over the entire study period (July 2002–December 2007).

Variable	1-D EnKF (%)	3-D EnKF (%)
Surface runoff	12	11
Subsurface runoff	135	78
Evaporation	30	29

the 1-D and 3-D EnKF simulations which is not found in other water cycle variables. The time series plots for all the water cycle variables from ensemble open loop and EnKF cases behave similarly though the values are different (figures not shown here).

4.3. Effect of the 3-D EnKF filter on state balancing

Model initialization plays a major role in controlling the evolution of the model state variables and fluxes through time (e.g. [39,69]) and data assimilation has been offered as a tool to provide balanced initial conditions without a need for long spin-ups. We verify the effect of the 3-D EnKF algorithm on the state balancing by performing several “cold start” simulations with different yet spatially homogenous initial soil moisture conditions and comparing those results with the “warm start” simulation (that is, a simulation with a restart file obtained through spin-up). A description of all the simulations is provided in Table 4. We choose three different homogeneous initial soil moisture conditions at all grid cells across the domain for this study: (a) dry start ($0.02 \text{ m}^3/\text{m}^3$), (b) intermediate start ($0.25 \text{ m}^3/\text{m}^3$) and (c) wet start ($0.48 \text{ m}^3/\text{m}^3$). An additional experiment is also performed with warm start initialization condition after simulating 5 year of model spin-up. The experiments are performed for the ensemble open loop without assimilation and for the 3-D EnKF assimilation integrations.

4.3.1. Sensitivity to the model spin-up

Fig. 11 shows 1 year of soil moisture time series results for the ensemble open loop and 3-D EnKF cases with different initialization conditions for all 4 soil layers. All the results are spatially averaged over 16 in situ locations. The dotted curves represent ensemble open loop and the solid curves represent the 3-D EnKF cases, respectively. The same color for both the dotted (ensemble open loop) and solid curves (3-D EnKF) corresponds to the same model initialization condition. The cold start simulations (open loop and 3-D EnKF cases) are compared to the respective target warm (spun up) start simulations and the time after which the cold started simulation converges to the warm started simulation is recorded in Table 5 as the recovery time for all the four soil moisture layers. All these experiments start on July 1, 2002 (summer start). Table 5 also shows the results for the simulations that start on January 1, 2003 (winter start) which are discussed later in Section 4.3.2.

Table 4

Description of the model experiments used to study the effect of the 3-D EnKF algorithm on state balancing. Experiments are started either in the summer ($X = S$, July 1, 2002) or in the winter ($X = W$, January 1, 2003).

Experiment name	Description
OL-LX	Open loop simulation with SM I.C. = $0.02 \text{ m}^3/\text{m}^3$
OL-MX	Open loop simulation with SM I.C. = $0.25 \text{ m}^3/\text{m}^3$
OL-HX	Open loop simulation with SM I.C. = $0.48 \text{ m}^3/\text{m}^3$
OL-RX	Open loop simulation with SM I.C. from spin-up simulations
3D-LX	3-D EnKF simulation with SM I.C. = $0.02 \text{ m}^3/\text{m}^3$
3D-MX	3-D EnKF simulation with SM I.C. = $0.25 \text{ m}^3/\text{m}^3$
3D-HX	3-D EnKF simulation with SM I.C. = $0.48 \text{ m}^3/\text{m}^3$
3D-RX	3-D EnKF simulation with SM I.C. from spin-up simulations

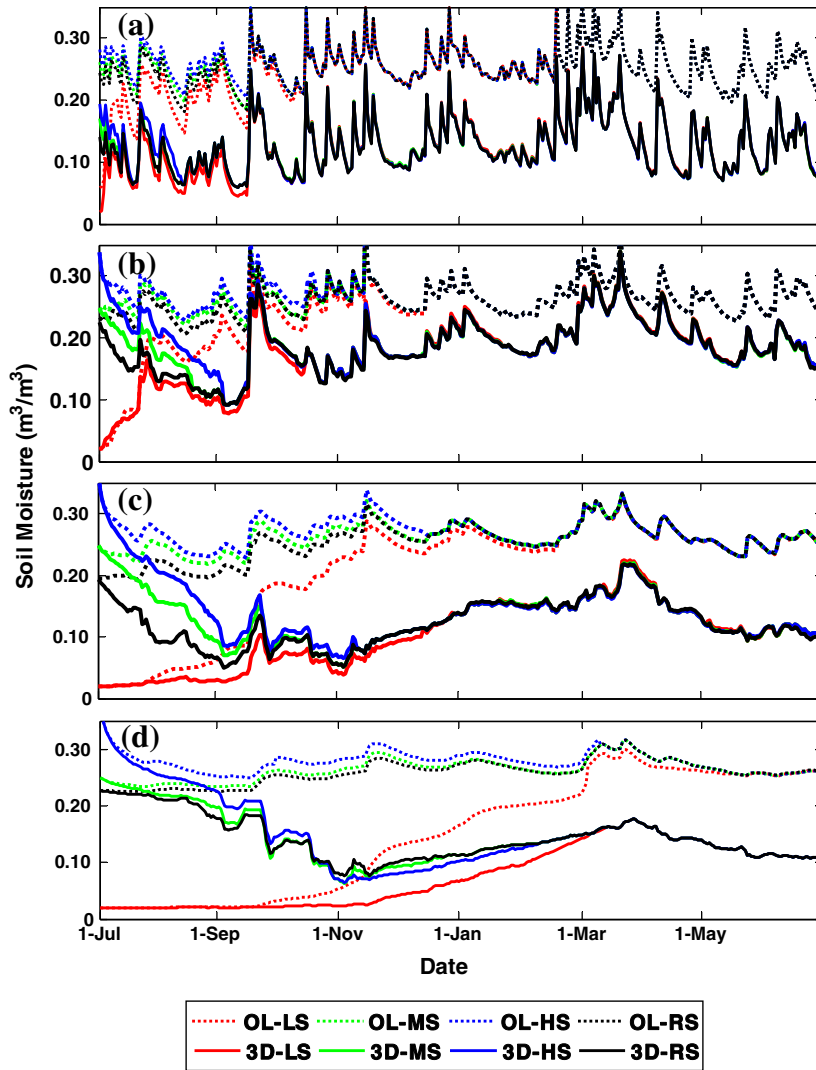


Fig. 11. Sensitivity results of the 3-D EnKF algorithm to the model initialization conditions and model spin-up for the (a) Layer 1, (b) Layer 2, (c) Layer 3 and (d) Layer 4. The simulations start in the summer (July 1, 2002). The results are spatially averaged over all the 16 in situ locations. Refer to Table 4 for the description of the experiments in the legend.

Table 5
The recovery time for the cold start runs with respect to the warm start run for all the four soil moisture layers for the 3-D EnKF algorithm. The table shows results for all the experiments starting in the summer (July 1, 2002) and in the winter (January 1, 2003).

Summer start			Winter start			
Approximate recovery time (in days) with respect to the OL-RS results			Approximate recovery time (in days) with respect to the OL-RW results			
OL-LS	OL-MS	OL-HS	OL-LW	OL-MW	OL-HW	
<i>Open loop simulations</i>						
Layer 1	103	61	78	61	31	38
Layer 2	163	90	115	93	45	52
Layer 3	229	132	167	303	192	250
Layer 4	322	185	253	365+	245	305
Summer start			Winter start			
Approximate recovery time (in days) with respect to the 3D-RS results			Approximate recovery time (in days) with respect to the 3D-RW results			
3D-LS	3D-MS	3D-HS	3D-LW	3D-MW	3D-HW	
<i>3-D EnKF simulations</i>						
Layer 1	74	24	42	13	33	33
Layer 2	105	48	64	80	45	55
Layer 3	166	102	134	250	166	200
Layer 4	257	164	217	320	203	230

Fig. 11a shows the results for the model top layer (0–0.1 m). The soil moisture climatologies of ensemble open loop and the 3-D EnKF cases are different because of the first order bias correction

to the 3-D EnKF results (only in the surface layer). However, the responses to the initialization conditions are very similar for both the ensemble open loop and the 3-D EnKF cases. The soil moisture val-

ues for the OL-HS and 3D-HS reach saturation after a few rain events and thereafter moisture is removed from the top layer through drainage. The soil moisture then decreases quickly and converges with the soil moisture time series from the OL-RS and 3D-RS experiments after 78 and 42 days, respectively. Similarly, the OL-MS and 3D-MS simulations take 61 and 24 days respectively to converge with OL-RS and 3D-RS simulations. In contrast to the wet simulations, the soil moisture values for the OL-LS and 3D-LS simulations increase gradually with subsequent precipitation events and therefore take more time (103 and 74 days) to recover.

Fig. 11b–d shows soil moisture for deeper layers. Note that no bias-correction is applied to the EnKF analyses in the deeper layers. The impact of observations on the sub-surface model layers depends on the strength of the vertical coupling of soil moisture between the different layers (decreasing strength with depth, [31]). Table 5 shows that it could take as long as a year or more for the cold-start runs to catch up with the restart run in deeper layers. Data assimilation is able to reduce the recovery time by 1–3 months by accelerating the state balancing.

This study also reveals that the performance of the EnKF algorithm can be affected by the model state initialization guess and that data assimilation does not eliminate the need for spin-up in land surface modeling. As indicated earlier, the wetter simulations (3D-HS and 3D-MS) converge with the 3D-RS simulation much before the 3D-LS simulation does. It implies that the Noah model is able to drain the wet soil layers faster than recharging the dryer soil layers and hence a wetter initialization condition is more likely a better cold start in case of the Noah model if the model spin-up is not possible.

4.3.2. Sensitivity to the model simulation start season

We perform an entire set of simulations (total 8) as above, but we start the simulations on January 1, 2003 as opposed to July 1, 2002 in the previous cases. The objective here is to verify whether the model simulation start season (summer (July 1, 2002) versus winter (January 1, 2003)) has any impact on the results discussed above in Section 4.3.1. The results are shown in Table 5.

In the winter start cases, the simulation results for the top two layers converge faster than those which start in the summer. This may be related to extreme atmospheric conditions that affect the ratio of evapotranspiration (ET) to precipitation. There is a long dry period (around 45 days; January 1 to February 15, 2003) at the beginning of these winter simulations which is followed by a long, continuous and heavy precipitation period (around 2 months; February 15 to April 15, 2003). These continuous extreme dry and wet conditions help the simulations for the top two layers (1st and 2nd layers) converge faster. This is the case for both the ensemble open loop and 3-D EnKF simulations. However, the impact of these continuous extreme dry and wet conditions is not sufficient to help the simulations for the two deepest layers (3rd and 4th layers) converge. The deeper layers do not freeze and they do not feel the direct impact of ET unlike surface layers. These deep layer simulations do not converge past these extreme dry and wet conditions, and take longer to converge as compared to those which start in the summer. Regardless of the influence of the atmospheric conditions, the beneficial impact of data assimilation on spin-up is again clear in the reduced convergence time.

5. Conclusions

This study describes the assimilation of coarse scale (25 km) satellite (AMSR-E) observations of surface soil moisture into a finer scale (1 km) land surface model (Noah) using the 1-D and 3-D EnKF schemes in LIS. The coarse satellite soil moisture observations are

retrieved from the AMSR-E passive microwave 10.65H GHz channel using the LSMEM microwave emission model. A priori first order forecast bias correction is performed for surface soil moisture within the assimilation procedure. We perform a priori observation partitioning of the satellite observations into 1 km grid cells for the 1-D EnKF algorithm and update the model forecast at each fine model grid without affecting any neighboring grid cells. On the other hand, we avoid the a priori observation partitioning when using the 3-D EnKF algorithm and perform the downscaling within the filter algorithm. Multiple coarse observation grid cells are assimilated simultaneously to update fine-scale model grid cells, based on a forecast error correlation length of 30 km and a covariance localization length of 75 km. All the simulation results are validated against the in situ soil moisture observations over the Little River Experimental Watershed, Georgia, USA.

Both the EnKF algorithms produce fine-scale results that are closer to the in situ data than either the model open loop or the satellite observations alone. The 3-D EnKF slightly outperforms the 1-D EnKF and better preserves realistic spatial patterns because of the colored spatial error correlations and the corresponding impact of multiple coarse observation grid cells. The ubRMSE is used as a 'bias-free' measure to validate the soil moisture skill in terms of relative temporal variations. The lowest ubRMSE values at most of the sites are found for the 3-D EnKF case. The mean correlation over all the sites increases from 0.7 for the open loop to 0.74 for the 1-D EnKF and 0.77 for the 3-D EnKF case. The improvements are significant at 95% confidence intervals.

The impact of the surface soil moisture assimilation is reflected in the deep layer soil moisture and other water cycle variables which are directly or indirectly controlled by the surface layer soil moisture. The correlation between the in situ observations and the model simulations at individual sites for the 2nd layer indicates an improvement of the 3-D EnKF results over the ensemble open loop case (Table 2). We assess the impact of the surface soil moisture assimilation on the water cycle variables (sub-surface runoff and evaporation). The change in the subsurface runoff caused by the surface moisture assimilation is found to be much larger than the changes in other water cycle variables. The impact of the surface layer soil moisture assimilation on the heat fluxes is negligible and not discussed here.

Finally, sensitivity studies are performed to verify the robustness of the data assimilation algorithm to the choice of the initial state conditions. It is clearly shown how the EnKF speeds up the state balancing (relative to the open loop integration), regardless of the chosen initial state values. Soil moisture in the top layer quickly reaches a stable moisture range, even when the model is initialized with far too wet or dry values, because the top layer soil moisture is directly constrained by the observations. However, the moisture content in the subsurface layers has a much longer recovery time, because the effect of the surface data assimilation reduces in deeper layers. The magnitude of the recovery time depends on the difference between the initial wetness and the mean soil moisture condition of the region, on the season (atmospheric conditions) and on the vertical coupling strength among the soil layers in a land surface model. Furthermore, the sensitivity results might depend on the frequency of assimilation. Perhaps most importantly, it is shown that the land surface model should be spun up before data assimilation, that is, data assimilation does not generally obviate the need for model spin-up.

The results in this study show that both the 1-D and 3-D EnKF algorithms improve the soil moisture results, but only the 3-D EnKF manages to preserve a realistic spatial pattern. At the same time, however, there is a tradeoff for the computational cost since 3-D EnKF simulations are computationally more expensive in the generation of colored perturbation fields for all state variables and forcings. Hence, the application of the 3-D EnKF algorithm

operationally over large spatial domain still remains a challenge regardless of its better performance.

Acknowledgements

We are thankful to the LIS software team and particularly to Sujay Kumar. We are grateful to Tom Jackson and Mike Cosh in USDA-ARS Laboratory at Beltsville, Maryland, and David Bosch in USDA-ARS Southeast Research Laboratory at Tifton, Georgia for providing all the in situ measurement data sets used in this study. This research work is supported by the NASA Grant NNX08AG90G. Gabriëlle De Lannoy was a postdoctoral researcher of the Research Foundation Flanders (FWO).

References

- [1] Andreadis K, Lettenmaier D. Assimilating remotely sensed snow observations into a macroscale hydrology model. *Adv Water Resour* 2006;29:872–86.
- [2] Bosch DD, Sheridan JM, Davis FM. Rainfall characteristics and spatial correlation for the Georgia Coastal Plain. *Trans Am Soc Agric Eng* 1999;42(6):1637–44.
- [3] Bosch DD, Lakshmi V, Jackson TJ, Choi M, Jacobs JM. Large scale measurements of soil moisture for validation of remotely sensed data: Georgia soil moisture experiment of 2003. *J Hydrol* 2006;323:120–37.
- [4] Bosch DD, Sheridan JM, Marshall LK. Precipitation, soil moisture, and climate database, Little River Experimental Watershed, Georgia, United States. *Water Resour Res* 2007;43:W09472. <http://dx.doi.org/10.1029/2006WR005834>.
- [5] Cashion J, Lakshmi V, Bosch D, Jackson TJ. Microwave remote sensing of soil moisture: evaluation of the TRMM microwave imager (TMI) satellite for the Little River Watershed, Tifton, Georgia. *J Hydrol* 2005;307(1–4):242–53.
- [6] Chen F, Mitchell K, Schaake J, Xue Y, Pan H-L, Koren V, et al. Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *J Geophys Res* 1996;101(D3):7251–66.
- [7] Chen F, Dudhia J. Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 Modeling system. Part II: Preliminary model validation. *Mon Weather Rev* 2001;129:587–604.
- [8] Clark MP, Slater AG, Barrett AP, Hay LE, McCabe GJ, Rajagopalan B, et al. Assimilation of snow covered area information into hydrologic and land-surface models. *Adv Water Resour* 2006;29:1209–21.
- [9] Crow WT, Van Loon E. Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed soil moisture. *J Hydrometeorol* 2006;7:421–32.
- [10] Dai Y, Zeng X, Dickinson RE, Baker I, Bonan GB, Bosilovich MG, et al. The common land model. *Bull Am Meteorol Soc* 2003;84:1013–23.
- [11] Dee DP, Todling R. Data assimilation in the presence of forecast bias: the GEOS moisture analysis. *Mon Weather Rev* 2000;128:3268–82.
- [12] Dee DP. Bias and data assimilation. *Q J R Meteorol Soc* 2005;131:3323–43.
- [13] De Lannoy GJM, Reichle RH, Houser PR, Pauwels VRN, Verhoest NEC. Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. *Water Resour Res* 2007;43:W09410. <http://dx.doi.org/10.1029/2006WR00544>.
- [14] De Lannoy GJM, Houser PR, Verhoest NEC, Pauwels VRN. Adaptive soil moisture profile filtering for horizontal information propagation in the independent column-based CLM2.0. *J Hydrometeorol* 2009;10(3):766–79.
- [15] De Lannoy GJM, Reichle RH, Houser PR, Arsenault KR, Verhoest NEC, Pauwels VRN. Satellite-scale snow water equivalent assimilation into a high-resolution land surface model. *J Hydrometeorol* 2010;11(2):352–69. <http://dx.doi.org/10.1175/2009JHM1192.1>.
- [16] De Lannoy GJM, Reichle RH, Houser PR, Arsenault K, Verhoest NEC, Pauwels VRN. Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado. *Water Resour Res* 2012;48:W01522. <http://dx.doi.org/10.1029/2011WR010588>.
- [17] Delworth TL, Manabe S. The influence of potential evaporation on the variabilities of simulated soil wetness and climate. *J Climate* 1988;1:523–47.
- [18] Dirmeyer PA. Problems in initializing soil wetness. *Bull Am Meteorol Soc* 1995;76:2234–40.
- [19] Ek MB, Mitchell KE, Lin Y, Rogers E, Grunmann P, Koren V, et al. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational Mesoscale Eta Model. *J Geophys Res* 2003;108. <http://dx.doi.org/10.1029/2002JD003296>.
- [20] Entekhabi D, Reichle RH, Koster RD, Crow WT. Performance metrics for soil moisture retrievals and application requirements. *J Hydrometeorol* 2010;11:832–40.
- [21] Evensen G. Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J Geophys Res* 1994;99(C5):10143–62.
- [22] Gao H, Wood EF, Drusch M, Crow WT, Jackson TJ. Using a microwave emission model to estimate soil moisture from ESTAR observations during SGP99. *J Hydrometeorol* 2004;5:49–63.
- [23] Gao H, Wood EF, Jackson TJ, Drusch M, Bindlish R. Using TRMM/TMI to retrieve surface soil moisture over the southern United States from 1998 to 2002. *J Hydrometeorol* 2006;7(1):23–38.
- [24] Hansen MC, Defries RS, Townshend JRG, Sohlberg R. Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int J Remote Sens* 2000;21:1331–64.
- [25] Huang C, Li X, Lu L. Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter. *Remote Sens Environ* 2008;112(4):1320–36. <http://dx.doi.org/10.1016/j.rse.2007.03.028>.
- [26] Hubbard RK, Berdanier CR, Perkins HF, Leonard RA. Characteristics of selected upland soils of the Georgia coastal plain. USDA Agricultural Research Service, 72, ARS-37; 1985.
- [27] Jackson TJ, Bosch DD, Cosh MH. Little River Experimental Watershed soil moisture network data for the AMSR-E calibration/validation program (digital media); 2007.
- [28] Jackson TJ, Cosh MH, Zhan X, Bosch DD, Seyfried MS, Starks PJ, et al. Validation of AMSR-E soil moisture products using watershed networks. In: Proceedings of the international geoscience and remote sensing symposium, July 31–August 4, 2006, Denver, CO; 2006. p. 432–5.
- [29] Kerr YH, Njoku EG. A semiempirical model for interpreting microwave emission from semiarid land surfaces as seen from space. *IEEE Trans Geosci Remote Sens* 1990;28:384–93.
- [30] Kumar SV, Peters-Lidard CD, Tian Y, Geiger J, Houser PR, Olden S, et al. LIS—an interoperable framework for high resolution land surface modeling. *Environ Model Softw* 2006;21:1402–15.
- [31] Kumar SV, Reichle RH, Peters-Lidard CD, Koster RD, Zhan X, Crow WT, et al. A land surface data assimilation framework using the Land Information System: description and application. *Adv Water Resour* 2008;31:1419–32.
- [32] Kumar SV, Reichle RH, Koster RD, Crow WT, Peters-Lidard CD. Role of subsurface physics in the assimilation of surface soil moisture observations. *J Hydrometeorol* 2009;10:1534–47.
- [33] Liu Q, Reichle RH, Bindlish R, Cosh MH, Crow WT, de Jeu R, et al. The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system. *J Hydrometeorol* 2011;12:750–65. <http://dx.doi.org/10.1175/JHM-D-10-05000>.
- [34] Mahrt L, Pan H-L. A two-layer model of soil hydrology. *Boundary-Layer Meteorol* 1984;29:1–20.
- [35] Mahfouf J-F. Assimilation of satellite-derived soil moisture from ASCAT in a limited-area NWP model. *Q J R Meteorol Soc* 2010;136(648):784–98. <http://dx.doi.org/10.1002/qj.602>.
- [36] Margulis SA, McLaughlin D, Entekhabi D, Dunne S. Land data assimilation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. *Water Resour Res* 2002;38(1299). <http://dx.doi.org/10.1029/2001WR001114>.
- [37] Miller DA, White RA. A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. *Earth Interact* 1998;2(2):1–26.
- [38] Mitchell KE, Lohmann D, Houser PR, Wood EF, Schaake JC, Robock A, et al. The multi-institution North American Land Data Assimilation System (NLDAS): utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J Geophys Res* 2004;109:D07S90. <http://dx.doi.org/10.1029/2003JD003823>.
- [39] Mo KC, Schemm JE, Kim H, Higgins WR. Influence of initial conditions on summer precipitation simulations over the United States and Mexico. *J Climate* 2006;19(15):3640–58.
- [40] Mocko DM, Sud YC. Refinements to SSIb with an emphasis on snow physics: evaluation and validation using GSWP and Valdai data. *Earth Interact* 200;15:1–31.
- [41] Moradkhani H, Sorooshian S, Gupta HV, Houser PR. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Adv Water Resour* 2005;28:135–47.
- [42] Njoku E. Updated daily AMSR-E/Aqua Daily L3 surface soil moisture, interpretive parameters, & QC EASE-Grids V002 [January 1, 2003 to December 31, 2003]. Boulder, CO, USA: National Snow and Ice Data Center; 2008 [Digital media].
- [43] Pan M, Wood EF, McLaughlin DB, Entekhabi D, Luo L. A multi-scale ensemble filtering system for hydrologic data assimilation: Part I. Implementation and synthetic experiment. *J Hydrometeorol* 2009;10(3):794–806. <http://dx.doi.org/10.1175/2009JHM1088.1>.
- [44] Parada LM, Liang X. Optimal multiscale Kalman filter for assimilation of near-surface soil moisture into land surface models. *J Geophys Res* 2004;109:D24109.1–21.
- [45] Perry CD, Vellidis G, Lowrance R, Thomas DL. Watershed-scale water quality impacts of riparian forest management. *J Water Resour Plann Manage* 1999;125:117–26.
- [46] Prigent C, Aires F, Rossow WB, Robock A. Sensitivity of satellite microwave and infrared observations to soil moisture at a global scale: relationship of satellite observations to in situ soil moisture measurements. *J Geophys Res* 2005;110:D07110. <http://dx.doi.org/10.1029/2004JD005087>.
- [47] Reichle RH, McLaughlin D, Entekhabi D. Hydrologic data assimilation with the ensemble Kalman filter. *Mon Weather Rev* 2002;130:103–14.
- [48] Reichle RH, Koster RD. Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. *J Hydrometeorol* 2003;4:1229–42.
- [49] Reichle RH, Koster RD. Bias reduction in short records of satellite soil moisture. *Geophys Res Lett* 2004;31:L19501. <http://dx.doi.org/10.1029/2004GL020938>.

- [50] Reichle RH, Koster RD, Dong J, Berg AA. Global soil moisture from satellite observations, land surface models, and ground data: implication for data assimilation. *J Hydrometeorol* 2004;5:430–42.
- [51] Reichle RH, Bosilovich MG, Crow WT, Koster RD, Kumar SV, Mahanama SPP, et al. Recent advances in land data assimilation at the NASA global modeling and assimilation office. In: Park Seon K, Xu Liang, editors. *Data assimilation for atmospheric, oceanic and hydrologic applications*. New York: Springer Verlag; 2009. p. 407–28. <http://dx.doi.org/10.1007/978-3-540-71056-1>.
- [52] Reichle RH, Kumar SV, Mahanama SPP, Koster RD, Liu Q. Assimilation of satellite-derived skin temperature observations into land surface models. *J Hydrometeorol* 2010;11:1103–22. <http://dx.doi.org/10.1175/2010JHM1262.1>.
- [53] Sahoo AK, Dirmeyer PA, Houser PR, Kafatos M. A study of land surface processes using land surface models over the Little River Experimental Watershed, Georgia. *J Geophys Res-Atmos* 2008;113:D20121. [1029/2007JD009671](http://dx.doi.org/10.1029/2007JD009671).
- [54] Sahoo AK, Houser PR, Ferguson C, Wood EF, Dirmeyer PA, Kafatos M. Evaluation of AMSR-E soil moisture results using the in situ data over the Little River Experimental Watershed, Georgia. *Remote Sens Environ* 2008;112(6):3142–52.
- [55] Sahoo AK, Pan M, Wood EF. Soil moisture retrieval using multi-channel passive microwave measurements through improved radiative transfer modelling. In: *Presented at remote sensing and hydrology 2010 symposium*, September 27–30, Jackson Hole, WY, USA; 2010.
- [56] Schaake JC, et al. An intercomparison of soil moisture fields in the North American Land Data Assimilation System (NLDAS). *J Geophys Res* 2004;109:D01S90. <http://dx.doi.org/10.1029/2002JD003309>.
- [57] Schär C, Luthi D, Beyerle U, Heise E. The soil–precipitation feedback: a process study with a regional climate model. *J Climate* 1999;12:722–41.
- [58] Seneviratne SI, Luthi D, Litschi M, Schär C. Land–atmosphere coupling and climate change in Europe. *Nature* 2006;443(7108):205–9.
- [59] Sheridan JM, Ferreira VA. Physical characteristic and geomorphic data for Little River Watersheds, Georgia. USDA-ARS, Southeast Watershed Research Laboratory Report 099201; 1992. 19pp.
- [60] Sheridan JM. Rainfall–streamflow relations for coastal plain watersheds. *Appl Eng Agric* 1997;13(3):333–44.
- [61] Sun CJ, Walker JP, Houser PR. A methodology for snow data assimilation in a land surface model. *J Geophys Res* 2004;109:D08108. <http://dx.doi.org/10.1029/2003JD003765>.
- [62] Vinnikov KY, Yezserkepova IB. Soil moisture: empirical data and model results. *J Climate* 1991;4:66–79.
- [63] Wagner W, Scipal K, Pathe C, Gerten D, Lucht W, Rudolf B. Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data. *J Geophys Res* 2003;108(D19):4611. <http://dx.doi.org/10.1029/2003JD003663>.
- [64] Wagner W, Blösch G, Pampaloni P, Calvet J-C, Bizzarri B, Wigneron J-P, et al. Operational readiness of microwave remote sensing of soil moisture for hydrologic applications. *Nordic Hydrol* 2007;38(1):1–20.
- [65] Walker JP, Houser PR. A methodology for initializing soil moisture in a global climate model: assimilation of near-surface soil moisture observations. *J Geophys Res* 2001;106:11761–74.
- [66] Walker JP, Houser PR. Requirements of a global near-surface soil moisture satellite mission: accuracy, repeat time, and spatial resolution. *Adv Water Resour* 2004;27:785–801.
- [67] Williams RG. Little River watersheds land use characteristics. USDA-ARS, SEWRL Lab. Note 098201; 1982.
- [68] Zhou Y, McLaughlin D, Entekhabi D, Crystal Ng G. An ensemble multiscale filter for large nonlinear data assimilation problems. *Mon Weather Rev* 2008;136:678–98.
- [69] Zhu H, Thorpe A. Predictability of extratropical cyclones: the influence of initial condition and model uncertainties. *J Atmos Sci* 2006;63(5):1483–97.