Accepted Manuscript

Assimilation and Downscaling of Satellite Observed Soil Moisture over the Little River Experimental Watershed in Georgia, USA

Alok Kumar Sahoo, Gabriëlle J.M. De Lannoy, Rolf H. Reichle, Paul R. Houser

PII: S0309-1708(12)00235-7
DOI: http://dx.doi.org/10.1016/j.advwatres.2012.08.007
Reference: ADWR 1943

To appear in: Advances in Water Resources

Please cite this article as: Sahoo, A.K., De Lannoy, G.J.M., Reichle, R.H., Houser, P.R., Assimilation and Downscaling of Satellite Observed Soil Moisture over the Little River Experimental Watershed in Georgia, USA, Advances in Water Resources (2012), doi: http://dx.doi.org/10.1016/j.advwatres.2012.08.007

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Assimilation and Downscaling of Satellite Observed Soil Moisture over the 
Little River Experimental Watershed in Georgia, USA

Alok Kumar Sahoo¹, Gabriëlle J. M. De Lannoy²³, Rolf H. Reichle³, Paul R. Houser⁴

¹Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ USA 08544
²Laboratory of Hydrology and Water Management, Ghent University, Coupure links 653, B-9000 Ghent, Belgium
³Global Modeling and Assimilation Office (Code 610.1), NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
⁴College of Sciences, George Mason University, Fairfax, Virginia, USA

*Corresponding Author: Alok K. Sahoo; Email: sahoo@princeton.edu
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.

Key Words: Data Assimilation, Kalman filter, Soil Moisture, Multi-scale, Little River Experimental Watershed, Satellite Observations
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 (Seneviratne et al., 2006; Schär et al., 1999). Soil moisture also influences the hydrologic cycles from local to global scales by redistributing rainfall into infiltration, surface runoff and evaporation at the earth surface (Delworth and Manabe, 1988; Vinnikov and Yeserkepova, 1991; Wagner et al., 2003); supplying water to the vegetation above the ground (Vinnikov and Yeserkepova, 1991); affecting the surface energy exchange (Prigent et al., 2005) and helping determine the Bowen ratio (Dirmeyer, 1995). 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 (Reichle et al., 2004; Sahoo et al., 2008a). 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 (Wagner et al., 2007).

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 to 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) 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. Dee and Todling, 2000; Reichle et al., 2004; Reichle and Koster, 2004; Dee, 2005; De Lannoy et al., 2007; Mahfouf, 2010; Liu et al., 2011).

Over the past few years, sequential data assimilation methods have been applied to soil moisture (e.g. Walker and Houser, 2001 & 2004; Crow and Van Loon, 2006; Pan et al., 2009; Liu et al., 2011), soil temperature (e.g. Huang et al., 2008; Reichle et al., 2010), snow (e.g. Sun et al., 2004; Andreadis and Lettenmaier, 2006; Clark et al., 2006; De Lannoy et al., 2012) 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. Reichle et al., 2002; Reichle and
Koster, 2003; Parada and Liang, 2004; Zhou et al., 2008; Pan et al., 2009) 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 observations (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. (2010) have used the same 3-D EnKF filter and downscaling method for snow data assimilation. However, our study differs from De Lannoy et al. (2010; 2012) 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. (2009), 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 (Figure 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://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 (Sheridan, 1997). 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 (Williams, 1982; Hubbard et al., 1985; Sheridan and Ferreira, 1992; Perry et al., 1999; Bosch et al., 1999; 2006). We define our experiment domain as a rectangular 100 km x 75 km area (~ 1
km\(^2\) fine scale resolution) as shown in Figure 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 (1990). 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. (2004; 2006) and Sahoo et al. (2008b). 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 (Sahoo et al., 2010). We also validated the same product at this LREW study location (Sahoo et al., 2008b) along with the official AMSR-E soil moisture product where the LSMEM product outperformed the official AMSR-E product (Njoku, 2008).

Our study domain contains 4 x 3 coarse 25 x 25 km\(^2\) 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 two to three 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

Figure 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 (Bosch et al., 2006) 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 (Figure 1) (Cashion et al., 2005).

Field observation data were obtained from the United States Department of Agriculture-Agricultural Research Service (USDA-ARS) located in Beltsville, Maryland (Jackson et al., 2006 & 2007) and USDA-ARS Southeast watershed research laboratory located in Tifton, Georgia (Bosch et al., 2007). The data used in this study include all available instantaneous soil moisture and precipitation measurements for the period 2002 to 2007 at sixteen out of the seventeen sites. These sixteen sites are 12, 16, 22, 26, 31, 32, 34, 39, 40, 43, 50, 62, 63, 66, 67 and 75 (Figure 1). A detailed description of these in-situ data can be found in Jackson et al. (2006) and Sahoo et al. (2008a). Again, unlike the satellite observations, the in-situ soil moisture data in this study are primarily used for validation and bias estimation purposes.

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 (Mahrt and Pan, 1984). 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 (Schaaeke et al., 2004). A detailed description of the Noah model is provided in Chen et al., (1996); Chen and Dudhia (2001) and Ek et al. (2003). 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; Mocko and Sud, 2001) and the CLM (Community Land Model, version 2.0; Dai et al., 2003) land surface models (Sahoo et al., 2008a).

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 (Kumar et al., 2006). For a comprehensive discussion of the LIS EnKF module, please refer to Kumar et al. (2008). The 3-dimentional EnKF (3-D EnKF; used in this study) algorithm was developed by the Global Modeling and Assimilation Office (GMAO/NASA, Reichle et al., 2009) and incorporated within LIS5.0 (after modifying LIS5.0 to enable the GMAO EnKF capability for horizontal error correlations, covariance localization, and distributed filtering; De Lannoy et al. 2010).

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 sixteen in-situ sites are used as the “truth” in this study and validation of the results is performed with respect to the in-situ observations. 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.
(Mitchell et al., 2004). 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 degree to 1 km² for the
simulations performed in this study. For land cover, we use the University of Maryland’s (UMD)
1-km global land cover product (Hansen et al., 2000). The sand, silt and clay fraction and soil
color data are generated at 1 km² resolution from the original Pennsylvania State University-
USDA State Soil Geographic Database (STATSGO) 1/120° resolution soil maps (Miller and
White, 1998). 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 (Kumar et al., 2008; Pan et al.,
2009; and De Lannoy et al., 2010), 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, stdv = 30
W/m²) and the near-surface air temperature (Ta, stdv = 2 °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,
stdv = 2) and the shortwave radiation (SW, stdv = 1.5). Cross-correlation between the forcing
perturbations is also included (LW–SW: -0.3, SW–P: -0.1, SW–Ta: 0.3, LW–P: 0.5, LW–Ta: 0.6, P–Ta: -0.1, assuring both positive definiteness 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. (2009) (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 (Evensen, 1994) and has been used frequently for land surface data assimilation (e.g. Margulis et al., 2002; Reichle et al., 2002; Moradkhani et al., 2005; De Lannoy et al., 2007 & 2012; Liu et al., 2011). 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 De Lannoy et al., 2010). The 3-D EnKF algorithm downscales the coarse observations within the assimilation scheme and uses multiple coarse observation grid cells (3D-Cm case of De Lannoy et al., 2010), as shown in Figure 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 $l_{\text{corr}} = 30$ km, which in turn covers at least the dimension of one coarse observation grid cell (De Lannoy et al., 2010)). 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 Figure 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:

$$\hat{x}_j^{k+} = \hat{x}_j^{k-} + K^k [y_j - \hat{y}_j]$$ (1)
Where the Kalman gain is defined as

$$ K^k = \text{Cov}[(\hat{x}^{-}, \hat{y}^{-})] \left[ \text{Cov}[(\hat{y}^{-}, \hat{y}^{-})] + R \right]^{-1} $$

(2)

In Equation (1), ‘\( \hat{x}^{+} \)’ is the a priori state estimate (or forecast) whereas ‘\( \hat{x}^{-} \)’ is the a posteriori state estimate (or analysis). The term ‘\( \hat{x}^{+} \)’ is a part of the total state vector ‘\( \hat{x}^{-} \)’ and contains the state variables at a single grid cell ‘\( k \)’. The term ‘\( \hat{x}^{+} \)’ in Equation (2) refers to the ensemble of forecasts ‘\( \hat{x}^{+} \)’ (cf. first term on the right hand side of the Equation (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 the Equation (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 the Equation (2) includes the error correlation between the ensemble of forecasted \( sm_1 \) over different coarse scale areas (\( \hat{y}^{-} \), size depends on the localization radius) and the ensemble of the eight fine scale state variables at a single fine scale model grid cell (\( \hat{x}^{-} \)). 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 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 Jul 1, 2002 through Dec 31, 2007 (prior to assimilation) is computed \((\hat{b} = \begin{bmatrix} b_1, 0, 0, \ldots, 0 \end{bmatrix}^T)\) 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.

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

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

\[
\hat{x}_j^{k+} = \hat{x}_j^{k+} + \hat{b}
\]  (4)
The analysis output (not fed back into the model) is the ensemble mean of the ensemble members \( \hat{x}_{j}^{k+} \) in Equation 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; Reichle and Koster, 2004), 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. (2007), 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 sixteen in-situ locations from 2002 to 2007 (Figure 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 (Sahoo et al., 2008a; 2008b). 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 m$^3$/m$^3$) prior to the data assimilation.

4.1.1. Temporal Evolution of the Spatially Averaged Top Layer Soil Moisture

Figure 3 shows the spatially averaged (over sixteen 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 to 2007. Figure 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 sixteen in-situ locations
(Figure 1). The corresponding in-situ total precipitation/irrigation data are shown in Figure 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.

Figure 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 Figure 4a, the number of data points is limited (1346 days versus 2010 days in other datasets) by the satellite revisit time (once every 2 to 3 days). The satellite observations show a small negative bias (-0.02 m$^3$/m$^3$) over this watershed. The grey highlighted box in Figure 4a shows that the satellite retrievals overestimate soil moisture for dry conditions and hence supports the findings in the time series plot in Figure 3. The ensemble open loop integration has a bias of 0.12 m$^3$/m$^3$ and a RMSE of 0.13 m$^3$/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 Figure 4b. Both the 1-D (Figure 4c) and 3-D (Figure 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 to 0.83) for all the soil moisture datasets with respect to the in-situ observations.

Figure 5 shows the temporal evolution of the spatial RMSE for the ensemble open loop and assimilation results, calculated over the sixteen in-situ stations. The spatial RMSE values range between 0.08 to 0.17 m$^3$/m$^3$ (open loop without bias correction) and 0.01 to 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 Figure 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.

Figure 6 shows the temporally averaged spatial maps of the analysis error standard
development (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 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
crude scale (25 km$^2$) 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 sixteen individual in-situ stations and
the nearby SCAN (SC) site. Figure 7 shows a comparison of the skill scores (bias,
ubRMSE≡sqrt(RMSE$^2$ – bias$^2$), 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 to 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 Figure 1 for station location). Figure 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 m$^3$/m$^3$ for the Station 32. The bias-corrected ensemble open loop results still contain positive bias (around 0.03 to 0.13 m$^3$/m$^3$). As discussed above we are here applying a first order bias correction (a constant value) is applied by determining the spatially averaged bias at sixteen 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 sixteen 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 m$^3$/m$^3$ respectively.

Figure 7b shows the time averaged ubRMSE values for all the simulation results for the same sixteen 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 (Entekhabi et al.,
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 Figure 7c and high correlation values are seen for 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 (Feb 26 to Mar 1, 2003) over our simulation domain. Figure 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 Figure 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 sixteen 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 (Reichle et al., 2009). For a consistently operating filter, the normalized innovations
sequence should be white, i.e. it 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 \(\sqrt{\text{Cov}[\hat{y}^m, \hat{y}^m] + [\hat{R}]_{m,m}}\). 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 \(m^3/m^3\) as the observation error standard deviation in this study.

Figure 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 Figure 9a are at the fine scale (1 km grid; total 7500 grid cells over the simulation domain) whereas the 3-D EnKF innovations in Figure 9b are at the coarse scale (25 km grid; total 12 grid cells over the simulation domain). The dotted vertical line in Figures 9a and 9b 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 sixteen 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
Figure 10 shows the time series for the moisture in the model 2\textsuperscript{nd} layer (Figure 10a), 3\textsuperscript{rd} layer (Figure 10b) and 4\textsuperscript{th} layer (Figure 10c) for the ensemble open loop, 1-D and 3-D EnKF simulations averaged over the sixteen 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.

Figure 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 a difference between the ensemble open loop and 3-D EnKF deep layer soil moisture values as high as 0.10 m\(^3\)/m\(^3\) (2\textsuperscript{nd} layer) to 0.15 m\(^3\)/m\(^3\) (4\textsuperscript{th} 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 (1\textsuperscript{st} layer: 0-0.1 m, 2\textsuperscript{nd}: 0.1-0.4 m, 3\textsuperscript{rd}: 0.4-1 m and 4\textsuperscript{th}: 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 sub-surface 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 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. Mo et al., 2006; Zhu and Thorpe, 2006) 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 m$^3$/m$^3$), (b) intermediate start (0.25 m$^3$/m$^3$) and (c) wet start (0.48 m$^3$/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

Figure 11 shows one 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 sixteen 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 Jan 1, 2003 (winter start) which are discussed later in Section 4.3.2.

Figure 11a shows the results for the model top layer (0 to 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 values 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.

Figure 11b-d show 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, Kumar et al., 2008). 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 Jan 1, 2003 as opposed to the 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 (Jan 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; Jan 1 to Feb 15, 2003) at the beginning of these winter simulations which is followed by a long, continuous and heavy precipitation period (around 2 months; Feb 15 to Apr 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.

**Acknowledgement**

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


Note 098201.


**List of Figures**

**Figure 1:** Little River Experimental Watershed (LREW), Tifton, Georgia, USA [from Cashion et al., 2005]. 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 sixteen 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 x 3) coarse-scale satellite grid cells and 7500 (100 x 75) fine-scale model grid cells. The sixteen in-situ stations are: 12, 16, 22, 26, 31, 32, 34, 39, 40, 43, 50, 62, 63, 66, 67 and 75.

**Figure 2:** A Schematic diagram of the 3-D EnKF approach illustrated for four coarse-scale pixels, each containing 4 x 6 fine-scale pixels.

**Figure 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 sixteen in-situ locations in the Little River Watershed.

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

**Figure 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 sixteen in-situ stations.

**Figure 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 m$^3$/m$^3$.

**Figure 7:** (a) Mean bias, (b) $ubRMSE = \sqrt{RMSE^2 - \text{bias}^2}$, 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 sixteen 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.

**Figure 8:** Spatial precipitation and soil moisture difference maps during a wet event for the period Feb 26 – Mar 1, 2003.

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

**Figure 10:** Daily time series of the station averaged moisture from (a) Layer 2 (0.1 to 0.4 m), (b) Layer 3 (0.4 to 1 m) and (c) Layer 4 (1 to 2 m) for all the ensemble open loop and EnKF simulations.
Figure 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 sixteen in-situ locations. Refer to Table 4 for the description of the Experiments in the Legend.

List of Tables

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

Table 2: Correlation between the model simulated 2\textsuperscript{nd} 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 sixteen in-situ locations. Also shown are the station average (AVG) correlation values.

Table 3: Percentage change in water cycle variables introduced by surface soil moisture assimilation averaged over the entire study period (July 2002 to December 2007).

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).

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 (Jul 1, 2002) and in the winter (Jan 1, 2003).
FIGURE 1
Observation Grid Scale

Model Grid Scale

Upscaling (H) of the state forecast to the coarse observation scale

Downscaling (K) of the innovation to the fine model scale

FIGURE 2
FIGURE 3
Figure 4
Figure 5
FIGURE 6
Figure

(a) Bias (SM) (m$^3$/m$^3$)

(b) ubRMSE (SM) (m$^3$/m$^3$)

(c) Correlation (SM) (-)

Open Loop
Open Loop (BC)
1-D EnKF
3-D EnKF

FIGURE 7
FIGURE 8
Figure 9

Bins for normalized innovations

Frequency (%)

-3 -2 -1 0 1 2 3

Mean: -0.23
Stdv: 0.72

(a)

Bins for normalized innovations

Frequency (%)

-3 -2 -1 0 1 2 3

Mean: -0.16
Stdv: 0.75

(b)
Figure 10

![Figure 10](image)
Soil Moisture ($m^3/m^3$)

Date

OL-LS  OL-MS  OL-HS  OL-RS  3D-LS  3D-MS  3D-HS  3D-RS

FIGURE 11
<table>
<thead>
<tr>
<th>Variable</th>
<th>Std dev (m$^3$/m$^3$ or K)</th>
<th>Cross correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sm$_1$</td>
<td>sm$_2$</td>
</tr>
<tr>
<td>sm$_1$</td>
<td>6.00E-3</td>
<td>0.6</td>
</tr>
<tr>
<td>sm$_2$</td>
<td>1.10E-4</td>
<td>1</td>
</tr>
<tr>
<td>sm$_3$</td>
<td>6.00E-5</td>
<td>0.4</td>
</tr>
<tr>
<td>sm$_4$</td>
<td>4.00E-5</td>
<td>0.2</td>
</tr>
<tr>
<td>st$_1$</td>
<td>1.00E-1</td>
<td>0</td>
</tr>
<tr>
<td>st$_2$</td>
<td>5.00E-2</td>
<td>0</td>
</tr>
<tr>
<td>st$_3$</td>
<td>5.00E-2</td>
<td>0</td>
</tr>
<tr>
<td>st$_4$</td>
<td>5.00E-2</td>
<td>0</td>
</tr>
<tr>
<td>Sites</td>
<td>Ensemble Open Loop</td>
<td>1-D EnKF</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------</td>
<td>----------</td>
</tr>
<tr>
<td>RG12</td>
<td>0.76</td>
<td>0.79</td>
</tr>
<tr>
<td>RG16</td>
<td>0.63</td>
<td>0.69</td>
</tr>
<tr>
<td>RG22</td>
<td>0.69</td>
<td>0.70</td>
</tr>
<tr>
<td>RG26</td>
<td>0.54</td>
<td>0.61</td>
</tr>
<tr>
<td>RG31</td>
<td>0.79</td>
<td>0.71</td>
</tr>
<tr>
<td>RG32</td>
<td>0.65</td>
<td>0.68</td>
</tr>
<tr>
<td>RG34</td>
<td>0.65</td>
<td>0.63</td>
</tr>
<tr>
<td>RG39</td>
<td>0.67</td>
<td>0.63</td>
</tr>
<tr>
<td>RG40</td>
<td>0.71</td>
<td>0.65</td>
</tr>
<tr>
<td>RG43</td>
<td>0.66</td>
<td>0.61</td>
</tr>
<tr>
<td>RG50</td>
<td>0.70</td>
<td>0.66</td>
</tr>
<tr>
<td>RG62</td>
<td>0.60</td>
<td>0.61</td>
</tr>
<tr>
<td>RG63</td>
<td>0.59</td>
<td>0.62</td>
</tr>
<tr>
<td>RG66</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>RG67</td>
<td>0.68</td>
<td>0.64</td>
</tr>
<tr>
<td>RG75</td>
<td>0.66</td>
<td>0.70</td>
</tr>
<tr>
<td>AVG</td>
<td>0.66</td>
<td>0.65</td>
</tr>
<tr>
<td>Variable</td>
<td>1-D EnKF</td>
<td>3-D EnKF</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Surface Runoff</td>
<td>12%</td>
<td>11%</td>
</tr>
<tr>
<td>Subsurface Runoff</td>
<td>135%</td>
<td>78%</td>
</tr>
<tr>
<td>Evaporation</td>
<td>30%</td>
<td>29%</td>
</tr>
</tbody>
</table>
**TABLE 4**

<table>
<thead>
<tr>
<th>Experiment Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OL-LX</td>
<td>Open loop simulation with SM I.C. = 0.02 m$^3$/m$^3$</td>
</tr>
<tr>
<td>OL-MX</td>
<td>Open loop simulation with SM I.C. = 0.25 m$^3$/m$^3$</td>
</tr>
<tr>
<td>OL-HX</td>
<td>Open loop simulation with SM I.C. = 0.48 m$^3$/m$^3$</td>
</tr>
<tr>
<td>OL-RX</td>
<td>Open loop simulation with SM I.C. from spin-up simulations</td>
</tr>
<tr>
<td>3D-LX</td>
<td>3-D EnKF simulation with SM I.C. = 0.02 m$^3$/m$^3$</td>
</tr>
<tr>
<td>3D-MX</td>
<td>3-D EnKF simulation with SM I.C. = 0.25 m$^3$/m$^3$</td>
</tr>
<tr>
<td>3D-HX</td>
<td>3-D EnKF simulation with SM I.C. = 0.48 m$^3$/m$^3$</td>
</tr>
<tr>
<td>3D-RX</td>
<td>3-D EnKF simulation with SM I.C. from spin-up simulations</td>
</tr>
</tbody>
</table>
TABLE 5

Open Loop Simulations

<table>
<thead>
<tr>
<th>Layer</th>
<th>OL-LS</th>
<th>OL-MS</th>
<th>OL-HS</th>
<th>OL-LW</th>
<th>OL-MW</th>
<th>OL-HW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer1</td>
<td>103</td>
<td>61</td>
<td>78</td>
<td>61</td>
<td>31</td>
<td>38</td>
</tr>
<tr>
<td>Layer2</td>
<td>163</td>
<td>90</td>
<td>115</td>
<td>93</td>
<td>45</td>
<td>52</td>
</tr>
<tr>
<td>Layer3</td>
<td>229</td>
<td>132</td>
<td>167</td>
<td>303</td>
<td>192</td>
<td>250</td>
</tr>
<tr>
<td>Layer4</td>
<td>322</td>
<td>185</td>
<td>253</td>
<td>365+</td>
<td>245</td>
<td>305</td>
</tr>
</tbody>
</table>

3-D EnKF Simulations

<table>
<thead>
<tr>
<th>Layer</th>
<th>3D-LS</th>
<th>3D-MS</th>
<th>3D-HS</th>
<th>3D-LW</th>
<th>3D-MW</th>
<th>3D-HW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer1</td>
<td>74</td>
<td>24</td>
<td>42</td>
<td>13</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Layer2</td>
<td>105</td>
<td>48</td>
<td>64</td>
<td>80</td>
<td>45</td>
<td>55</td>
</tr>
<tr>
<td>Layer3</td>
<td>166</td>
<td>102</td>
<td>134</td>
<td>250</td>
<td>166</td>
<td>200</td>
</tr>
<tr>
<td>Layer4</td>
<td>257</td>
<td>164</td>
<td>217</td>
<td>320</td>
<td>203</td>
<td>230</td>
</tr>
</tbody>
</table>
Highlights

> In this study we analyze the performance of a 3-D EnKF technique. > Both 1-D and 3-D algorithms improve model forecast soil moisture over our study area. > Surface moisture assimilation impacts deep layer moisture and water cycle variables. > EnKF speeds up the state balancing relative to the open loop integration.