

## NOTES AND CORRESPONDENCE

### Revising the Ensemble-Based Kalman Filter Covariance for the Retrieval of Deep-Layer Soil Moisture

SHU-WEN ZHANG

*Key Laboratory for Semi-Arid Climate Change of the Ministry of Education, College of Atmospheric Sciences, Lanzhou University, and Key Laboratory of Arid Climate Changing and Reducing Disaster of Gansu Province, Lanzhou, China, and Department of Atmospheric Sciences, The University of Arizona, Tucson, Arizona*

XUBIN ZENG

*Department of Atmospheric Sciences, The University of Arizona, Tucson, Arizona*

WEIDONG ZHANG

*Key Laboratory for Semi-Arid Climate Change of the Ministry of Education, College of Atmospheric Sciences, Lanzhou University, and Key Laboratory of Arid Climate Changing and Reducing Disaster of Gansu Province, Lanzhou, China*

MICHAEL BARLAGE

*Research Applications Laboratory, NCAR, Boulder, Colorado*

(Manuscript received 5 January 2009, in final form 25 August 2009)

#### ABSTRACT

Previous studies have demonstrated that soil moisture in the top layers (e.g., within the top 1-m depth) can be retrieved by assimilating near-surface soil moisture observations into a land surface model using ensemble-based data assimilation algorithms. However, it remains a challenging issue to provide good estimates of soil moisture in the deep layers, because the error correlation between the surface and deep layers is low and hence is easily influenced by the physically limited range of soil moisture, probably resulting in a large noise-to-signal ratio. Furthermore, the temporally correlated errors between the surface and deep layers and the nonlinearity of the system make the retrieval even more difficult. To tackle these problems, a revised ensemble-based Kalman filter covariance method is proposed by constraining error covariance estimates in deep layers in two ways: 1) explicitly using the error covariance at the previous time step and 2) limiting the increase of the soil moisture error correlation with the increase of the vertical distance between the two layers. This method is then tested at three separate point locations representing different precipitation regimes. It is found that the proposed method can effectively control the abrupt changes of error covariance estimates between the surface layer and two deep layers. It significantly improves the estimates of soil moisture in the two deep layers with daily updating. For example, relative to the initial background error, after 150 daily updates, the error in the deepest layer reduces to 11.4%, 32.3%, and 27.1% at the wet, dry, and medium wetness locations, only reducing to 62.3%, 80.8%, and 47.5% with the original method, respectively. However, the improvement of deep-layer soil moisture retrieval is very slight when the updating frequency is reduced to once every three days.

---

#### 1. Motivation

The importance of a realistic soil moisture initialization in land-atmosphere coupled global models for seasonal-to-interannual forecasts is widely recognized (e.g., Koster et al. 2004). However, at present, no operational networks

---

*Corresponding author address:* Shu-Wen Zhang, College of Atmospheric Sciences, Lanzhou University, Lanzhou Gansu 73000, China.  
E-mail: zhangsw@lzu.edu.cn

of ground-based instruments are available at regional or global scales to provide soil moisture observations for land surface initialization. Satellite microwave sensors can provide global near-surface soil moisture estimates, but the measurement is limited to the top few centimeters of soil. By assimilating these near-surface soil moisture estimates into land surface models, the soil moisture profile in the top soil layers (e.g., within the top 1-m depth) can be realistically estimated with land data assimilation schemes (e.g., Walker et al. 2001; Zhang et al. 2005; Sabater et al. 2007). The ensemble-based Kalman filters are widely used in the land data assimilation system (LDAS), because they are better adapted to compute an approximation to the “true” state error covariance through a cloud of sampled members, without any need for hypotheses about the linearity of the land model (Reichle et al. 2002; Crow and Wood 2003; Zhou et al. 2006).

So far, most soil moisture profile estimates are confined to the top 1-m depth (e.g., Entekhabi et al. 1994; Hoeben and Troch 2000; Walker et al. 2001) by assimilating near-surface soil moisture observation into a land surface model, probably because of the relatively high error correlation between near-surface soil moisture observations and analysis points within this depth. However, it is difficult to produce good filter update for the deep-layer soil moisture by using the ensemble-based error covariance estimate, even if a large ensemble size is used. One main reason is that the error correlation between the surface and deep layers is low and hence is very easily influenced by the limited range of soil moisture, resulting in a large noise-to-signal ratio (Hamill et al. 2001). In particular, when the soil is either very wet or very dry, many ensemble members can be beyond the two limits of soil moisture—that is, zero and saturation. To make each of them physically reasonable, the unphysical member has to be discarded or replaced with a new member. This can result in large errors in the ensemble-based error covariance.

The nonlinearity of the system could be another reason. Kalman filters are based on the assumption that errors at two depths correlate linearly to each other and add linear corrections to the state estimate, although ensemble-based filters do allow nonlinear dynamics in state propagation from one time step to another. However, deep-layer soil moisture errors may not respond linearly to surface errors, because soil moisture is governed by the nonlinear Richards equation. Therefore, the linear “covariance” may not be able to describe this nonlinear relationship, and the linear update may not be useful.

Furthermore, temporally correlated errors may make the retrieval even more difficult. The surface serves as the

boundary condition for the soil moisture Richards equation to force changes downward, causing a time lag in errors between the surface and the deep layers. In other words, the “system noise” is correlated in time between the surface and the deep layers. Therefore, the most basic optimality condition (i.e., temporally uncorrelated noise) for Kalman filters would be directly violated.

Soil moisture retrieval in deep layers is important for land–atmosphere interaction studies, because soil moisture there is available for vegetation transpiration through vegetation roots, and it has a longer memory than near-surface soil moisture and hence affects atmospheric prediction at longer (e.g., weekly to even seasonal) time scales. The purpose of this paper is to propose a revised ensemble-based Kalman filter error covariance method (hereafter, referred to as the new method) for soil moisture retrieval in deep layers based on the considerations mentioned earlier. Although the new method does not fully solve these problems, it can effectively reduce their negative effects on the error covariance estimate. It is also worth noting that the new method is very easy to implement, with only a minor revision to the background error covariance at the deep layers before each update.

Section 2 describes the two assimilation algorithms [the ensemble square root filter (EnSRF) and the new method], land surface model, experimental locations, and uncertainties in model inputs and measurement errors. Section 3 presents the results along with sensitivity tests and further discussions. The conclusions are presented in section 4.

## 2. Methodology

### a. Assimilation approach

#### 1) THE ENSEMBLE SQUARE ROOT FILTER

The ensemble-based assimilation methodology has taken two general forms: stochastic filters (e.g., Evensen 1994; Burgers et al. 1998) and deterministic filters (e.g., Whitaker and Hamill 2002). The latter is designed to avoid one source of sampling errors associated with the use of “perturbed observations” in the stochastic filter. Different forms of deterministic filters have been proposed, such as the ensemble transform Kalman filter (Bishop et al. 2001), the ensemble adjustment Kalman filter (Anderson 2001), and the EnSRF (Whitaker and Hamill 2002).

In this study, the EnSRF is used for its convenience. Here, only a short introduction is given on the EnSRF but further details can be found in Whitaker and Hamill (2002). First, let  $\theta$  be an  $m$ -dimensional state vector, and express  $\theta$  as the sum of an ensemble mean (denoted by an overbar) and the deviations from the ensemble mean

(denoted by a prime), that is,  $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}} + \boldsymbol{\theta}'$ . Then, the analysis step for the EnSRF is formulated as

$$\bar{\boldsymbol{\theta}}^a = \bar{\boldsymbol{\theta}}^b + \mathbf{K}(\bar{\mathbf{y}} - \mathbf{H}\bar{\boldsymbol{\theta}}^b) \quad \text{and} \quad (1)$$

$$\boldsymbol{\theta}'^a = \boldsymbol{\theta}'^b - \tilde{\mathbf{K}}\mathbf{H}\boldsymbol{\theta}'^b, \quad (2)$$

where superscripts  $a$  and  $b$  denote the analysis quantity and background model forecast, respectively;  $\bar{\mathbf{y}}$  is the observation vector;  $\mathbf{H}$  is the linear observation operator,  $\mathbf{K}$  is the traditional Kalman gain given by

$$\mathbf{K} = \mathbf{P}^b \mathbf{H}^T (\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R})^{-1}, \quad (3)$$

and  $\tilde{\mathbf{K}}$  is given by

$$\tilde{\mathbf{K}} = \mathbf{P}^b \mathbf{H}^T \left[ \left( \sqrt{\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R}} \right)^{-1} \right]^T \left( \sqrt{\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R} + \sqrt{\mathbf{R}}} \right)^{-1}, \quad (4)$$

where  $\mathbf{R}$  is the observation-error covariance matrix, the superscript T means transpose, and  $\mathbf{P}^b$  is the sample background covariance matrix defined by

$$\mathbf{P}^b = \overline{\boldsymbol{\theta}'^b \boldsymbol{\theta}'^{bT}} \equiv \frac{1}{M-1} \sum_{i=1}^{i=M} \boldsymbol{\theta}'^b \boldsymbol{\theta}'^{bT}, \quad (5)$$

where  $M$  is the ensemble size. For an individual observation and assuming  $\tilde{\mathbf{K}} = \mathbf{K}/\alpha$ , then  $\alpha = 1 + \sqrt{\mathbf{R}/(\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R})}$  (Whitaker and Hamill 2002).

In our Observing System Simulation Experiment (OSSE), only one soil moisture observation at the second layer of the Community Land Model, version 3 (CLM3) is assimilated with  $\mathbf{H} = (0, 1, 0, \dots, 0)^T$ , so that only the second column of the background-error covariance matrix  $\mathbf{P}^b$  [i.e.,  $\mathbf{P}^b \mathbf{H}^T$  in (3)] is useful for assimilation. For simplicity, denote the second column of  $\mathbf{P}^b$  at time  $t_n$  by a vector  $\mathbf{u}^n$  with the elements  $u_i^n = p_{i,2}^n$ , where the subscript  $i$  refers to 1 of the 10 layers of the CLM and  $p_{i,2}^n$  is the entry in the  $i$ th row and the second column of  $\mathbf{P}^b$  at  $t_n$ . In the new method, only the covariance estimates in the second column need revising.

## 2) THE NEW METHOD

In general, the soil moisture error correlation between two layers should decrease with the increase of the vertical distance between them. On the basis of this assumption, in the first step, the potentially large change of error correlation at the deep layers is constrained by

$$|\rho_i^n| \leq |\rho_{i-1}^n|, \quad (6)$$

where  $\rho_i^n$  is the error correlation of soil moisture between the observation point and the analysis point in the

$i$ th layer (with soil layers numbering from top to bottom) at time  $t_n$ . Note that Eq. (6) is only imposed on the bottom three layers, because the error correlations at those layers with the observation point are relatively low. These error correlations may be easily influenced by the physically limited range of soil moisture (particularly when the soil is very dry or very wet) and also probably by the nonlinearity of the system, resulting in large noise-to-signal ratios (Hamill et al. 2001). If constraint (6) is violated in a layer  $i$  at  $t_n$ , let  $|\rho_i^n| = |\rho_{i-1}^n|$  and one revised error covariance estimate  $u_i^n$  is calculated accordingly:

$$\left| \frac{u_i^n}{\sigma_i^n \sigma^n} \right| = |\rho_i^n| = |\rho_{i-1}^n|. \quad (7)$$

Here,  $\sigma^n$  and  $\sigma_i^n$  are the standard deviations in the observational point and the  $i$ th layer at  $t_n$ , respectively. Notice that the sign of the error correlation and covariance in (7) remains the same as the original.

In contrast, if constraint (6) is satisfied, then  $u_i^n$  is simply taken as  $u_i^n$ , which is the background-error covariance estimate between the observation and the layer  $i$  at  $t_n$ .

Regarding the temporally correlated errors, as stated in section 1, the soil moisture error is correlated in time between the surface and bottom layers, so that the error covariance at the former time step might contain useful observation information for data assimilation at the present time step. On the basis of this consideration, another  $u_i^n$  is calculated as a linear combination of error covariance estimates at  $t_n$  and previous  $t_{n-1}$ :

$$u_i^n = (1 - \omega)u_i^{n-1} + \omega u_i^n. \quad (8)$$

Here,  $\omega$  is a relaxation factor, which is determined based on trial and error (to be discussed later). A similar approach was used by Xu et al. (2008), but they sampled a series of (preferably three in their experiment) perturbed state vectors to approximate the error covariance.

Lastly, of  $u_i^n$  and  $u_i^{n-1}$ , the larger one (in absolute value) is chosen as the new background error covariance estimate in layer  $i$  at  $t_n$  to prevent the underestimates of the background-error covariance, because the underestimation has a more severe effect on analysis error than overestimation (Daley 1991).

### *b. Land surface model, observations, and experimental locations*

In the OSSE, the land surface model used is the CLM, in which the vertical movement of soil moisture is governed by the Richards equation. The CLM is a one-column

TABLE 1. Summary of uncertainties in the inputs and observational errors for the land surface simulation experiment.

Variable	Specified nominal value	Uncertainties in replicates
Humidity	Qian et al. (2006)	Temporally uncorrelated multiplicative uniform noise: relative humidity: $U(0.9, 1.1)$ , solar radiation: $U(0.9, 1.1)$ , wind speed: $U(0.7, 1.3)$
Solar radiation		
Wind speed		
Air temperature	Qian et al. (2006)	Temporally uncorrelated additive uniform noise $U(-4, 4 \text{ K})$
Precipitation	Qian et al. (2006)	Temporally uncorrelated multiplicative lognormal noise with a mean of 1.0 and a standard deviation equal to 35% of the nominal value
Leaf area index (LAI)	Monthly LAI = 1.67, 4.07, 4.78, 4.23, and 1.88, respectively, from May to September	Multiplicative uniform noise $U(0.85, 1.15)$ for leaf area index (LAI) without change with time (i.e., static)
Soil fraction (sand–clay)	CLM surface dataset	Uniform additive noise $U(-10, 10; -10, 10)$ for sand and clay percentages without change with time (i.e., static)
Initial volumetric soil moisture	0.29 at all layers (poorly guessed)	Vertically uncorrelated additive zero-mean Gaussian noise with variance of $0.15^2$ (at dry and medium wetness area) or $0.1^2$ (at wet area) for each of the 10 layers
Volumetric soil moisture observation	Simulated true value	Temporally uncorrelated additive zero-mean Gaussian noise with variance of $0.05^2$

land surface model and has 10 soil layers with center points located at approximately 0.7, 2.8, 6.2, 12, 21, 37, 62, 104, 173 and 286 cm below the surface, respectively (Zeng et al. 2002; Oleson et al. 2004). The input to the CLM at each time step (five minutes) includes incident solar radiation, incident longwave radiation, precipitation, temperature, horizontal wind components, specific humidity, and pressure (Qian et al. 2006).

Because soil moisture in deep soil layers takes tens of years to spin up and stabilize, the CLM control simulation is spun up for 100 years, forced by cycling the 1998 National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis data (Qian et al. 2006). In this OSSE, the final spun-up results are chosen as the true states, whereas the “observations” are obtained by adding random errors with the specified statistics into the true model states.

To test the applicability of the new method at different precipitation regimes, three separate point locations have been chosen with the total precipitation of 568 (37.0°N, 86.2°W), 104 (31.4°N, 105.0°W), and 404 mm (45.5°N, 96.5°W), respectively, from 1 May to 30 September 1998 (hereafter referred to as wet, dry, and medium wetness locations, respectively). The new method is then separately tested at each location. In other words, no horizontal error correlation among them is used during data assimilation. The land cover and soil texture (e.g., sand and clay percentages) data at the medium wetness location are used for all three locations to isolate the influence of different precipitation amounts. The plant functional type is  $C_3$  grass. For the 10 CLM layers, the percentages of sand in the soil from the top to bottom layer are 18, 18, 18, 18, 17, 16, 16, 15, 20, and 20, respectively, and the percentages of clay are 36, 36, 36, 35, 35, 32, 31, 30, 24, and 24, respectively.

### c. Uncertainties in model inputs and measurement errors

In the OSSE, the replicates of the land surface states are generated from random values of uncertain model inputs; Table 1 lists measurement errors and all uncertain model inputs for this application with 10 initial background state variables (i.e., soil moisture at each of the 10 CLM layers) from the forecasts, 5 from the atmospheric forcing data (i.e., precipitation, downward solar radiation, air temperature, humidity, and wind) and 3 from vegetation and soil texture parameters (i.e., sand and clay percentages). A similar approach has been used in the past (e.g., Zhou et al. 2006).

For each uncertainty scenario, 40 ensemble members are used. To roughly represent soil moisture data that are expected to be available from satellite missions planned for the near future, only near-surface soil moisture observations are assimilated. Because layer 1 in the CLM is very thin and the soil moisture measurement from satellite microwave sensors is the averaged soil moisture of the top few centimeters, the layer-2 soil moisture with the midpoint at the depth of 2.8 cm might be much closer to the satellite observation; thus, soil moisture in layer 2 is assimilated. The updating frequency is once a day, and sensitivity test results using an updating frequency of once every three days will also be discussed. The overall assimilation length is 153 days from 1 May to 30 September.

## 3. Results

### a. Comparison of the estimated soil moisture from the different methods

First, the relaxation factor  $\omega$  needs to be calibrated for the new method. Figure 1 plots the corresponding errors

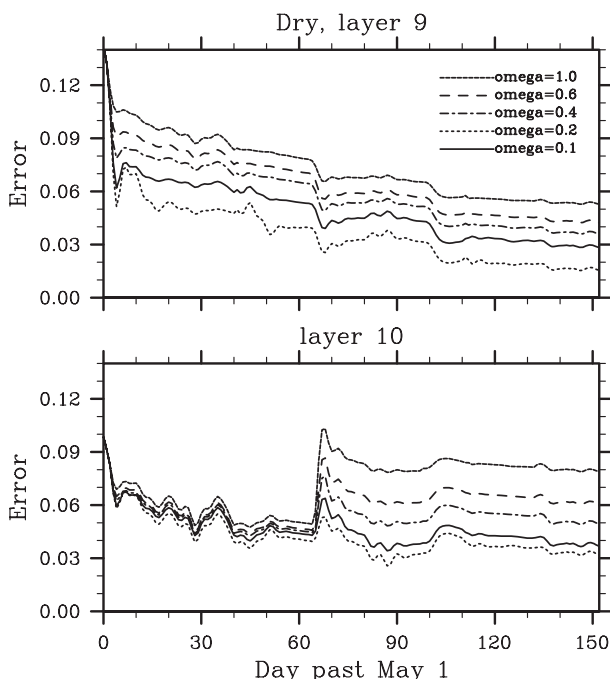


FIG. 1. Comparison of the absolute errors of volumetric soil moisture estimates with the new method using five different relaxation factors in (top) layer 9 and (bottom) layer 10 at the dry location.

in the last two layers at the dry location by using five different relaxation factors in Eq. (8). It is obvious that 0.2 is an optimal relaxation factor for this location. This is also the case for the other two locations (results not shown), so the relaxation factor is taken as 0.2 in the following tests.

Figure 2 compares the estimated soil moisture profiles from the EnSRF and the new method after 0, 50, 100, and 150 daily updates. To demonstrate the benefit of assimilating the observations into the CLM, Fig. 2 also shows the open-loop results (i.e., a 40-member ensemble CLM integration with perturbations but without data assimilation). It is clear that both data assimilation algorithms can retrieve the top 1-m soil moisture profile at the three locations, consistent with previous studies (e.g., Entekhabi et al. 1994; Hoeben and Troch 2000; Walker et al. 2001). However, for the two deep layers (i.e., layer 9 and layer 10), the soil moisture estimates with the new method are much closer to the truth than those from the EnSRF. Because the initial background errors are not the same at the three locations (see Table 1), a relative error (i.e., the ratio between the analysis error and the initial background error) is also adopted for a fair comparison. At the final update, the relative error at layer 10 reduces to 11.4%, 32.3%, and 27.1% at the wet, dry, and medium wetness locations

with the new method, respectively, but only to 62.3%, 80.8%, and 47.5% with the EnSRF and to 93.0%, 74.9% and 79.3% from the open-loop simulation, respectively.

Sensitivity tests also show that further increasing the ensemble members (e.g., using 80, 120, and 200 members) does not further improve the performance, probably because the original ensemble size (40) is already larger than the total number of model states plus uncertainties in vegetation and soil parameters as well as atmospheric forcing (Table 1).

To understand why the new method can give a better result at the two deep layers, the evolution of error and error covariance is investigated. Figure 3 shows the absolute errors of the ensemble mean in layer 9 (first row) and layer 10 (second row) from the EnSRF, the new method, and open-loop simulation. To explain the influences of precipitation on the soil moisture retrieval, the rainfall time series at the three locations are added into Fig. 3 (last row).

For the wet location (first column in Fig. 3), in contrast to the estimates in the layers within the top 1-m depth, the retrieval speed in the two deep layers is slow; after several update cycles, soil moisture estimates in the two deep layers are even worse than those from the open-loop simulation. However, as the update proceeds, the new method gives better estimates of soil moisture there than the other two methods. For example, the volumetric soil moisture error in layer 10 suddenly increases from 0.072 on the 50th day to 0.124 on the 52nd day with the EnSRF, whereas the error only increases from 0.064 to 0.085 with the new method. This can be explained by the change of error covariance in the first column of Fig. 4, in which the corresponding error covariance between the observation layer and layer 10 abruptly increases by an order of magnitude with the EnSRF—that is, from  $0.98 \times 10^{-4}$  to  $8.3 \times 10^{-4}$ , whereas it increases by much less, from  $1.0 \times 10^{-4}$  to  $2.5 \times 10^{-4}$ , with the new method. A similar abrupt change also appears with the EnSRF on the 11th and 91st days in layer 10 as well as on the 80th and 100th days in layer 9 (first column of Fig. 4). In addition to these days, the error covariance estimates with the new method generally change smoothly.

For the dry location (see the second column of Fig. 3), the errors in layer 9 with the new method decrease quickly from the 4th day and remain much smaller than those using the other two methods. In layer 10, the new method successfully prevents the abrupt error increase on the 67th day in the EnSRF and gives good estimates of soil moisture. In contrast, the errors in layer 10 with the EnSRF become even larger than those of the open-loop simulation from this day forward. The better performance of the new method is again associated with the revised covariance estimates (the second column of

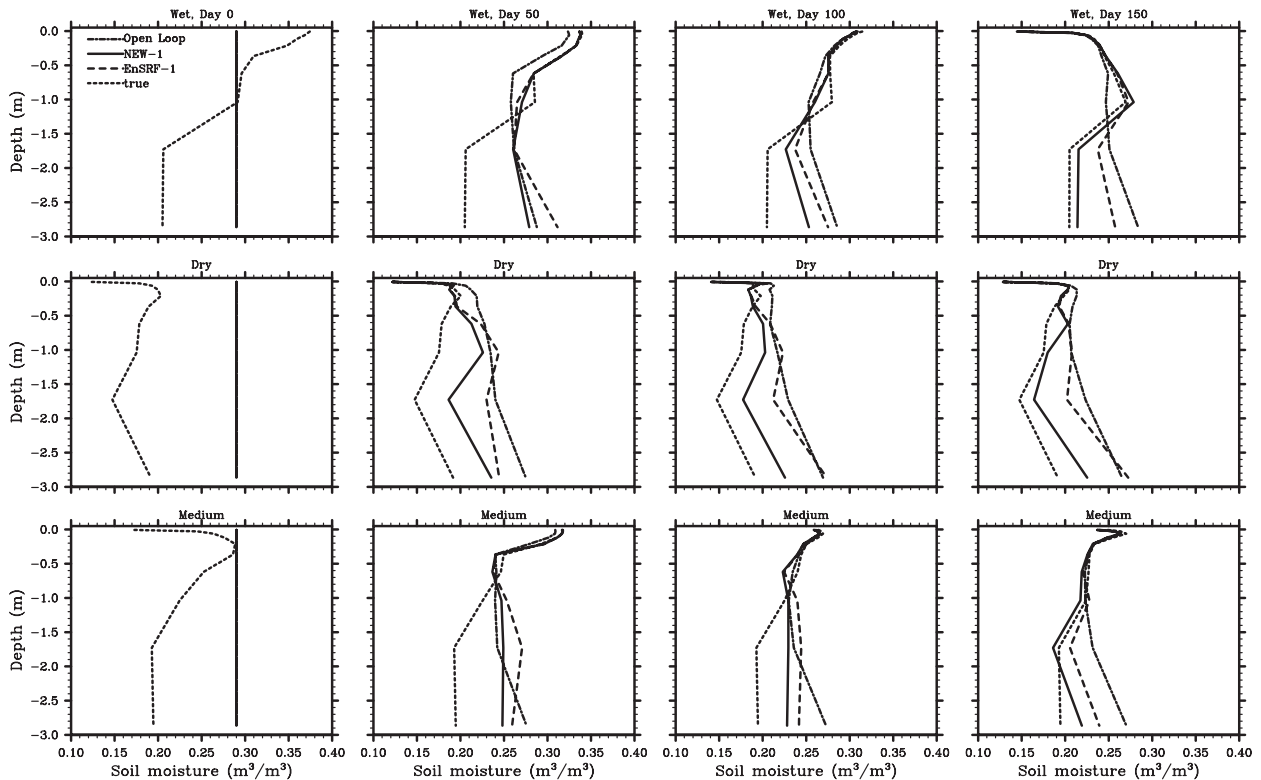


FIG. 2. Comparison at days 0, 50, 100, and 150 of two estimated soil moisture profiles by assimilating the observations once a day with the EnSRF (EnSRF-1) and the new method (NEW-1) with both the true soil moisture profile (true) and the open-loop simulation (Open Loop) at the (top) wet, (middle) dry, and (bottom) medium wetness locations.

Fig. 4); for example, on these two days (4th and 67th), the error covariance estimates with the EnSRF abruptly drops from  $8.8 \times 10^{-4}$  to  $1.6 \times 10^{-4}$  (in layer 9 on the 4th day) and from  $0.49 \times 10^{-4}$  to  $-5.4 \times 10^{-4}$  (in layer 10 on the 67th day), respectively, whereas those estimates of the new method drop from  $8.8 \times 10^{-4}$  to  $6.5 \times 10^{-4}$  and from  $0.51 \times 10^{-4}$  to  $-0.75 \times 10^{-4}$ , respectively.

The conclusions for the medium wetness area (last column of Fig. 3) are very similar to those at the wet and dry locations: the soil moisture estimates in the two deep layers with the new method are better than those with the other methods, because the new method leads to a more smooth change of the covariance estimates (last column of Fig. 4).

### b. Sensitivity tests and discussions

#### 1) THE RELATIVE IMPORTANCE OF TWO CONSTRAINTS

Section 2a explains the reasons why the two constraints [i.e., Eqs. (6) and (8)] are needed, whereas section 3a shows how the new method improves the deep-layer soil moisture retrieval. Additional tests are done here to explore which constraint is more impor-

tant at the three different locations using one constraint at a time.

Results from these sensitivity tests show that Eq. (6) is often needed to constrain the covariance value at the wet location, especially after a heavy rain, whereas it is seldom used at the dry location. This might be related to the upper limit of soil moisture: at the wet location after a heavy rain, such as on days 51 and 52, near-surface soil moisture is almost saturated and highly skewed toward the wet end, whereas at the dry location, CLM soil moisture is not dry enough, so it does not skew toward the dry end.

In contrast to Eq. (6), Eq. (8) always makes a large positive contribution to the estimates of the deep-layer soil moisture at the dry location, and on average, the positive contribution gradually decreases from the dry location to the medium wetness location and then to the wet location. When soil becomes dry, its hydraulic conductivity is small and the downward propagation of near-surface soil moisture perturbations is slow. Therefore, at the dry location, properly taking the temporally correlated error into consideration is very important for the correct estimates of deep-layer soil moisture.

Comparison of the results using Eq. (6), Eq. (8), and both equations together also shows that, although Eq. (8)

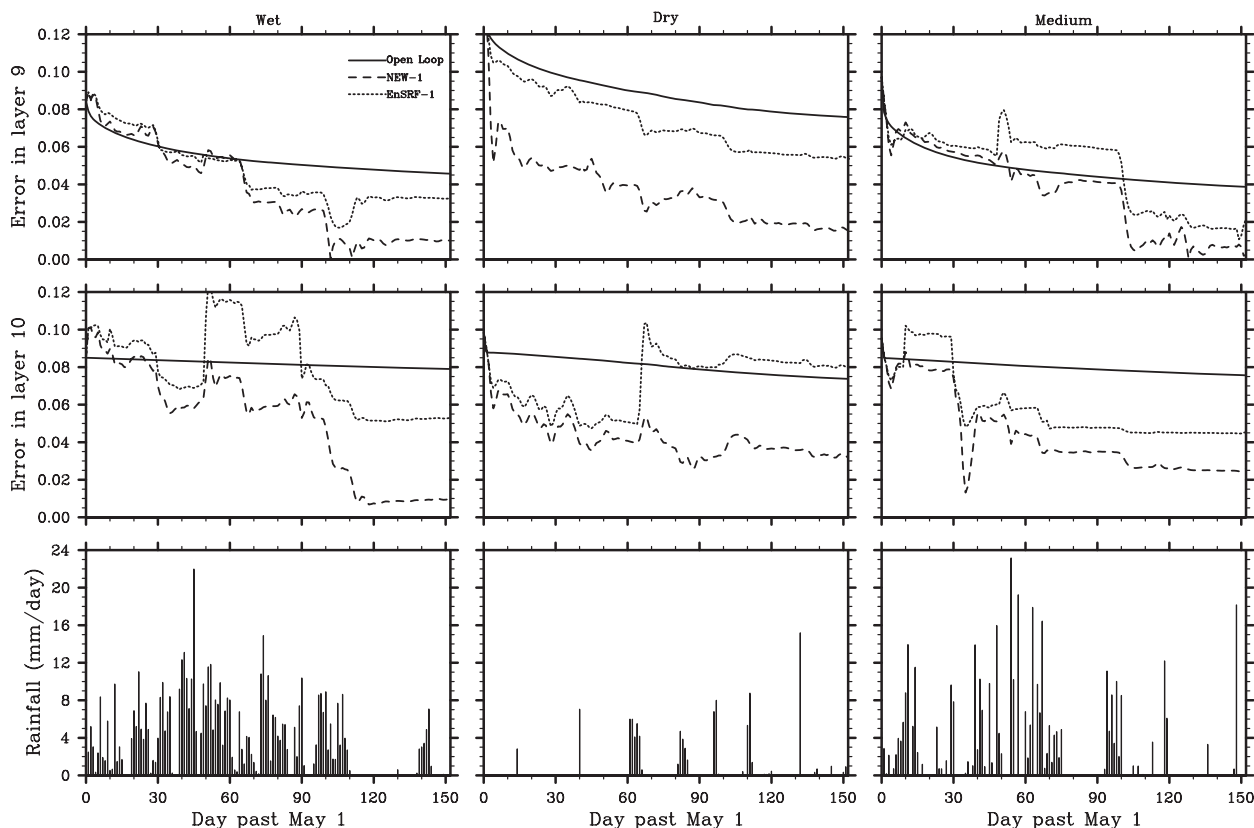


FIG. 3. Time series of the absolute errors of volumetric soil moisture for the ensemble mean using an updating frequency of once a day with the EnSRF-1, NEW-1, and Open Loop in (top) layer 9 and (middle) layer 10; (bottom) the rainfall time series is also shown. (left to right) The columns represent the results at the wet, dry, and medium wetness locations.

is the main reason for the improvement of deep-layer soil moisture estimates at the dry location, both equations are necessary for deep soil moisture estimates in the medium wetness and wet locations.

## 2) UPDATING FREQUENCY AND AMPLITUDE OF THE UNCERTAINTY

When the updating frequency is reduced to once every three days, the soil moisture profile retrievals are worse than those from assimilation once a day, and except at the dry location, there is no improvement of the two deep-layer soil moisture retrievals with the new method as compared with those from the EnSRF (Fig. 5). Further sensitivity tests show that Eq. (8) is the main reason for the retrieval improvement at the dry location, but it does not contribute to improving deep soil moisture at the medium wetness and wet locations. The reason for this might be that the useful information obtained at the last update almost disappears at the medium wetness and wet locations after 3-day model integration, so that constraint (8) is not needed. As mentioned earlier, when soil becomes dry, the downward propagation of near-surface soil moisture perturbations is slow. Thus, the useful in-

formation from the last update is not completely lost, so Eq. (8) is still slightly helpful at the dry location.

In contrast to Eq. (8), Eq. (6) does not contribute to the improvement of the deep-layer soil moisture retrieval at any of the three locations with an update of once every three days. The reason for this might be that the model itself can adjust the spatial inconsistency of the error correlations (resulting from the last update) during the 3-day model integration, so that constraint (6) is not needed.

Additional sensitivity tests have also been done with reduced amplitudes of uncertainties in precipitation, soil moisture observation, and soil texture. These factors do not change the earlier mentioned conclusions, even though reduced uncertainties slightly improve the estimates of the soil moisture profile (not shown here).

## 4. Conclusions

As demonstrated in the past, soil moisture in the top layers (e.g., within the top 1-m depth) can be retrieved based on the assimilation of near-surface soil moisture once a day using the ensemble-based Kalman filter (or

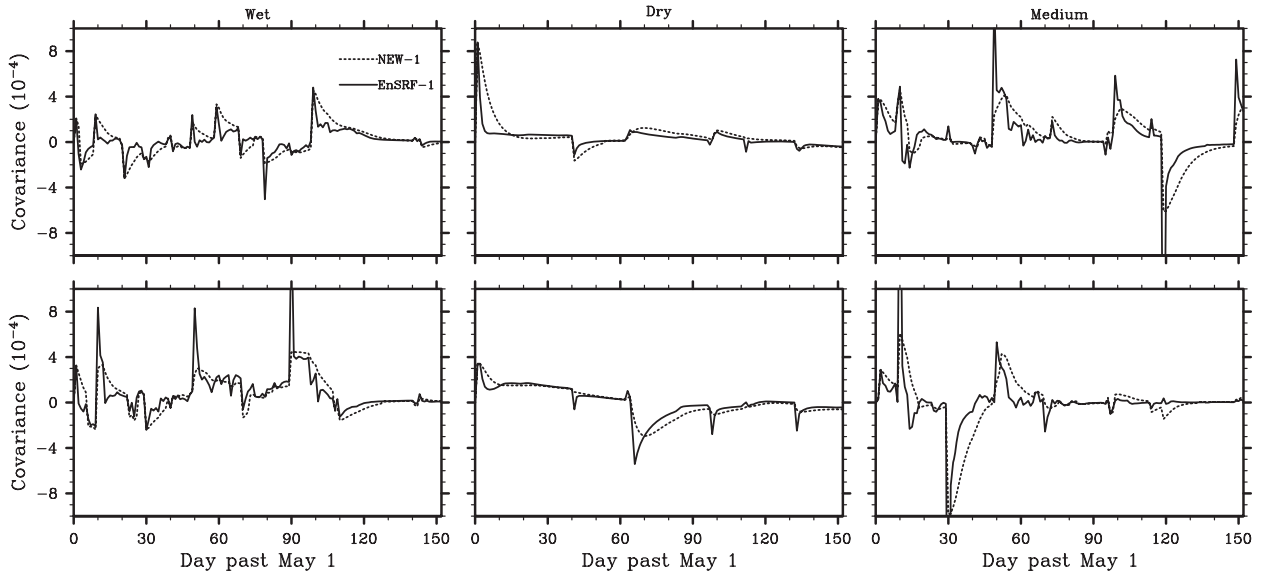


FIG. 4. Same as Fig. 3 but for the background error covariance estimates from the first two methods with an updating frequency of once a day. The covariance estimate is not needed for the open-loop simulation.

other assimilation methods). However, from literature searches, no studies have been reported on retrieving soil moisture at depths below 1 m. The relatively low error correlation between the surface and the deep layer, the limited range of soil moisture with the high skewness toward the wet or dry end, the temporally correlated soil moisture error between the surface and bottom layers, and the effect due to the nonlinearity of the system all make it difficult to estimate the deep-layer soil moisture. To reduce these negative effects, the new

method is proposed in this study, and the results from the OSSEs show that, compared with the standard ensemble square root filter (EnSRF; Whitaker and Hamill 2002), the new method significantly improves the estimation of deep-layer soil moisture when the updating frequency is once a day.

Sensitivity tests show that 1) Eq. (6) (i.e., the soil moisture error correlation should not increase with the increase of the vertical distance between the two layers) can effectively constrain the large change of the ensemble-based

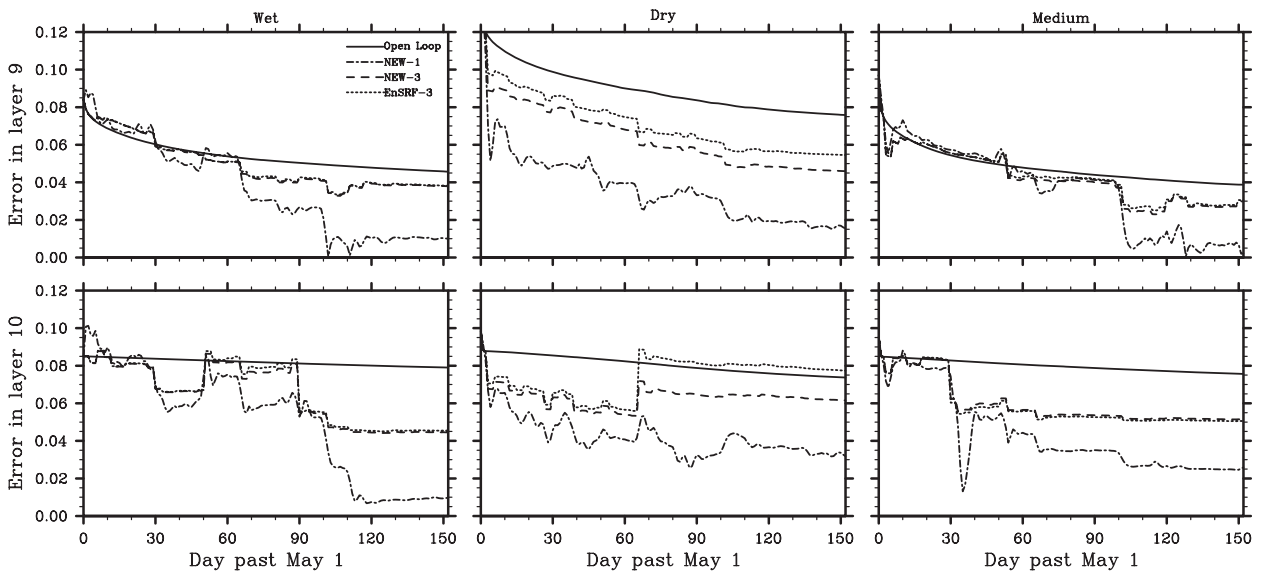


FIG. 5. Same as Fig. 3 but using an updating frequency of once every three days with the EnSRF (EnSRF-3) and the new method (NEW-3). For comparison, the results from the new method with an update of once a day (NEW-1) are also plotted.



Kalman filter error covariance when soil moisture skews toward the wet or dry end and 2) Eq. (8) (i.e., explicitly using the error covariance at the previous time step) always makes a positive contribution to the correct estimates of the deep-layer soil moisture. Although Eq. (8) is the main reason for the improvement of deep-layer soil moisture estimates at the dry location, Eqs. (6) and (8) are both necessary for deep soil moisture estimates in the medium wetness and wet locations.

The new method is sensitive to the updating frequency. For example, when the updating frequency is reduced to once every three days, the improvement of deep-layer soil moisture retrieval from the new method is minimal at the medium wetness and wet locations compared with the EnSRF. Therefore, alternative methods still need to be developed to improve the soil moisture retrieval with an updating frequency of once every three days. It is also worth noting that the results here are based on a particular OSSE setup using the CLM in which the soil moisture-based Richards equation is used, so the new method needs further tests using different land surface models and different observational data. For instance, recent studies (Oleson et al. 2008; Zeng and Decker 2009) have demonstrated that soil is generally too dry (particularly in deep layers) in CLM3 (e.g., as indicated by the initial soil moisture profile in Fig. 2), and different approaches have been proposed to resolve this model deficiency in the new version (i.e., CLM3.5; Oleson et al. 2008; Decker and Zeng 2009). Future study is still needed to see how CLM3.5 (or its future version, CLM4) would affect the conclusions of this study.

*Acknowledgments.* This work was supported by the National Natural Science Foundation of China (Grants 40775065 and 40475012), NASA (Grant NNG06GA24G), and NSF (Grant ATM0634762). The three anonymous reviewers are thanked for their insightful comments and suggestions.

#### REFERENCES

- Anderson, J. L., 2001: An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Bishop, C. H., B. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Crow, W. T., and E. F. Wood, 2003: The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using Ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. *Adv. Water Resour.*, **26**, 137–149.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- Decker, M., and X. Zeng, 2009: Impact of modified Richards equation on global soil moisture simulation in the Community Land Model (CLM3.5). *J. Adv. Model. Earth Syst.*, **1**, doi:10.3894/JAMES.2009.1.5.
- Entekhabi, D., H. Nakamura, and E. G. Njoku, 1994: Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multi-frequency remotely sensed observations. *IEEE Trans. Geosci. Remote Sens.*, **32**, 438–448.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear QG model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162.
- Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.
- Hoeben, R., and P. A. Troch, 2000: Assimilation of active microwave observation data for soil moisture profile estimation. *Water Resour. Res.*, **36**, 2805–2819.
- Koster, R. D., and Coauthors, 2004: Realistic initialization of land surface states: Impacts on subseasonal forecast skill. *J. Hydrometeorol.*, **5**, 1049–1063.
- Oleson, K. W., and Coauthors, 2004: Technical description of the community land model (CLM). NCAR Tech. Note NCAR/TN-461+STR, 186 pp. [Available online at [http://www.cgd.ucar.edu/tss/clm/distribution/clm3.0/TechNote/CLM\\_Tech\\_Note.pdf](http://www.cgd.ucar.edu/tss/clm/distribution/clm3.0/TechNote/CLM_Tech_Note.pdf).]
- , and Coauthors, 2008: Improvements to the Community Land Model and their impact on the hydrological cycle. *J. Geophys. Res.*, **113**, G01021, doi:10.1029/2007JG000563.
- Qian, T., A. Dai, K. E. Trenberth, and K. W. Oleson, 2006: Simulation of global land surface conditions from 1948 to 2004. Part I: Forcing data and evaluations. *J. Hydrometeorol.*, **7**, 953–975.
- Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, 2002: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 103–114.
- Sabater, J. M., L. Jarlan, J.-C. Calvet, F. Bouyssel, and P. De Rosnay, 2007: From near-surface to root-zone soil moisture using different assimilation techniques. *J. Hydrometeorol.*, **8**, 194–206.
- Walker, J. P., G. R. Willgoose, and J. D. Kalma, 2001: One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: A comparison of retrieval algorithms. *Adv. Water Resour.*, **24**, 631–650.
- Whitaker, J., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- Xu, Q., L. Wei, H. Lu, C. Qiu, and Q. Zhao, 2008: Time-expanded sampling for ensemble-based filters: Assimilation experiments with a shallow-water equation model. *J. Geophys. Res.*, **113**, D02114, doi:10.1029/2007JD008624.
- Zeng, X., and M. Decker, 2009: Improving the numerical solution of soil moisture-based Richards equation for land models with a deep or shallow water table. *J. Hydrometeorol.*, **10**, 308–319.
- , M. Shaikh, Y. Dai, R. E. Dickinson, and R. Myneni, 2002: Coupling of the Common Land Model to the NCAR Community Climate Model. *J. Climate*, **15**, 1832–1854.
- Zhang, S. W., H. R. Li, W. D. Zhang, C. J. Qiu, and X. Li, 2005: Estimating soil moisture profile by assimilation of near-surface observations with the ensemble Kalman filter (EnKF). *Adv. Atmos. Sci.*, **22**, 936–945.
- Zhou, Y. H., D. McLaughlin, and D. Entekhabi, 2006: Assessing the performance of the ensemble Kalman filter for land surface data assimilation. *Mon. Wea. Rev.*, **134**, 2128–2142.