ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 22, NO. 6, 2006, 936–945

Estimating the Soil Moisture Profile by Assimilating Near-Surface Observations with the Ensemble Kalman Filter (EnKF)

ZHANG Shuwen*[1,2] (张述文), LI Haorui[2] (李昊睿), ZHANG Weidong[3] (张卫东), QIU Chongjian[1,2] (邱崇践), and LI Xin[4] (李 新)

[1]*School of Atmospheric Sciences, Lanzhou University, Lanzhou* 730000

[2]*Key Laboratory of Arid Climatic Changing and Reducing Disaster of Gansu Province, Lanzhou* 730020

[3]*School of Physical Sciences and Technology, Lanzhou University, Lanzhou* 730000

[4]*Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou* 730000

(Received 24 January 2005; revised 29 June 2005)

ABSTRACT

The paper investigates the ability to retrieve the true soil moisture profile by assimilating near-surface soil moisture into a soil moisture model with an ensemble Kalman filter (EnKF) assimilation scheme, including the effect of ensemble size, update interval and nonlinearities in the profile retrieval, the required time for full retrieval of the soil moisture profiles, and the possible influence of the depth of the soil moisture observation. These questions are addressed by a desktop study using synthetic data. The "true" soil moisture profiles are generated from the soil moisture model under the boundary condition of 0.5 cm d^{-1} evaporation. To test the assimilation schemes, the model is initialized with a poor initial guess of the soil moisture profile, and different ensemble sizes are tested showing that an ensemble of 40 members is enough to represent the covariance of the model forecasts. Also compared are the results with those from the direct insertion assimilation scheme, showing that the EnKF is superior to the direct insertion assimilation scheme, for hourly observations, with retrieval of the soil moisture profile being achieved in 16 h as compared to 12 days or more. For daily observations, the true soil moisture profile is achieved in about 15 days with the EnKF, but it is impossible to approximate the true moisture within 18 days by using direct insertion. It is also found that observation depth does not have a significant effect on profile retrieval time for the EnKF. The nonlinearities have some negative influence on the optimal estimates of soil moisture profile but not very seriously.

Key words: soil moisture, ensemble Kalman filter, insertion, land data assimilation

1. Introduction

The role of soil moisture in the root zone is widely recognized as a key parameter in meteorology, hydrology and agriculture (Yeh et al., 1984; Koster et al., 2000). Adequate knowledge of the soil moisture is necessary to the understanding and prediction of the reciprocal influences between climate, weather and land surface process (Ma et al., 2000; Guo and Wang, 2003). Soil moisture can be obtained from point measurements, hydrological models and remote sensing techniques, with each having its own advantages and disadvantages (Schmugge et al., 1980; Zhang et al., 2004). The point measurement method is accurate but the representativeness of the spatial distribution is very poor due to the large spatial and temporal variability of soil moisture. A hydrological model may calculate the spatial distribution and temporal evolution of soil moisture but the results generally deviate from the true soil moisture distribution with the time integrations because of uncertainties in the forcing atmospheric data, the nonlinear nature of land-atmosphere interactions, and the high inhomogeniety of soil properties, vegetation and precipitation. As for remote sensing methods, microwave techniques have been widely used to infer soil moisture because the

*E-mail: zhangsw@lzu.edu.cn

atmosphere and clouds are relatively transparent to radiation in the low microwave frequencies. In recent years, some different bands of passive microwave radiometry have been put into orbit (for example, SSM/I, TMI and AMSR-E) and can frequently provide information on the soil moisture over a large scale. One drawback of microwave sensors is that they only provide soil moisture information in the top few centimeters (Njoku and Entekhabi, 1995). As we know, the top soil moisture changes very quickly because of evaporation and rainfall while the land-atmosphere interaction processes generally depend on the soil moisture profile to depths considerably larger than a few centimeters. How to fully use the information from the different channels and give accurate estimates of the soil moisture profile is a difficult inverse problem. Modern data assimilation theory provides a possible solution for the problems by merging the above three methods (Errico, 1999; Errico et al., 2000). Different assimilation algorithms have been used for retrieval of the soil moisture profile based on the near-surface soil moisture observations (Houser et al., 1998; Reichle et al., 2001; Crow and Wood, 2003). Among them, Entekhabi et al. (1994) first used the extended Kalman filter (EKF) technique to retrieve a 1-m soil moisture profile with simulated data and compared the results with both the true profile and those from the open loop. Walker et al. (2001) compared two assimilation schemes, viz. the direct insertion and EKF, in the context of retrieval rates. For nonlinear dynamics, the EKF requires the linearizations of the models and measurement operators, which may result in the failure of the EKF scheme. The ensemble Kalman filter (EnKF) (Evensen, 1994) is an alternative to the EKF and can directly calculate the state error covariance matrix by propagating an ensemble of states from which the required covariance information is obtained at the time of the update. Thus, it relatively easy to cope with models and measurement equations including thresholds and other nonlinearities. Because of these, it has been widely applied to oceanographic and meteorological problems (Houtekamer and Mitchel, 1998). On the contrary, in the field of land process, only a few studies have been carried out. Because the land surface models typically have different characteristics (e.g. dissipative in nature) with the atmospheric movements, we need to further investigate the soil moisture ensemble Kalman filtering problem. Reichle et al. (2002a) first applied the EnKF to the retrieval problem of soil moisture distributions by assimilating synthetic surface brightness temperatures into a land surface model. Later, Crow and Wood (2003) extended the EnKF methodology to a real-data case based on the electronically scanned thinned array radiometer (ESTAR) measurements during the 1997 Southern Great Plains Hydrology Experiment (SGP97).

To assess the performance of the EKF and EnKF for soil moisture estimation, Reichle et al. (2002b) conducted a twin experiment with their results showing that the EnKF is a robust and promising approach due to its flexibility in covariance modeling, so the EnKF is slightly superior to the EKF in the context of performance. In terms of accuracy, the ensemble filter may be more accurate than the EKF since covariances are calculated by propagating model states with a fully nonlinear model rather than using the assumption of linearity (Hamill and Whitaker, 2001). After comparing the performance of the EnKF to a weak-constraint variational algorithm (a kind of optimal smoother), Reichle et al. (2002a) indicated that the EnKF is a flexible and robust data assimilation option with no need to compute adjoint models or derivatives that can handle a wide range of model errors, but it is too early to make a definitive comparison between the two methods. Anderson (2001) has also done some preliminary experiments that suggest that an ensemble-based approach may also be better than 4DVAR (Four dimensional variational analysis). Based on the above considerations, this paper continues to use the EnKF to investigate the ability to retrieve the 1-dimensional soil moisture profile by assimilating near-surface soil moisture into the soil moisture model with focus on the required time for full retrieval of the soil moisture profiles, the effect of the update interval, ensemble size and nonlinearities on the profile retrieval, and the possible influence of the depth of the soil moisture observations.

Since we mainly focus on methodology, the algorithm is tested in a desktop study using synthetic data. Synthetic datasets are generated using the same soil moisture model used to retrieve the soil moisture profile from surface observations. As explained by Walker et al. (2001), using synthetic datasets can eliminate experimental errors in measuring the soil moisture profile, as well as in estimating the soil properties. Furthermore, using the same model for the generation and retrieval of profile data can eliminate model errors due to the neglect of thermally induced moisture transport. Beginning with an intentionally poor initial guess of the soil moisture profile, the data assimilation schemes are used to retrieve the full profiles with hourly and daily updates, respectively. In order to demonstrate the ability of the EnKF data assimilation scheme to extract the information content of the observations, the retrieved profile is compared not only with the true profile but also with that from the direct insertion assimilation scheme because of its simplicity and easy implementation.

2. Soil moisture equation

The flow of soil moisture in a vertical soil column is a very complicated process. For simplicity, almost all land surface models ignore the effect of thermal vapor movement so the water flux is only driven by hydraulic gradient, thus the resulting equation is a one-dimensional Richards' equation. If we retrieve both the soil moisture and soil temperature profile at the same time based on observed surface radiobrightness (or surface soil moisture) and surface temperature (or infrared surface temperature) (Entekhabi et al., 1994; Walker et al., 2001), or based only on observed radiobrightness (or surface soil moisture) (Galantowitz et al., 1999; Li and Islam, 1999), we should consider the coupled flow of heat and moisture. However, if we directly assimilate the surface moisture observation and at the same time specify the evaporation at the surface, the heat transfer processes will have no influence on the soil moisture flux in the simplified land surface models. As explained in the introduction, we mainly concentrate on the assimilation algorithm so, for simplicity, we choose not to include the soil heat transfer equation for extracting the soil temperature profile in our test. As we know, Richards' equation has three forms: the soil water matric potential based or ψ-based form, the moisture content based or θ-based form, and the mixed form. No exact solutions of Richards' equation for general boundary and initial conditions are known so numerical methods are used. According to the study of Celia et al. (1990), we adopt the mixed form of Richards' equation which has the following form:

$$\frac{\partial\theta}{\partial t}=\frac{\partial}{\partial z}\left[k(\psi)\frac{\partial\psi}{\partial z}\right]+\frac{\partial k(\psi)}{\partial z}\ , \tag{1}$$

where θ is the volumetric moisture content, ψ is the soil water matric potential, $k(\psi)$ is the unsaturated hydraulic conductivity and z is the vertical coordinate (distance from the surface, positive downward). The Clapp and Hornberger relationships are used to express the dependences of $k(\psi)$ and ψ on θ:

$$k=k_{\rm sat}\left(\frac{\theta}{\theta_{\rm sat}}\right)^{2b+3}\ , \tag{2}$$

$$\psi=\psi_{\rm sat}\left(\frac{\theta}{\theta_{\rm sat}}\right)^{-b}\ . \tag{3}$$

Here, $\theta_{\rm sat}$ and $\psi_{\rm sat}$ are the values at saturation of soil water content and matric potential, respectively, and b is a non-dimensional exponent.

The numerical solution of Eq. (1) is based on the finite volume method (Versteeg and Malalasekera, 1998). Each node is surrounded by a control volume whose boundaries are located in the middle of two adjacent nodal points. The soil moisture at each node represents the averaged value over its control volume and the water fluxes are at the boundaries of the control volume. Accordingly, the soil moisture equation (1) is discretized into the following form:

$$\frac{\Delta z_i}{\Delta t}\Delta\theta_i=q_{i-1,n+1}-q_{i,n+1}\ , \tag{4}$$

where Δz_i is the thickness of layer $i, \Delta t$ is the constant time step of integration, $\Delta\theta_i=\theta_{i,n+1}-\theta_{i,n}$, the suffix n identifies variables at time $n\Delta t$, and q_i is the flux of water at the interface between layer i and $i+1$ (positive downward), defined as

$$q_i=-k_{i+1/2}\left(\frac{\psi_{i+1}-\psi_i}{z_{i+1}-z_i}-1\right)\ . \tag{5}$$

The boundary conditions are $q_0=-q_{\rm seva}$ evaporation for the first layer and q_m=0 for the bottom soil layer $m.\psi$ and k are nonlinear functions of θ so that $q=f(\theta_i,\theta_{i+1})$. With $q_{i,n+1}$ approximated as

$$q_{i,n+1}=q_{i,n}+\frac{\partial q_{i,n}}{\partial\theta_i}\Delta\theta_i+\frac{\partial q_{i,n}}{\partial\theta_{i+1}}\Delta\theta_{i+1}\ . \tag{6}$$

The water balance for the i-th layer is

$$-\frac{\partial q_{i-1,n}}{\partial\theta_{i-1}}\Delta\theta_{i-1}+\left[\frac{\Delta z_i}{\Delta t}-\frac{\partial q_{i-1,n}}{\partial\theta_i}+\frac{\partial q_{i,n}}{\partial\theta_i}\right]\Delta\theta_i + \frac{\partial q_{i,n}}{\partial\theta_{i+1}}\Delta\theta_{i+1}=q_{i-1,n}-q_{i,n}\ , \tag{7}$$

which is a tridiagonal system of equations for $\Delta\theta$ (Bonan, 1996). The water balance for the first soil layer (i=1) is

$$\left(\frac{\Delta z_1}{\Delta t}-\frac{\partial q_{0,n}}{\partial\theta_1}+\frac{\partial q_{1,n}}{\partial\theta_1}\right)\Delta\theta_1+\frac{\partial q_{1,n}}{\partial\theta_1}\Delta\theta_2=q_{0,n}-q_{1,n}\ . \tag{8}$$

For the bottom layer ($i=m$), the water balance is

$$-\frac{\partial q_{m-1,n}}{\partial\theta_{m-1}}\Delta\theta_{m-1}+\left(\frac{\Delta z_m}{\Delta t}-\frac{\partial q_{m-1,n}}{\partial\theta_m}+\frac{\partial q_{m,n}}{\partial\theta_m}\right)\Delta\theta_m = q_{m-1,n}-q_{m,n}\ . \tag{9}$$

Denote by $\boldsymbol{\theta}_{n,f}$ the predicted vector composed of soil moistures from all m layers at $t=n\Delta t$. When the analysis of the soil moisture profile at the previous time $t=(n-1)\Delta t$, denoted by $\boldsymbol{\theta}_{n,a}$, is known, based on Eqs. (7), (8) and (9), $\boldsymbol{\theta}_{n,f}$ is calculated in the following vector form:

$$\boldsymbol{\theta}_{n,f}=\boldsymbol{F}_n(\boldsymbol{\theta}_{n,a})+q_n\ , \tag{10}$$

where $\boldsymbol{F}_n$ is the model operator and $\boldsymbol{q}_n$ the summation of errors in the model formulation and the forcing data.

3. Retrieval algorithms

3.1 *The Ensemble Kalman filter*

The Kalman filter (KF) is a statistical assimilation method that updates all the predicted variables at the same time based on the relative magnitudes of the covariances of both the predicted model variables and the observations. For nonlinear applications, a linearized model operator, known as the extended Kalman filter (EKF), is needed to calculate the evolution of the uncertainties in the state estimates. The EnKF is an alternative to the EKF and can circumvent the expensive integration of the state error covariance matrix in the EKF. It is also easy to implement even if the models and measurement operator include thresholds and nonlinearities in contrast to the adjoint-based 4DVAR.

The EnKF needs to be initialized with the input of an ensemble of initial condition fields $\boldsymbol{\theta}_{0,i,a}(i = 1,\ldots,N)$ with mean $\boldsymbol{\theta}_{0,a}$ and covariance $\boldsymbol{P}_{0,a}$, and then each ensemble member is integrated by using a corresponding ensemble of N random realizations of error fields $\boldsymbol{q}_{n,i}$:

$$\boldsymbol{\theta}_{n,i,f} = \boldsymbol{F}_n(\boldsymbol{\theta}_{n,i,a}) + \boldsymbol{q}_{n,i}\ , \quad i = 1,\ldots,N \tag{11}$$

where $\boldsymbol{\theta}_{n,i,f}$ and $\theta_{n,i,a}$ refer to the state estimates of the i-th ensemble member before and after the update, respectively. The error covariance matrix for the forecast estimate, $\boldsymbol{P}_{n,f}$, is approximated by

$$\boldsymbol{P}_{n,f} = \overline{(\boldsymbol{\theta}_{n,i,f} - \boldsymbol{\theta}_{n,f})(\boldsymbol{\theta}_{n,i,f} - \boldsymbol{\theta}_{n,f})^{\mathrm{T}}} \tag{12}$$

where the overbar denotes an expectation value and $\boldsymbol{\theta}_{n,f}$ is the ensemble mean:

$$\boldsymbol{\theta}_{n,f} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\theta}_{n,i,f}\ . \tag{13}$$

If we assemble all observations taken at time $t = n\Delta t$ into the measurement vector $\boldsymbol{d}_n$, the measurement process can be written as

$$\boldsymbol{d}_n = \boldsymbol{H}_n\boldsymbol{\theta}_{n,i,f} + \varepsilon_{n,i}\ , \tag{14}$$

where a linear relationship is assumed and $\varepsilon_{n,i}$ is an ensemble of perturbed observations with a mean equal to zero and a covariance equal to the covariance for measurements, $\boldsymbol{R}_n$ (Burgers et al., 1998). If observations are available at time $t = n\Delta t$, we update each ensemble member by using a linear combination of forecast model states and the observation:

$$\boldsymbol{\theta}_{n,i,a} = \boldsymbol{\theta}_{n,i,f} + \boldsymbol{K}_n(\boldsymbol{d}_n - \boldsymbol{H}_n\boldsymbol{\theta}_{n,i,f} - \varepsilon_{n,i})\ , \quad i = 1,\ldots,N \tag{15}$$

where the Kalman gain is equal to

$$\boldsymbol{K}_n = \boldsymbol{P}_{n,f}\boldsymbol{H}_n^{\mathrm{T}}(\boldsymbol{H}_n\boldsymbol{P}_{n,f}\boldsymbol{H}_n^{\mathrm{T}} + \boldsymbol{R}_n)^{-1}\ . \tag{16}$$

In the EnKF, the state error covariance is never explicitly needed, but parts or all of it can be estimated at any time from the ensemble.

3.2 *Direct insertion*

The direct insertion data assimilation is the simplest method in which the predicted soil moisture is directly substituted, i.e. a hard update, by the observed moistures at the same point without consideration of observational errors whenever the observation is available (Li and Islam, 1999; Walker et al., 2001), so it is very easy to implement. Only through the physical infiltration and exfiltration processes can this surface information be transferred into the deeper layers, thus it will take a relatively long time for the deeper soil to "feel" the variation of soil moisture at the top layer.

4. Test and application

In order to explore the relative merits of the data assimilation methods, a case study is presented. The same soil moisture equation is used to generate and retrieve the profile data. The initial soil moisture condition is assumed to be 0.40 cm^3 cm^{-3} throughout the 1 m deep soil. The boundary condition is 0.5 cm d^{-1} evaporation at the surface and no water flux at the bottom. To test the assimilation schemes, we initialize the soil moisture model with an intentionally poor initial guess of 0.35 cm^3 cm^{-3} throughout the profile. A value of 5% variation in the system state is added into the model forecasts to represent the model errors (Walker et al., 2001). More details on the simulation environment are given in Table 1.

We know that the microwave observation data are not a predicted variable of the near-surface soil moisture but the brightness temperature, while the brightness temperature generally is a complicated function of the soil moisture, soil temperature, observation frequency, soil texture and soil bulk density (Njoku and Entekhabi, 1995; Owe et al., 2001). For simplicity and concentration on the retrieval algorithm, we do not adopt the radiobrightness estimation model but directly use the "true" soil moisture as observations for that time and depth. Because the thermally induced moisture transport is neglected and the evaporation at the surface is a fixed value, the heat transfer equation is not needed in our assimilation scheme. In general, microwave measurements of soil moisture are limited to the top few centimeters (less than 10 cm) of the soil column, but actual sensing depth will depend on the magnitude of the surface moisture and its shape. To investigate whether different observation depths have an influence on the profile retrieval or not, three representative observation depths (2, 6, and 10 cm) of soil moisture observation are chosen.

To initiate the assimilation scheme, an initial ensemble needs to be specified. Under ideal conditions for applications of the EnKF, the initial ensemble should be chosen to properly represent the error statistics of the initial guesses for the model state. However,

Table 1. Parameters and conditions used in the profile retrieval.

Soil parameters		Retrieval conditions	
Soil type	Clay loam	Depth	100 cm
Soil moisture at saturation	0.476 cm^3 cm^{-3}	Number of nodes	50
Hydraulic conductivity at saturation	25 cm d^{-1}	Boundary condition	0.5 cm d^{-1}
Exponent b	8.52	Initial condition	0.40 cm^3 cm^{-3}
Matric potential at saturation	−63.0 cm	Bad initial guess	0.35 cm^3 cm^{-3}

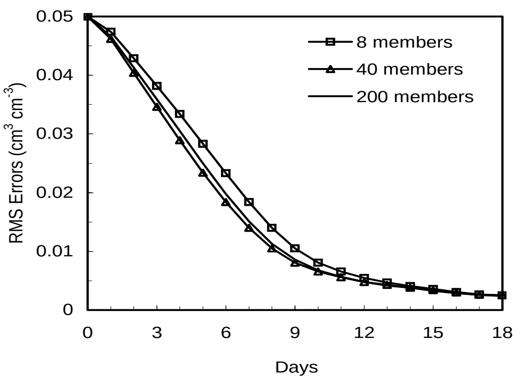

Fig. 1. RMS errors in EnKF soil moisture predictions for various ensemble sizes.

a modest mis-specification of the initial ensemble normally does not influence the results very much over a long time (Evensen, 2003). In this study, the generated initial ensemble of perturbed soil moisture satisfies a normally-distributed random field with mean zero and variance 0.05^2, and the spatial correlation function has a exponential form with a decorrelation length of 1 m. To prevent an updated ensemble with a variance that is too low, one should add perturbations with the correct statistics to the observations and generate an ensemble of observations that then is used in updating the ensemble of model states (Burgers et al., 1998). The added ensemble of perturbed observations is a normally-distributed random field with mean zero and variance 0.02^2.

In the EnKF, the required covariance information is calculated from an ensemble of states at the time of the update, so it is necessary to introduce enough ensemble members to obtain satisfactory estimates. From the theoretical viewpoint, the greater the number of members, the more accurate the estimated state error covariance, but in the real application of the method, it is impossible to use a very large ensemble size, so an appropriate size should be chosen. Figure 1 shows the time evolution of the actual rms errors for 8, 40, and 200 ensemble members by using the updates once a day. The time evolutions of the errors for 40 and 200 ensemble members are quite similar so an ensemble of 40 members should be large enough to estimate the state error covariance with reasonable accuracy. Based on the above consideration, the ensemble with 40 members is adopted in this study.

4.1 *Updating once every hour*

The EnKF assimilation scheme updates instantaneously the entire profile every hour, based on the relative magnitudes of the covariances of the model forecasts and the observations, because it is able to add or subtract mass from the system from more than just the node at the observational depth. Full retrieval of the soil moisture profile using the EnKF algorithm is shown in Fig. 2. As the observation depth increases, profile retrieval becomes relatively slow, but the differences between using different observation depths are very small. A similar result has been reported by Walker et al. (2001). Full soil moisture profile retrieval approximately requires about 16 hours of wall clock time.

For comparison, we plot the results from the direct insertion assimilation scheme in Fig. 3. The direct insertion assimilation scheme performs an instantaneous replacement of the model prediction and with the true soil moisture only at the observational position. For the different observation depths, the required times for the full retrieval of soil moisture profile are almost the same at about 12 days; that is to say, the observation depth has almost no effect on the profile time for the direct insertion assimilation scheme. In comparison with the EnKF assimilation algorithm in Fig. 2, full retrieval of the soil moisture profile using the direct insertion assimilation scheme is very slow.

4.2 *Updating once every day*

It is unrealistic that the soil moisture observation is taken once every hour for any practical application of profile retrieval. At best we may expect a repeat coverage of once a day. Thus, we continue to evaluate the ability of estimating the soil moisture profile but with a daily update. Figure 4 indicates that full retrieval of the soil moisture profile takes about 16, 15, and 14 days with the EnKF assimilation methods for the respective observation depths of 2 cm, 6 cm, and 10 cm. Because the direct insertion data assimilation is not able to track the true soil moisture profile within

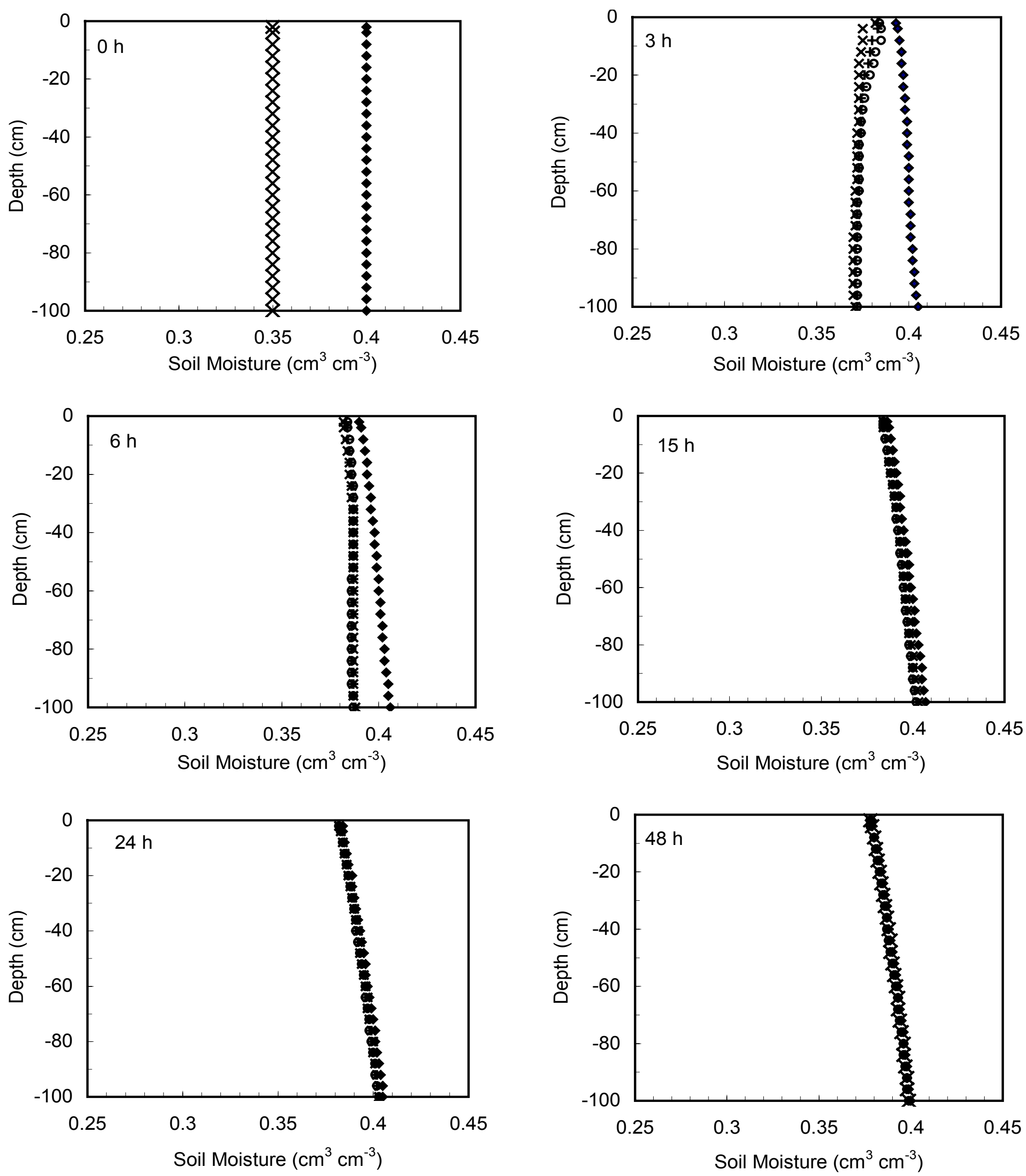

Fig. 2. Comparison of simulated soil moisture profiles using the EnKF for observation depths of 2(×), 6(+), and 10 cm (○) with the true values (♦) for hourly observations.

20 days when the predicted moisture is updated once a day, the retrieval results are not shown. It is obvious that the EnKF assimilation scheme is superior to the direct insertion assimilation scheme, however, it also takes a relatively long time to approximate the true soil moisture profile.

One interesting result is that the required update times for full retrieval of soil moisture profile are almost equal to those for updating once an hour, although the profile retrieval time length is completely different, i.e. one is achieved in about 15 days while the other is achieved in about 15 hours. Therefore, if we want to approximate the true soil moisture profile in a short time based solely on surface soil moisture conditions, the frequency of observation is very important and should not be too small. On the other hand, we should revise our method to accelerate the convergence speed for the retrieval of soil moisture profile.

4.3 *Nonlinearities and deviations from Gaussian distributions*

We know that the application of the EnKF does not explicitly require model linearity or Gaussian error statistics, but the Kalman filter will cease to be an

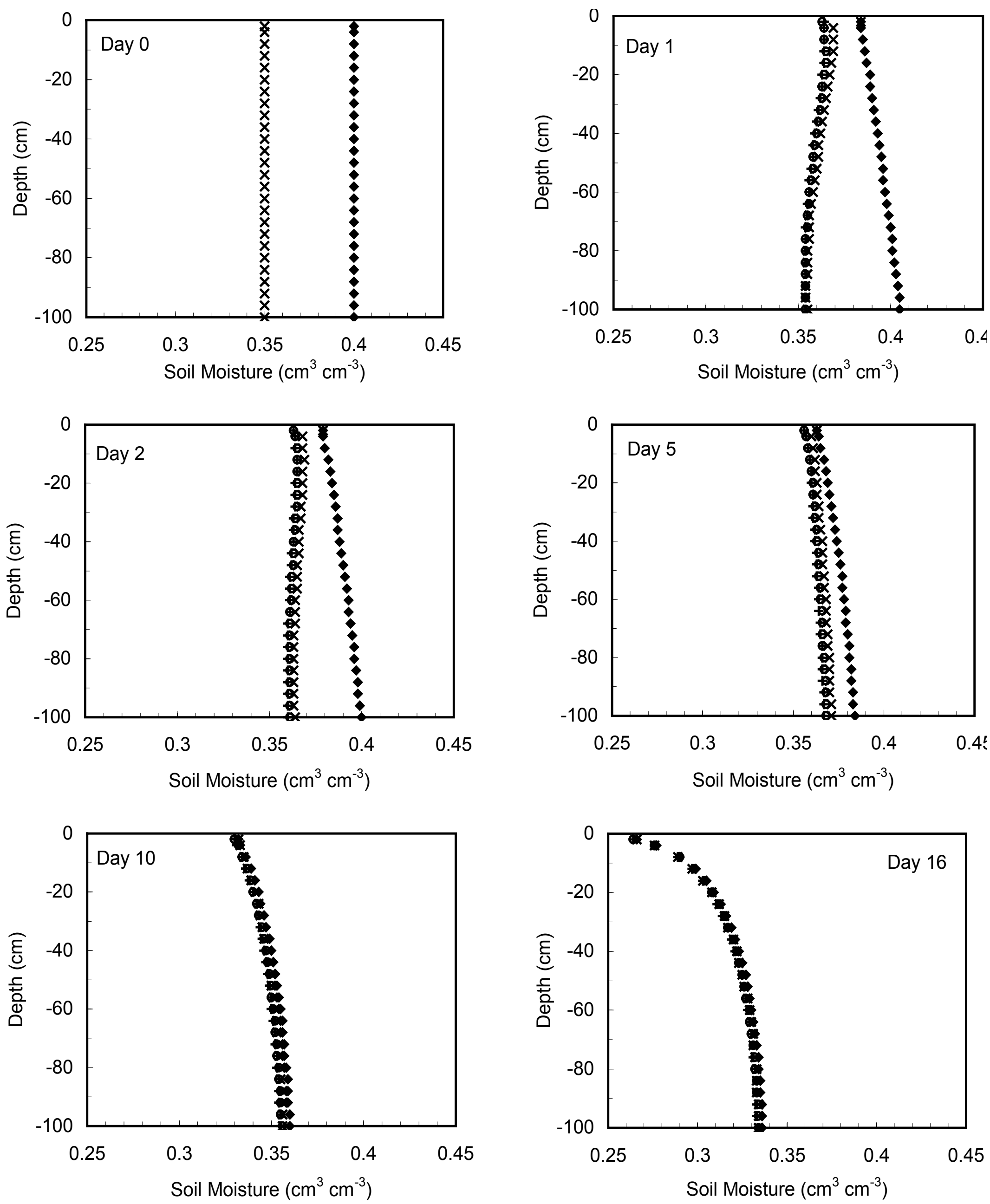

Fig. 3. Comparison of simulated soil moisture profiles using the direct insertion for observation depths of 2(×), 6(+), and 10 cm (○) with the true values (♦) for hourly observations.

optimal data assimilation if either condition is not met. Nonlinearities, whether differential or not, are likely to induce asymmetries in the sample distribution of the ensemble members (Reichle et al., 2002a). In order to test the asymmetries, we use the skewness coefficient defined as

$$s = E\{[\theta - E(\theta)]^3\}/\sigma_\theta^3 \,, \quad (17)$$

where $E(\cdot)$ is the expectation operation and σ_θ is the standard deviation of soil moisture members. In order to reduce the possible consequence of sampling error we use a large ensemble whose size is 200. Figure 5 shows the curves of skewness coefficient with time at three representative depths of 10, 50, and 100 cm by using the updating of the observation at the depth of 2 cm once a day. All initially guessed soil moistures have been set as Gaussian distributions, but with the updates the ensemble tends to be a little wetter so the sknewness coefficients at the three depths are slightly

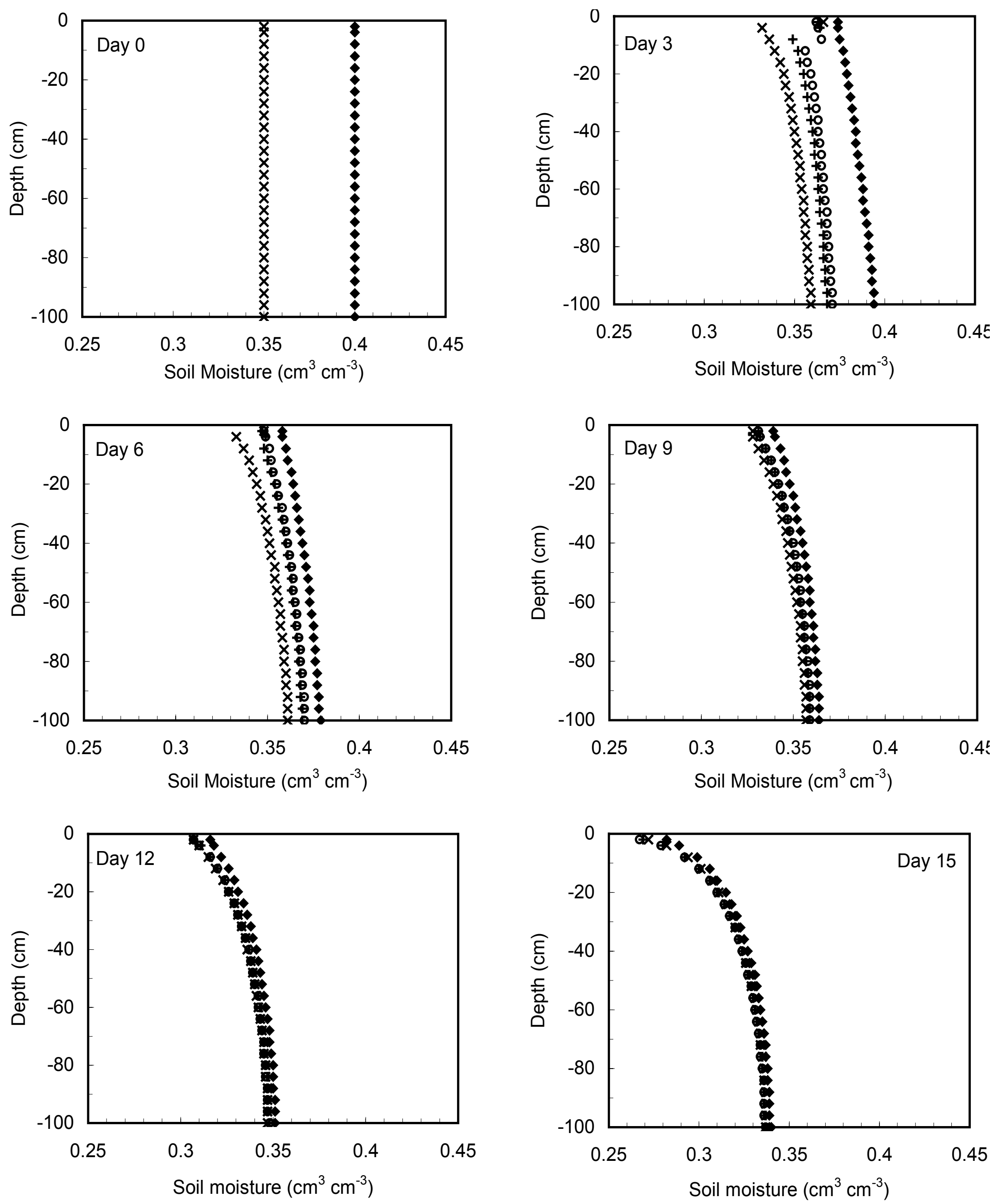

Fig. 4. Comparison of simulated soil moisture profiles using the EnKF for observation depths of 2(×), 6(+), and 10 cm (○) with the true values (♦) for daily observations.

increased. Fortunately, the skewness coefficients do not deviate from zero too much within the assimilation period so the EnKF can still extract spatial and temporal trends in the root-zone (with in 1-m depth) soil water content, however, it takes a relatively long time to acquire the true soil moisture profile. Besides the nonlinearities, the lower bound of soil moisture is another likely reason causing the non-Gaussian conditional probability distribution functions (PDFs).

Further insight on the significance of non-Gaussian behavior can be gained by looking at the distribution of soil moisture across the ensemble at three selected locations just before each update. Figure 6 shows the ensemble distribution of the soil moisture for three representative depths 10 cm, 50 cm, and 100 cm. In general, the figures appear to be reasonably symmetric, however, small differences between the ensemble median and ensemble mean at the three depths can be observed, which inform us of the existence of ensemble skewness as in Fig. 5.

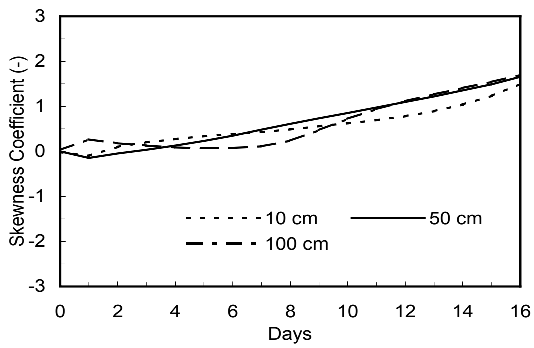

Fig. 5. Filter-derived skewness coefficients of 200 ensemble members for the three different depths.

5. Summary and Discussion

The results demonstrate that the EnKF is an effective hydrological data assimilation scheme for the estimation of soil moisture profile from the surface soil moisture observations with a very easy implementation. We also compare the performance of the EnKF to the direct insertion assimilation scheme and find that the EnKF is superior to the direct insertion assimilation scheme. The superiority of the ensemble Kalman filter lies in its ability to update the entire profile, while the direct insertion assimilation scheme only directly substitutes the measurement for the corresponding soil moisture. More ensemble members might be better, but in this study we find that an ensemble with 40 members is enough to represent the error covariance. If the update is carried out once a day, the direct insertion assimilation scheme is not able to realize full retrieval while the EnKF assimilation scheme can still track the true soil moisture profile. However, it will take a long time to acquire the true soil moisture profile, so we need to further revise the EnKF data assimilation scheme for its future application. This problem is under our consideration and the results from a new EnKF data assimilation scheme are promising. The observation depth does not have an obvious influence on the profile retrieval time in contrast to the great importance of repeat coverage frequency. Owing to nonlinearities in the soil moisture model and the bound ary of soil moisture, the required Gaussian distributions cannot be exactly met, but the influence is not very serious in this study because the skewness coefficients do not deviate from zero very seriously.

This synthetic study has shown that the soil moisture data at the near surface is very useful for correcting errors in the model forecast of soil moisture profile resulting from poor initialization. The results show the feasibility of using the EnKF data assimilation scheme to solve the inverse problem associated with soil moisture profile retrieval based on the near-surface observations, however, more research is needed to better understand the role of nonlinearities and the asymmetries in the conditional forecast probability density function. For example, when the soil is very wet or dry, the soil moisture PDF presents obvious skewness. In practical application, it will be difficult to select the proper ensemble size with good spatial and temporal correlation functions for expressing the model error fields also. All of these questions should be properly answered before the EnKF can be applied to the retrieval of a field soil moisture profile.

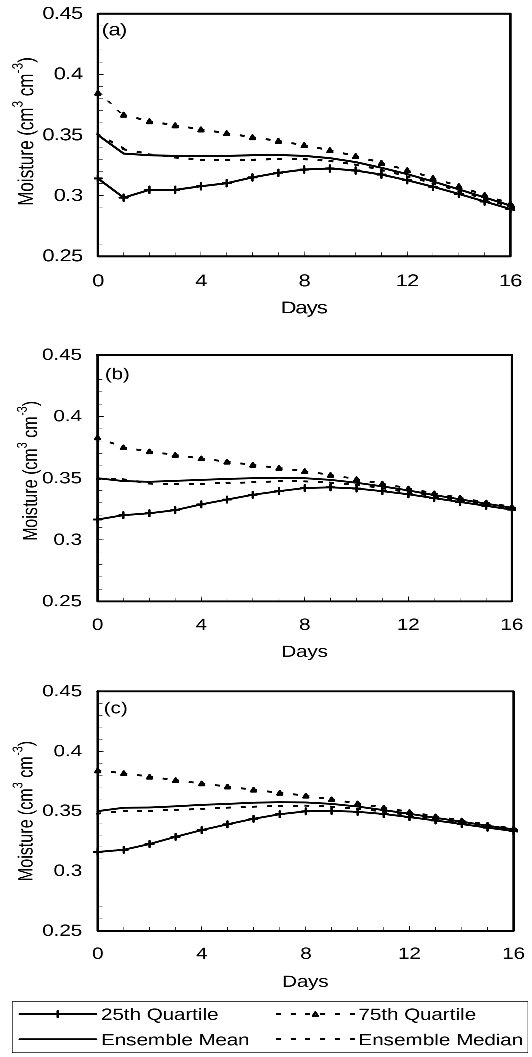

Fig. 6. Distribution of the ensemble members with updates (a) at the observational depth of 10 cm, (b) at the depth of 50 cm, and (c) at the depth of 100 cm.

Finally, if we directly assimilate the microwave brightness observations, the coupled heat and moisture transport models should be adopted. Moreover, the surface boundary condition is a complicated function of precipitation, wind speed, temperature, and specific humidity, and soil temperature and moisture etc. and it often contains large uncertainties, so we should also investigate the possible influence of the errors in the forcing data that have been ignored in this test.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant Nos. 40475012, 90202014, and 2001CB309404).

REFERENCES

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

Bonan, G. B., 1996: A land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: Technical description and user's guide. NCAR Tech. Note NCAR/TN-417+STR, 150pp.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman Filter. *Mon. Wea. Rev.*, **126**, 1719–1724.

Celia, M. A., E. T. Bouloutas, and R. L. Zarba, 1990: A general mass conservative numerical solution for the unsaturated flow equation. *Water Resour. Res.*, **26**, 1483–1496.

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. *Advances in Water Resources*, **26**, 137–149.

Entekhabi, D., H. Nakamura, and E. G. Njoku, 1994: Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. *IEEE Trans. Geosci. Remote Sens.*, **32**, 438–448.

Erric, R. M., 1999: Workshop on assimilation of satellite data. *Bull. Amer. Meteor. Soc.*, **80**, 463–471.

Erric, R. M., G. Ohring, J. Derber, and J. Hoiner, 2000: NOAA-NASA-DoD workshop on satellite data assimilation. *Bull. Amer. Meteor. Soc.*, **81**, 2457–2462.

Evensen, G., 1994: Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**(C5), 10143–10162.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367.

Galantowicz, J. F., D. Entekhabi, and E. G. Njoku, 1999: Test of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness. *IEEE Trans. Geosci. and Remote Sens.*, **37**(4), 1860–1870.

Guo Weidong, and Wang Huijun, 2003: A case study of the improvement of soil moisture initialization in IAP-PSSCA. *Adv. Atmos. Sci.*, **20**, 845–848.

Hamill, T. M., and J. S. Whitaker, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.

Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C. Goodrich, 1998: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resour. Res.*, **34**, 3405–3420.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 769–811.

Koster, R. D., M. J. Suarez, and M. Heiser, 2000:Variance and predictability of precipitation at seasonal to interannual timescales. *Journal of Hydrometeorolog.*, **1**, 26–46.

Li, J., and S. Islam, 1999: Estimation of soil moisture profile and surface fluxes partitioning from sequential assimilation of surface layer soil moisture. *J. Hydrol.*, **220**, 86–103.

Ma Zhuguo, Wei Helin, and Fu Congbin, 2000: Relationship between regional soil moisture variation and climatic variability over East China. *Chinese J. Atmos. Sci.*, **58**, 278–287.

Njoku, E. G., and D. Entekhabi, 1995: Passive microwave remote sensing of soil moisture. *J. Hydrol.*, **184**, 101–130.

Owe, M., R. de Jeu, and J. Walker, 2001: A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Geosci. Remote Sens.*, **39**, 1643–1653.

Reichle, R. H., D. Entekhabi, and D. B. McLaughlin, 2001: Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach. *Water Resour. Res.*, **37**, 2353–2364.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, 2002a: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 103–114.

Reichle, R. H., J. P. Walker, R. D. Koster, and P. R. Houser, 2002b: Extended versus ensemble Kalman filtering for land data assimilation. *Journal of Hydrometeorology*, **3**, 728–740.

Schmugge, T. J., T. J. Jackson, H. L. McKim, 1980: Survey of methods for soil moisture determination. *Water Resour. Res.*, **16**, 961–979.

Versteeg, H. K., and W. Malalasekera, 1998: *An Introduction to Computational Fluid Dynamics*. Longman Group, 257pp.

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. *Advances in Water Resources*, **24**, 631–650.

Yeh, T. C., R. I. Wetherald, and S. Manabe, 1984: The effect of soil moisture on the short term climate and hydrology change—A numerical experiment. *Mon. Wea. Rev.*, **112**, 474–490.

Zhang, S.-W., C.-J. Qiu, and Q. Xu, 2004: Estimating soil water contents from soil temperature measurements by using an adaptive Kalman filter. *J. Appl. Meteor.*, **43**, 379–389.