JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 113, D14127, doi:10.1029/2007JD009650, 2008

A land surface soil moisture data assimilation system based on the dual-UKF method and the Community Land Model

Xiangjun Tian,[1] Zhenghui Xie,[1] and Aiguo Dai[2]

Received 28 November 2007; revised 10 April 2008; accepted 18 June 2008; published 29 July 2008.

[1] Many studies have shown the deficiencies of the extended Kalman filter (EKF), even though it has become a standard technique used in nonlinear estimation. In the EKF method, the state distribution is propagated analytically through the first-order linearization of the nonlinear system, which can introduce large errors in variable estimation and may lead to suboptimal performance and sometimes divergence of the filter. The unscented Kalman filter (UKF) addresses these problems using a deterministic sampling approach to capture the posterior mean and covariance accurate to the third order for any nonlinearity, while the dual-UKF method uses two UKF filters (one for state variables and one for parameters, in contrast to only one filter in the usual UKF) to simultaneously optimize the model states and parameters using observational data. In this paper, we employ the dual-UKF method to account for the effects of land surface subgrid-scale heterogeneity and soil water thawing and freezing and implement it into the NCAR Community Land Model version 2.0 to build a data assimilation system for assimilating satellite observations of soil moisture. Experiments for two sites in north and south China show that this dual-UKF-based assimilation system outperforms the usual UKF- and EKF-based methods in reproducing the temporal evolution of daily soil moisture, especially under freezing conditions. Furthermore, the improvement also propagates, albeit to a lesser extent, to lower layers where observations are unavailable.

Citation: Tian, X., Z. Xie, and A. Dai (2008), A land surface soil moisture data assimilation system based on the dual-UKF method and the Community Land Model, *J. Geophys. Res.*, *113*, D14127, doi:10.1029/2007JD009650.

1. Introduction

[2] Soil moisture plays a vital role in land-atmosphere interactions. It is a key state variable for many hydrologic and climate studies: Soil moisture content affects surface evaporation, runoff, albedo, emissivity, and portioning of sensible and latent heat fluxes. Some studies [e.g., *U.S. National Research Council*, 1994] show that soil moisture's effect on the atmosphere is secondary only to that of sea surface temperature (SST) on a global scale and even exceeds SST's effect over land. According to *Chahine* [1992], more than 65 percent of the precipitation over land comes from continental evaporation, which is strongly affected by soil moisture content. Accurate knowledge of spatial and temporal variations of soil moisture is needed for weather predictions and climate studies. This knowledge is, however, limited by a lack of measurements of soil moisture over many land areas [*Robock et al.*, 2000].

[3] To improve our knowledge of soil moisture variability, there have been large efforts to create estimates of soil moisture fields using land surface models forced with realistic precipitation and other atmospheric data, such as the Global Soil Wetness Project (http://grads.iges.org/gswp/) [*Dirmeyer et al.*, 1999], the North America Land Data Assimilation System (NLDAS) [*Mitchell et al.*, 2004], Global Land Data Assimilation System (GLDAS) (http://ldas.gsfc.nasa.gov/), and others [e.g., *Nijssen et al.*, 2001; *Qian et al.*, 2006; *Sheffield and Wood*, 2008]. However, because these efforts currently do not assimilate observational soil moisture data, they are not data assimilation systems in a conventional sense.

[4] In land data assimilation, Kalman filter methods, such as the extended Kalman filter (EKF) [*Entekhabi et al.*, 1994] and *Ensemble Kalman Filter* (EnKF) [*Evensen*, 2003], are the most frequently used optimization algorithms. In the EKF, the state variables can only propagate through the first-order linearization of the nonlinear system, which can introduce large errors in variable estimation, especially for some highly nonlinear systems like the soil water hydrodynamic equation. These errors could result in suboptimal performance and sometimes divergence of the filter. The unscented Kalman filter (UKF) [*Wan and van der Merwe*, 2001] addresses these problems using a deterministic sampling approach to capture the posterior mean and covariance accurate to the third order for any nonlinearity when propagated through the true nonlinear system. It represents an alternative to the EKF or EnKF, and provides superior (almost equivalent) performance at an equivalent (less) computational complexity. The UKF has been applied to nonlinear estimation problems in a variety of fields such

[1] Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China.
[2] National Center for Atmospheric Research, Boulder, Colorado, USA.

0148-0227/08/2007JD009650$09.00

as global positioning system [*van der Merwe and Wan*, 2004], space craft attitude estimation [*Crassidis and Markley*, 2003], ballistic missile tracking [*Saulson and Chang*, 2004], object tracking in image analysis system [*Chen et al.*, 2002]. The application of the UKF may be divided into state estimation, parameter estimation, and dual estimation (i.e., estimating both the state and parameters at the same time), referred to as the dual UKF. The dual-UKF method consists of two usual UKF filters (one for the states and the other for the parameters), in contrast to only one filter in the usual UKF, that simultaneously optimize both model states and parameters using observational information, thereby leading to improved simulations. In a land surface study, *Gove and Hollinger* [2006] applied the dual-UKF method for simultaneous state and parameter estimation for land surface-atmosphere flux exchanges.

[5] Some widely used land surface models, such as the National Center for Atmospheric Research (NCAR) Community Land Model Version 2.0 (CLM2) [*Bonan et al.*, 2002; *Oleson et al.*, 2004], are designed to consider the model subgrid heterogeneity and soil water thawing and freezing, so that soil moisture of different fractions within one model grid box can have large differences. Furthermore, the CLM2 computes both the liquid and solid contents of soil moisture, which are then used to derive the total soil moisture content. Soil moisture observations, however, usually provide only total soil moisture content. How to use the total soil moisture content from observations to forecast a new model state (i.e., the liquid and solid water contents for different fractions) is a nontrivial task, especially when the subgrid heterogeneity and liquid and solid phases need to be considered. It is difficult to consider all these aspects simultaneously using a single state filter like in the usual UKF and EKF methods. Although we used observations from stations in the following experiments, our system is designed for assimilating gridded satellite data sets in which the observations represent the mean value of the model gridcell. Of course, it seems more appropriate to use the gridded satellite based soil moisture data to do such evaluations, but unfortunately, whose quality is still expected to be improved greatly compared with the observation soil moisture data in situ.

[6] In this paper, we employ the dual-UKF method to consider the land surface subgrid-scale heterogeneity and soil water thawing and freezing, and implement it into the CLM2 to build a soil moisture data assimilation system primarily for assimilating satellite observations. In this system, liquid soil moisture content in a given fraction of a model grid is assimilated through the state filtering, while solid moisture content in the same fraction and solid and liquid moisture contents in the other fractions are optimized by the parameter filter. To our knowledge, there have been few studies to account for both subgrid-scale heterogeneity and soil water thawing and freezing in land surface data assimilation. Preliminary assimilation results show that the dual-UKF-based assimilation framework assimilates soil moisture more effectively and precisely than the usual UKF- and EKF-based assimilation systems that do not consider model's subgrid-scale heterogeneity and soil water phase changes. Assimilated soil moisture can be improved greatly not only in the layers where observations are assimilated directly, but also, to a lesser extent, in the layers where no observations are available.

2. Model and Methods

2.1. CLM2 and Its Soil Water Calculation

[7] We used the land model CLM2 [*Bonan et al.*, 2002] in this study. The CLM2, an NCAR version of the Common Land Model [*Dai et al.*, 2003], is a comprehensive, global land surface model that is used as the land component of the Community Climate System Model (CCSM). The CLM2 considers the subgrid-scale heterogeneity by subdividing each grid cell into a number of fractions. Each fraction contains a single land cover type. By default, each grid box is divided into up to five fractions for a dominant vegetation type, a secondary vegetation type, a fraction with bare soils, a wetland fraction, and an inland water fraction. Energy and water balance calculations are performed over each fraction at every time step and each fraction maintains its own prognostic variables. The fractions within a grid cell respond to the mean conditions from the overlying atmospheric grid box, and this grid box, in turn, responds to the area-weighted mean fluxes of heat and moisture from the fractions. The fractions within a grid cell do not interact with each other directly. The CLM2 has one vegetation layer, 10 unevenly spaced vertical soil layers, and up to 5 snow layers (depending on the total snow depth). It computes soil temperature and soil water content in the 10 soil layers to a depth of 3.43 m in every fraction within one grid cell.

[8] The conservation of liquid water mass for one-dimensional vertical water flow in a soil column in the CLM2 is expressed as

$$\frac{\partial \theta_{liq}}{\partial t} = -\frac{\partial q}{\partial z} - E - R_{fm}, \tag{1}$$

where θ_{liq} is the volumetric soil moisture content (m^3/m^3), q is the vertical soil water flux (mm s^{-1}) (note there is no horizontal movements of water within soils in the CLM2), E is the evapotranspiration rate, which is partitioned into each soil layer on the basis of the effective fraction of roots within the layer in the CLM2, and R_{fm} is the melting (negative) or freezing (positive) rate, and z is the depth from the soil surface. Both q and z are positive downward.

[9] The soil water flux q is described by Darcy's law [*Darcy*, 1856]:

$$q = -K_c \frac{\partial(\psi + z)}{\partial z}, \tag{2}$$

where K_c is the hydraulic conductivity (mm s^{-1}), and ψ is the soil matric potential (mm). The CLM2 computes liquid volumetric soil moisture content in the 10 soil layers by (1) and (2) in every cell fraction and then adjusts the solid and liquid volumetric soil moisture contents again on the basis of the energy (released from freezing soil water or consumed by melting) budget, which is finally followed by updating the whole gridcell total soil moisture content

(θ_g) for each layer as the sum of area-weighted soil moisture averaged over all the fractions within the grid cell, that is

$$\theta_g = \sum_{j=1}^{M} a_j(\theta_{liq,j} + \theta_{ice,j}), \tag{3}$$

where M is the number of fractions in the grid cell, a_j is the area-based weighting factor, and $\theta_{liq,j}$, $\theta_{ice,j}$ are, respectively, the liquid and ice volumetric soil moisture content in the jth fraction of the grid cell.

2.2. Application of the Dual Unscented Kalman Filter

2.2.1. Unscented Transformation

[10] The basic framework for the UKF involves estimation of the state of a discrete, nonlinear dynamic system [*Julier and Uhlmann*, 2004],

$$x_{k+1} = F(x_k, u_k) + v_k, \tag{4}$$

$$y_k = H(x_k) + \eta_k, \tag{5}$$

where x_k represents the unobserved state of the system (e.g., the liquid soil moisture content θ_{liq} in equation (1)) at time step k, u_k is a known input (e.g., parameters K_c and ψ in the hydrodynamic equations (1) and (2)), and y_k is the observed signal (e.g., the observed total soil moisture content). The process noise v_k (e.g., the infiltration errors) partly drives the dynamic system, and the observation error is given by η_k (e.g., the errors in the observed soil moisture content). F and H each represents a nonlinear function.

[11] Let n_x be the dimension of the state space (for example, n_x = 10 soil layers in the CLM2) and $\bar{x}$ and P_x be the mean and covariance of x_k, respectively. Select $2n_x + 1$ sigma points (sampling members) by letting $\lambda = \alpha^2(n_x + \kappa) - n_x$. Here, the parameter $0 \le \alpha \le 1$ controls the spread of the sigma points and weighting for higher-order moments, while parameter $\kappa \ge 0$ is not critical and is often set to zero [*van der Merwe*, 2004]. The smaller the value of α, the narrower the sigma point spread and the less likely to pick up anomalous effects in the distribution and the drawing χ_i [*Julier and Uhlmann*, 2004]:

$$\chi_0 = \bar{x} \tag{6}$$

$$\chi_i = \bar{x} + \left(\sqrt{(n_x + \lambda)P_x}\right)_i, \quad i = 1, \cdots, n_x, \tag{7}$$

$$\chi_i = \bar{x} - \left(\sqrt{(n_x + \lambda)P_x}\right)_i, \quad i = n_x + 1, \cdots, 2n_x, \tag{8}$$

$$W_0^{(m)} = \frac{\lambda}{n_x + \lambda}, \qquad i = 0, \tag{9}$$

$$W_0^{(c)} = \frac{\lambda}{n_x + \lambda} + (1 - \alpha^2 + \beta), \quad i = 0, \tag{10}$$

$$W_i^{(m)} = W_i^{(c)} = \frac{1}{2(n_x + \lambda)}, \quad i = 1, \cdots, 2n_x. \tag{11}$$

[12] The sigma point set (sampling set) $S = \{\chi_i, W_i^{(j)}, i = 0, \ldots 2n_x, j \in (m,c)\}$ is composed of the sigma points χ_i (equations (6), (7), and (8) and their respective mean (m) weights $W_i^{(m)}$ and covariance (c) weights $W_i^{(c)}$). The parameter $\beta \ge 0$ also controls the spread of the sigma points and weighting for higher-order moments. For Gaussian distributions, $\beta = 2$ is optimal [*Julier and Uhlmann*, 2004].

[13] Given an arbitrary nonlinear function F in (4), the unscented transformation can be performed through following two steps [*van der Merwe*, 2004]: (1) Apply the nonlinear transformation F to each sigma point, namely, $y_i = F(\chi_{k,i}, u_k)$, $i = 0, \ldots, 2n_x$; and (2) calculate the mean, covariance and cross covariance, respectively, of the transformed sigma points as

$$\bar{y} = \sum_{i=0}^{2n_x} W_i^{(m)} y_i, \quad P_y = \sum_{i=0}^{2n_x} W_i^{(c)} (y_i - \bar{y})(y_i - \bar{y})',$$

$$P_{xy} = \sum_{i=0}^{2n_x} W_i^{(c)} (x_i - \bar{x})(y_i - \bar{y})'.$$

2.2.2. Dual Unscented Kalman Filter

[14] One general method, namely dual estimation (e.g., the dual UKF) [*van der Merwe*, 2004; *Wan and van der Merwe*, 2001], has been developed to simultaneously estimate the states and parameters from the noisy measurements through Kalman filtering. The dual-UKF approach can be implemented within the unscented Kalman filter framework [*Wan and Nelson*, 2001; *Wan and van der Merwe*, 2001], with two usual UKF filters (one for states and the other for parameters) to simultaneously optimize the model states and parameters using observational information. In the dual UKF, the cross covariances are not explicitly estimated and are assumed to be zero.

2.2.3. State Filter

[15] The dual-UKF system for state with additive stochastic noise is [*Wan and van der Merwe*, 2001; *van der Merwe*, 2004]

$$x_k = F(x_{k-1}; \overline{w}_{k-1}) + v_{k-1}, \tag{12}$$

$$y_k = H(x_k; \overline{w}_{k-1}) + \eta_k \tag{13}$$

with the noise sequences v_{k-1} and η_k as defined in equations (4) and (5). The $\overline{w}_{k-1}$ (i.e., $u_{k-1} = w_{k-1}$ in equation (4)) are estimates of the parameters from the previous time step in by the parameter filter and are treated as constants in the state filter. The state filter recursions are given in the following steps [*van der Merwe*, 2004; *Wan and van der Merwe*, 2001]:

[16] 1. Initialization with

$$\overline{x}_0 = E(x_0), P_{x_0} = E\left[(x_0 - \overline{x}_0)(x_0 - \overline{x}_0)'\right].$$

[17] 2. Sigma point calculation and time update

$$\chi_{k-1} = \left[\overline{x}_{k-1}, \overline{x}_{k-1} + \sqrt{(n_x+\lambda)P_{x_{k-1}}}, \overline{x}_{k-1} - \sqrt{(n_x+\lambda)P_{x_{k-1}}}\right]$$
$$\overset{*}{\chi}_{k|k-1} = F(\chi_{k-1}; \overline{w}_{k-1})$$
$$\overline{x}_{k|k-1} = \sum_{i=0}^{2n_x} W_i^{(m)} \overset{*}{x}_{i,k|k-1}$$
$$P_{x|k-1} = \sum_{i=0}^{2n_x} W_i^{(c)} \left(\overset{*}{\chi}_{i,k|k-1} - \overline{x}_{k|k-1}\right)\left(\overset{*}{\chi}_{i,k|k-1} - \overline{x}_{k|k-1}\right)' + Q_k$$
$$\chi_{k|k-1} = \left[\overline{x}_{k|k-1}, \overline{x}_{k|k-1} + \sqrt{(n_x+\lambda)P_{x_{k|k-1}}}, \overline{x}_{k|k-1} - \sqrt{(n_x+\lambda)P_{x_{k|k-1}}}\right]$$
$$Y_{k|k-1} = H\left(\chi_{k|k-1}; \overline{w}_{k-1}\right)$$
$$\overline{y}_{k|k-1} = \sum_{i=0}^{2n_x} W_i^{(m)} y_{i,k|k-1}.$$

[18] 3. Measurement update

$$P_{y_k} = \sum_{i=0}^{2n_x} W_i^{(c)} \left(y_{i,k|k-1} - \overline{y}_{k|k-1}\right)\left(y_{i,k|k-1} - \overline{y}_{k|k-1}\right)' + R_k$$
$$P_{x_k y_k} = \sum_{i=0}^{2n_x} W_i^{(c)} \left(\chi_{i,k|k-1} - \overline{x}_{k|k-1}\right)\left(y_{i,k|k-1} - \overline{y}_{k|k-1}\right)'$$
$$K_k = P_{x_k y_k} P_{y_k}^{-1}$$
$$\overline{x}_k = \overline{x}_{k|k-1} + K_k\left(y_k - \overline{y}_{k|k-1}\right)$$
$$P_{x_k} = P_{x_{k|k-1}} - K_k P_{y_k} K_k',$$

where R_k and Q_k are also the measurement and process noise covariances, respectively.

[19] In our formulation, the process model F is given by equations (1) and (2) and their discrete format and the state variable x_k is *the liquid volumetric soil moisture content* in each fraction of one grid cell, while the measurement model H can be an arbitrary operator (including linear and nonlinear, in our experiments, it is simply a real matrix). The input variable x_k (the sum of the liquid and solid volumetric soil moisture) of the measurement model equation (13) is usually based on the whole-grid value when satellite observations are used. Thus, in our data assimilation system, the state variable x_k of the process model is actually different from the input signal of the measurement model: The former is a variable defined on the grid fractions, while the latter is defined on the whole grid cell. This issue is addressed further below.

2.2.4. Parameter Filter

[20] The UKF can be used for parameter estimation and approximation for both clean and noisy time series. In the last subsection, we mentioned that the state variable (θ_{liq}) of the process model is defined differently from that of the measurement model, so we have to transform the measurement model (H) from the grid cell into the grid fractions because the process model F is just on the grid fractions. The variables of the process model and the measurement model have the following relationship (cf. equation (3)):

$$\theta_g = \sum_{j=1}^{M} a_j\left(\theta_{liq,\,j} + \theta_{ice,\,j}\right).$$

[21] For any given grid fraction i

$$\theta_g = \sum_{j=1,\,j\neq i}^{M} a_j\left(\theta_{liq,\,j} + \theta_{ice,\,j}\right) + a_i\left(\theta_{liq,\,i} + \theta_{ice,\,i}\right)$$
$$= \sum_{j=1,\,j\neq i}^{M} a_j\left(\theta_{liq,\,j} + \theta_{ice,\,j}\right) + a_i\theta_{liq,\,i} + a_i\theta_{ice,\,i}, \quad (14)$$

[22] From equations (3) and (14), the measurement model can be rewritten as follows:

$$y_k = H\left(\sum_{j=1,\,j\neq i}^{\mathrm{M}} a_j\left(\theta_{liq,\,j} + \theta_{ice,\,j}\right) + a_i\theta_{liq,\,i} + a_i\theta_{ice,\,i}\right) + \eta_k,$$
$$y_k = \overline{H}\left(\theta_{liq,i}, w_k\right) + \eta_k, \quad (15)$$

where $\overline{H}(\theta_{liq}, w_k) = H(\sum_{j=1,\,j\neq i}^{M} a_j(\theta_{liq,\,j} + \theta_{ice,\,j}) + a_i\theta_{liq,\,i} + a_i\theta_{ice,\,i})$.

[23] For each time step, the CLM2 does the simulation loop through all the fractions within the simulated grid cell. When the assimilation is being conducted within a given fraction, it is hard to know whether the assimilation is done or not for all the other fractions within the same grid cell. To address this problem, we take other fractions' liquid and ice soil moisture in the same grid cell as parameters that can be optimized during the assimilation procedure for the given fraction i, then the w_k in (12) and (13) is composed of ($\theta_{liq,j}$, $\theta_{ice,j}$, $\theta_{ice,i}$) ($j \neq i$).

[24] The equations for the parameter filter, which is coupled with the measurement model in the dual setting usually, are

$$w_k = w_{k-1} + v_{k-1}$$
$$y_k = H(F(\overline{x}_{k-1}; w_k), w_k) + \eta_k$$

[25] In our soil moisture data assimilation system, we transform it into the following format

$$w_k = \left(\theta_{liq,\,j}, \theta_{ice,\,j}, \theta_{ice,\,i}\right)_{j\neq i} + v_{k-1},$$
$$y_k = \overline{H}\left(\theta_{liq,i}, w_k\right) + \eta_k,$$

where the v_k and η_k are the process and measurement noises, with covariances Q_{w_k} and R_{w_k} [*van der Merwe*, 2004; *Wan and van der Merwe*, 2001], respectively. The parameter vector (($\theta_{liq,j}$, $\theta_{ice,j}$, $\theta_{ice,i}$) ($j \neq i$).) in our soil moisture data assimilation can be then updated through the parameter filter recursion [*van der Merwe*, 2004; *Wan and van der Merwe*, 2001].

[26] In the CLM2, the volumetric heat capacity C_i (J m^{-3} K^{-1}) for soil is from work by *De Vries* [1963] and depends on the heat capacities of the soil solid, liquid water and ice constituents

$$C_i = C_{s,i}\left(1 - \theta_{sat,i}\right) + \frac{w_{ice,i}}{\Delta z_i} C_{ice} + \frac{w_{liq,i}}{\Delta z_i} C_{liq} \quad (16)$$

where i is the soil layer number, $C_{s,i}$ is the heat capacity of soil solids, C_{liq} and C_{ice} are the specific capacities of liquid water and ice, respectively, $\theta_{sat,i}$ is the ith layer saturated soil moisture, $w_{ice,i}$ and $w_{liq,i}$ are the mass of ice and liquid

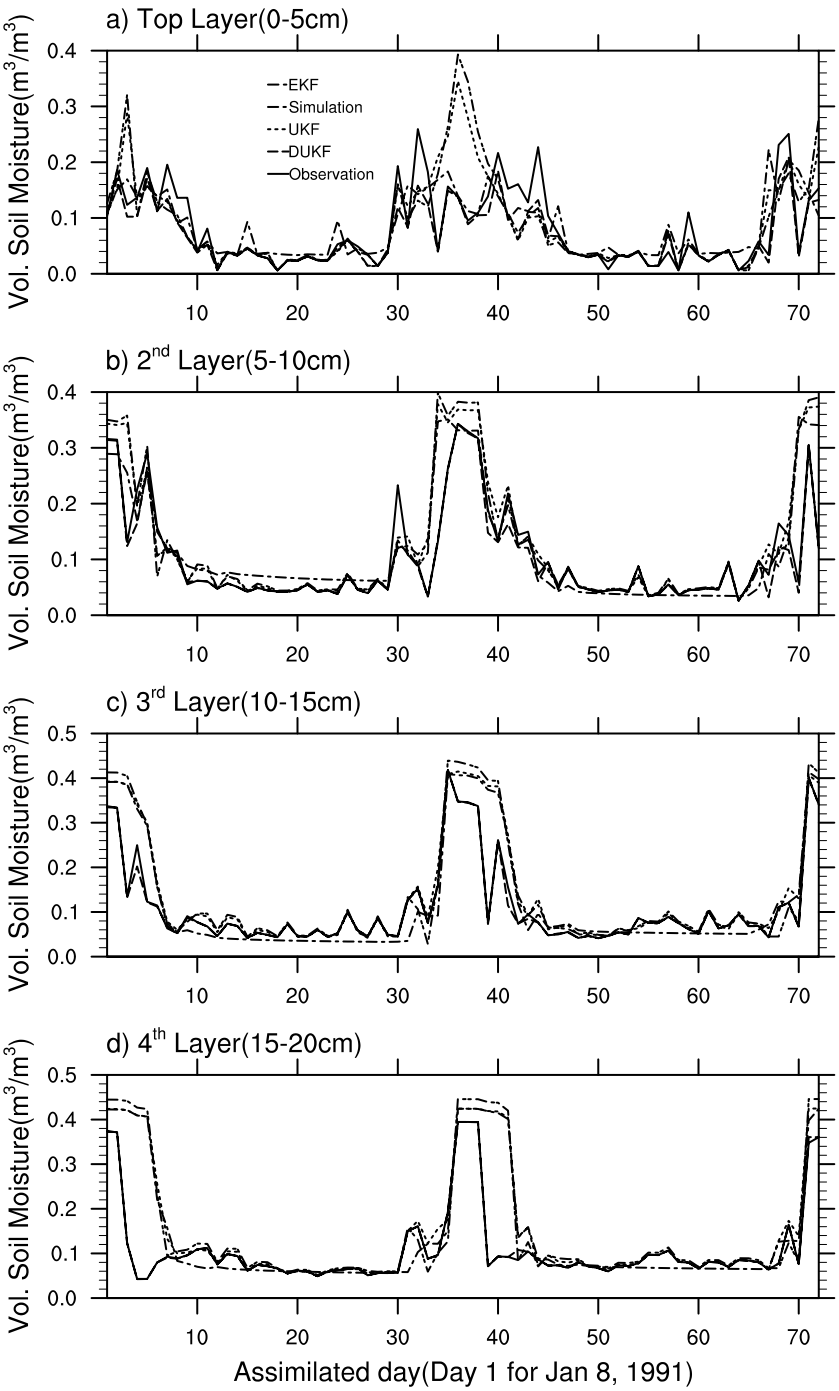

Figure 1. Time series of assimilated daily volumetric soil moisture (m3/m3) at Xilinguo (39.080°N, 105.380°E) in North China from 8 January 1991 to 28 December 1992 from observations and three assimilation systems and a CLM2 simulation forced with observed atmospheric forcing for the top four soil layers.

water (kg m^{-2}). The one dimensional soil temperature equation is

$$c\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left[\lambda\frac{\partial T}{\partial z}\right], \tag{17}$$

where T is soil temperature, λ is thermal conductivity. One can find the difference of assimilated (simulated) soil moisture definitely makes the volumetric heat capacity C_i differently from (16) and results in different soil temperature simulations on the basis of (17).

3. Preliminary Assimilation Experiments

[27] In this section, we use the atmospheric forcing data extracted from a global data set at 3-hourly and T62 ($\sim$1.875°) resolution [*Qian et al.*, 2006; *Tian et al.*, 2007]

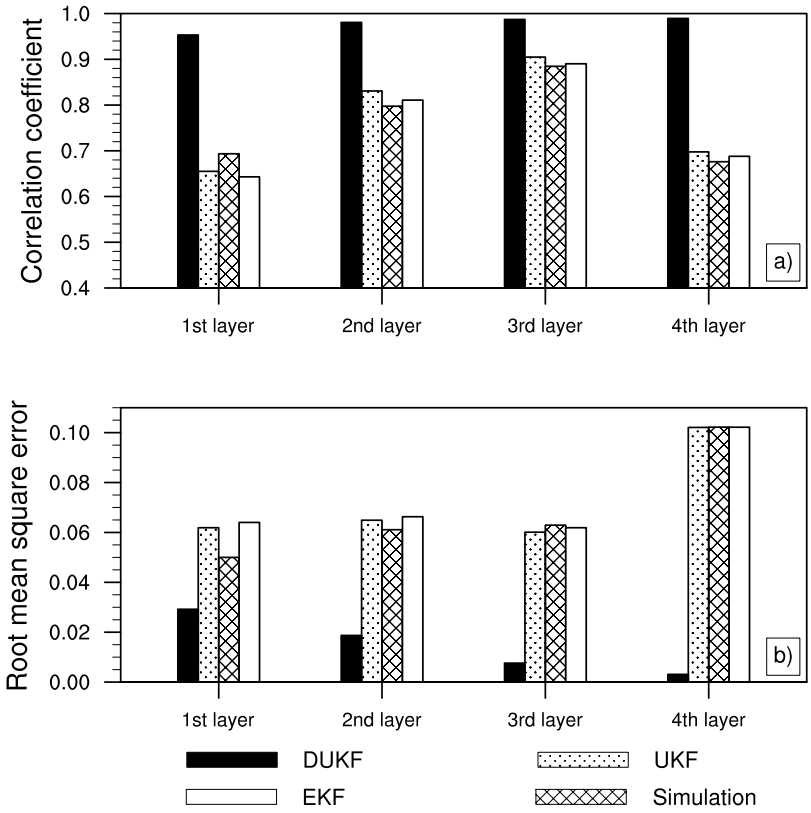

Figure 2. (a) Correlation coefficients between the observed and assimilated or simulated soil moisture shown in Figure 1 (the sample size is 72). (b) Root mean square (RMS) errors (m^3/m^3) for assimilated or simulated soil moisture content shown in Figure 1.

to drive the assimilation system at a station (Xilinguo, 39.08°N, 105.38°E) in north China to assimilate the observed soil moisture [*Li et al.*, 2005] from 1991 to 1992. We chose this station partially because it is located in the seasonally frozen soil zone, which allows us to investigate how the dual-UKF-based system that considers the subgrid-scale heterogeneity and soil water phase changes performs during the thawing and freezing periods. Through numerical sensitivity experiments, we chose $\alpha = 0.9$, $\kappa = 0$, $\beta = 2.0$ in our dual-UKF-based assimilation system (DUKF). The model gridcell is 0.01° longitude by 0.01° latitude in all our CLM2 experiments. Because our goal is to develop a system for assimilating gridded satellite observations, we assume that the station observations represent the mean of this model gridcell. It can be argued, however it is one of few acceptable ways to evaluate a land data assimilation system: (1) Partly limited by our poor knowledge of the complicated land conditions (such as subgrid-scale heterogeneity, soil ice and snow, etc.), the quality of the gridded satellite-based soil moisture requires considerable improvements, especially under freezing conditions. (2) If the model grid resolution is chosen considerably high (for example, 0.01° longitude by 0.01° latitude), the observations in situ can be supposed to represent the mean of this model gridcell, which is very usual in land data assimilation evaluations. This enables us to examine how this mean value is reproduced by the DUKF method in comparison with the usual UKF- and EKF-based systems that cannot, in contrast to the DUKF method, consider the model subgrid-scale heterogeneity and the phase change in soil water. The total soil moisture content is supposed to be the model forecast (state) variable in the UKF- and EKF-based assimilation systems, while the liquid and solid moisture contents for each fraction are the forecast states in the DUKF method. The same atmospheric forcing data were used to drive the CLM2, and the observed soil moisture data from the station were assimilated using the three different methods, except for the simulation case in which no soil moisture data were assimilated.

[28] The linearization of the soil moisture equation (1) in the EKF method follows the format in work by *Zhang et al.* [2006]. Volumetric soil moisture content at this station was measured three times each month on the 8th, 18th, 28th day (at the same local time) at 11 vertical layers, two 5-cm layers from 0 cm down to 10 cm depth and nine 10-cm layers from 10 cm down to 1 m. Soil moisture in only the top four layers is assimilated in our experiments. Data from 8 January 1991 to 28 December 1992 were used here.

3.1. Soil Moisture

[29] Figure 1a shows the time series of assimilated daily soil moisture content from 1991 to 1992 (three times each month on the 8th, 18th, 28th day) for the station Xilinguo for the top four layers compared with the station daily observations and CLM2 simulations without assimilating the observed soil moisture data. For all the top four layers, the DUKF-based system assimilates the observed soil moisture evolution better than the usual UKF- and EKF-

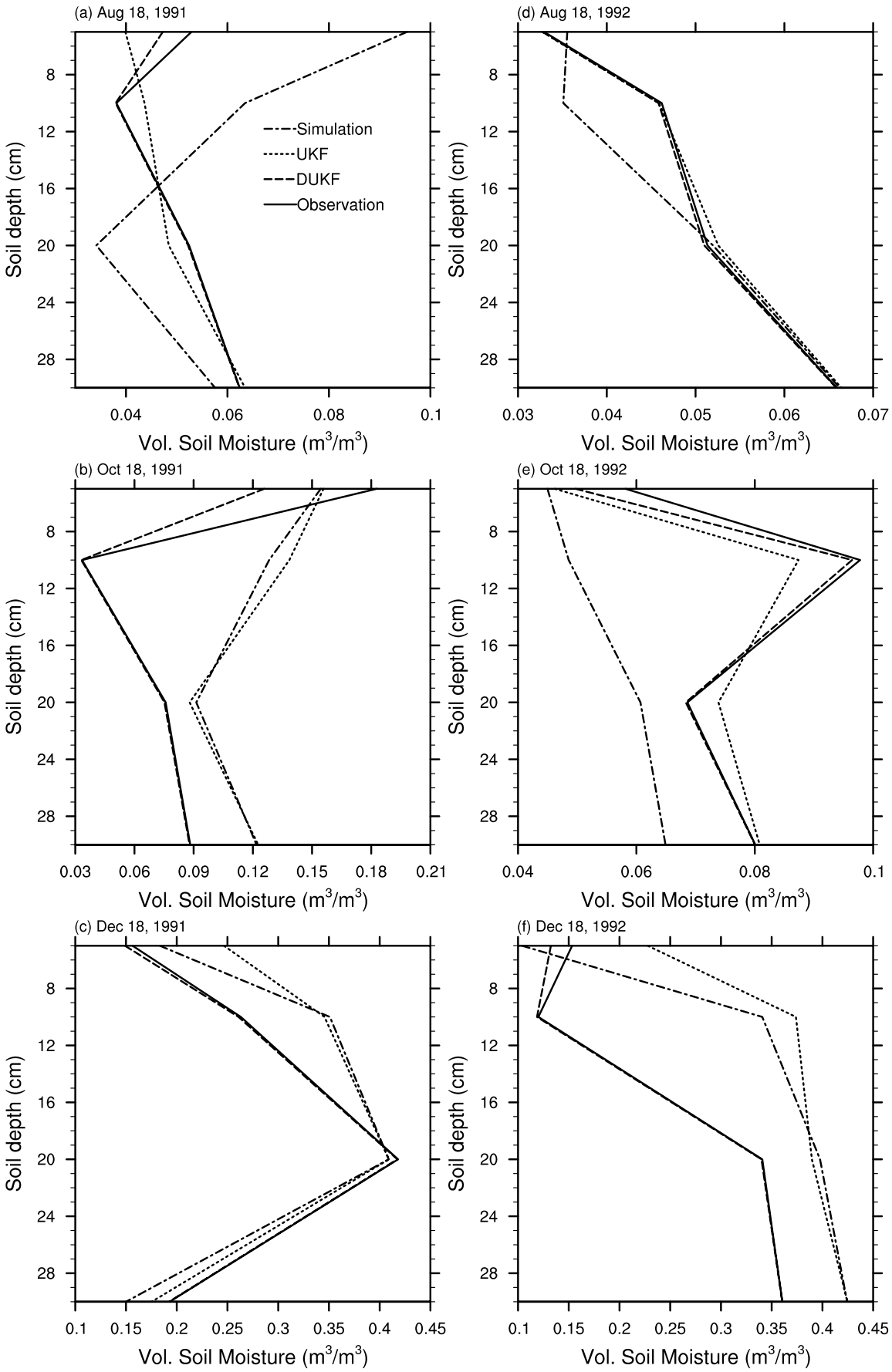

Figure 3. Observed and assimilated or simulated soil moisture profiles on 6 selected days, Xilinguo site.

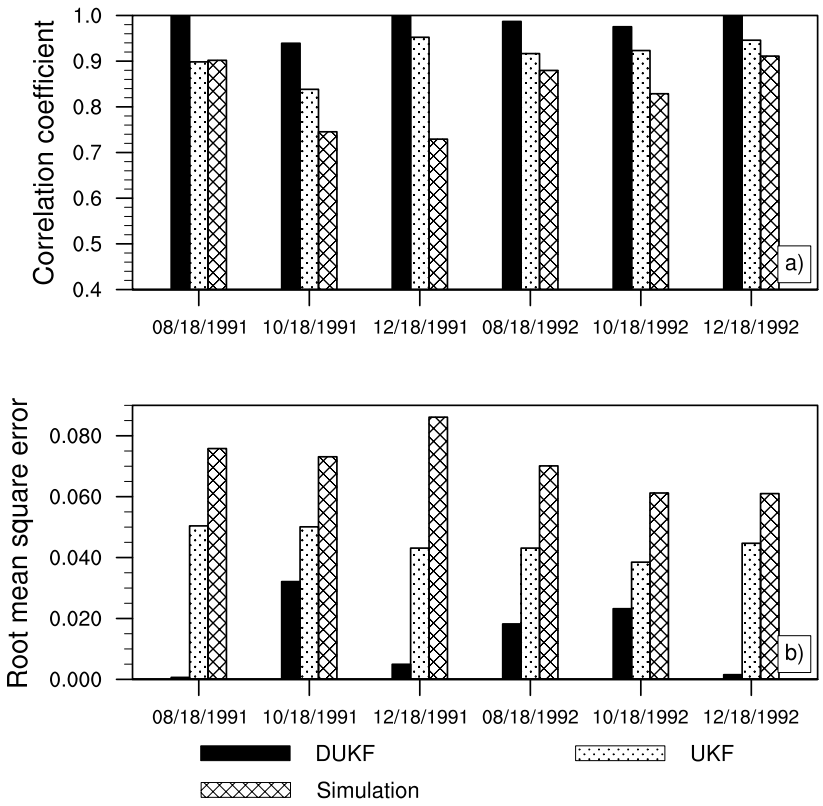

Figure 4. (a) Spatial correlation coefficients between the observed and assimilated or simulated soil moisture profiles shown in Figure 3. (b) Root mean square errors (m^3/m^3) for the assimilated or simulated soil moisture profiles shown in Figure 3.

based systems, whose performances are not better than the climate-forced CLM2 simulation (without assimilated the soil moisture data). The correlation coefficients (r) between the observed and assimilated or simulated daily soil moisture during 1991–1992 are shown in Figure 2a, with the root mean square (RMS) errors shown in Figure 2b. The soil moisture content from the DUKF method correlates with the observations considerably higher ($r > 0.95$) than the other three cases ($r < 0.91$), while its RMS errors are much smaller. For the first layer, the UKF and EKF cases are even

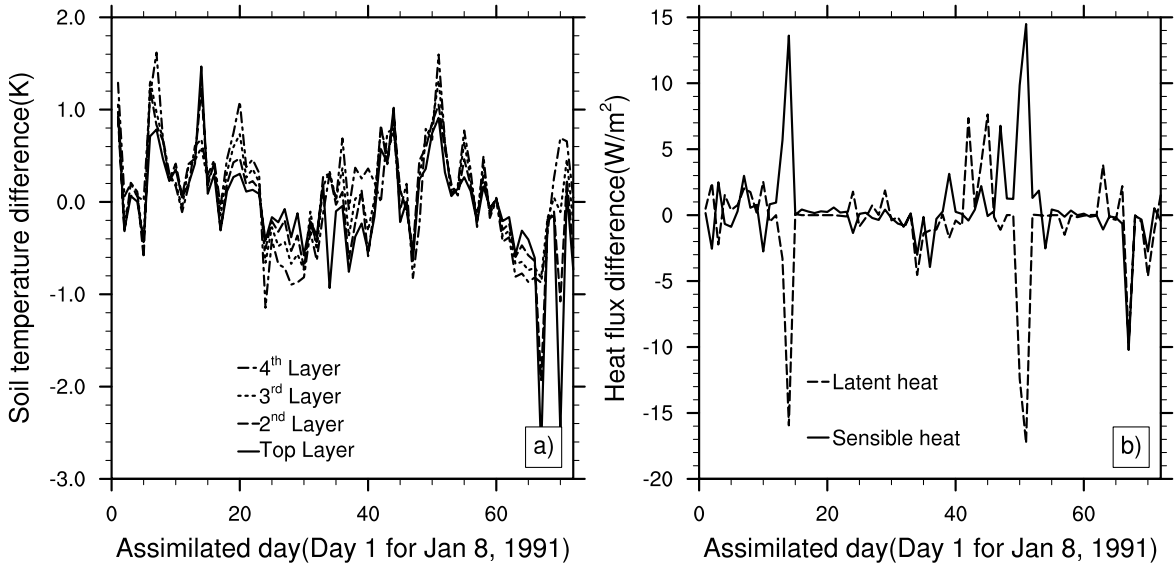

Figure 5. (a) The UKF minus DUKF difference of assimilated daily soil temperatures from the two assimilation systems at the Xilinguo site (39.08°N, 105.38°E) from 8 January 1991 to 28 December 1992 for the top four layers. (b) Same as Figure 5a but for surface latent and sensible heat fluxes.

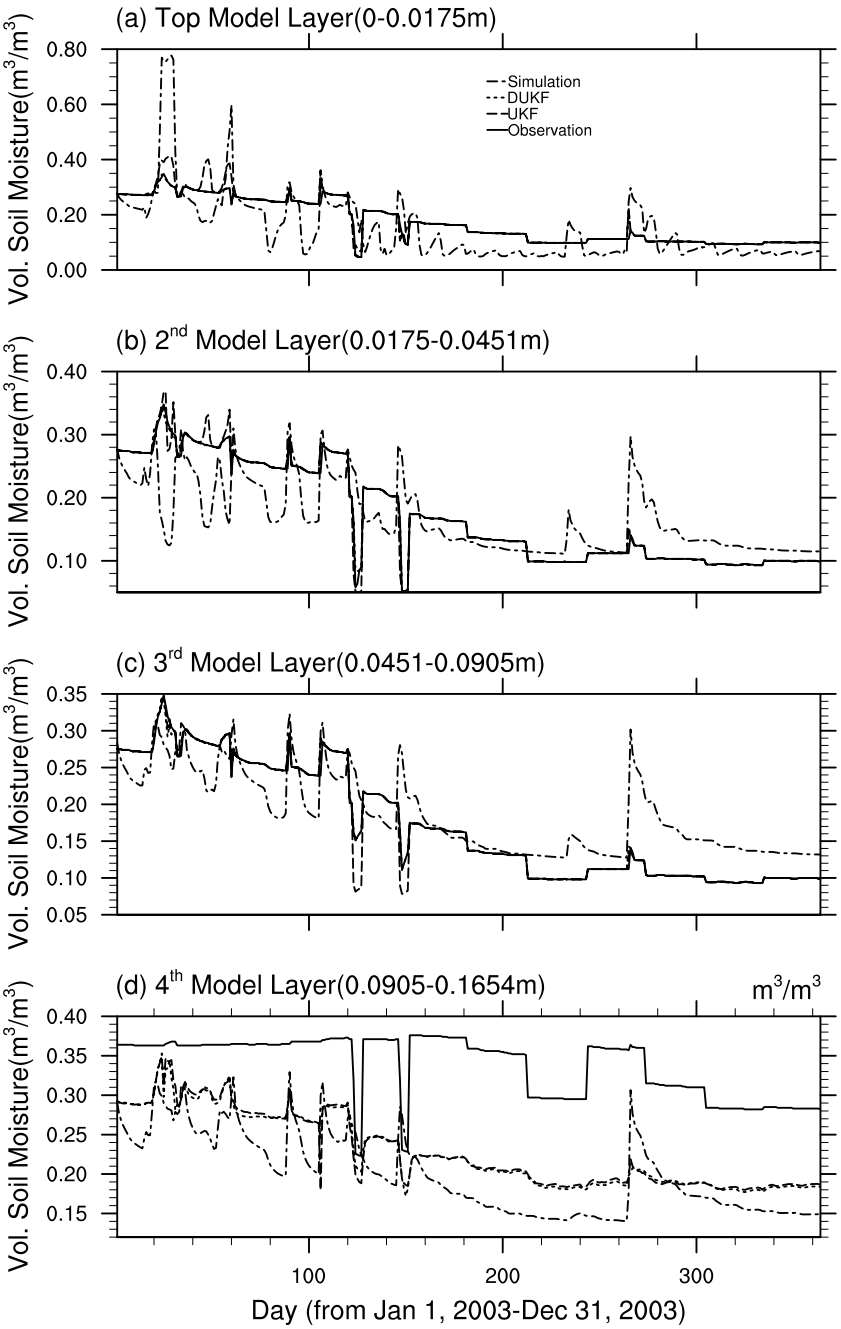

Figure 6. Time series of daily volumetric soil moisture (m^3/m^3) at site Qianyanzhou (26°44′N, 115°04′E) in south China from 1 January 2003 to 31 December 2003 for the top four soil layers from observations and CLM2-based assimilations/simulations.

worse than the climate-forced pure simulation. The EKF-based system performs slightly worse than the usual UKF method, which may be explained by the fact that the soil water hydrodynamic equation (1) is a highly nonlinear system and the linearization operator used in the EKF method can only propagate with the first-order precision. For simplicity, we exclude the EKF case in the following discussions

[30] Soil moisture profiles from 6 selected days during different seasons are shown in Figure 3. Figure 3 shows that the dual-UKF-based system assimilates the soil moisture profiles better than the usual UKF method during both the ice-free (Figures 3a and 3d) and freezing (Figures 3c and 3f) conditions. During the cold season, the existence of soil ice plays a significant role, but the usual UKF-based system cannot account for its influence, resulting in poor performance (Figures 3c and 3f). The (spatial) correlation coefficients (Figure 4a) and RMS errors (Figure 4b) between the observed and assimilated or simulated soil moisture profiles of Figure 3 further quantify the superior performance of the DUKF method.

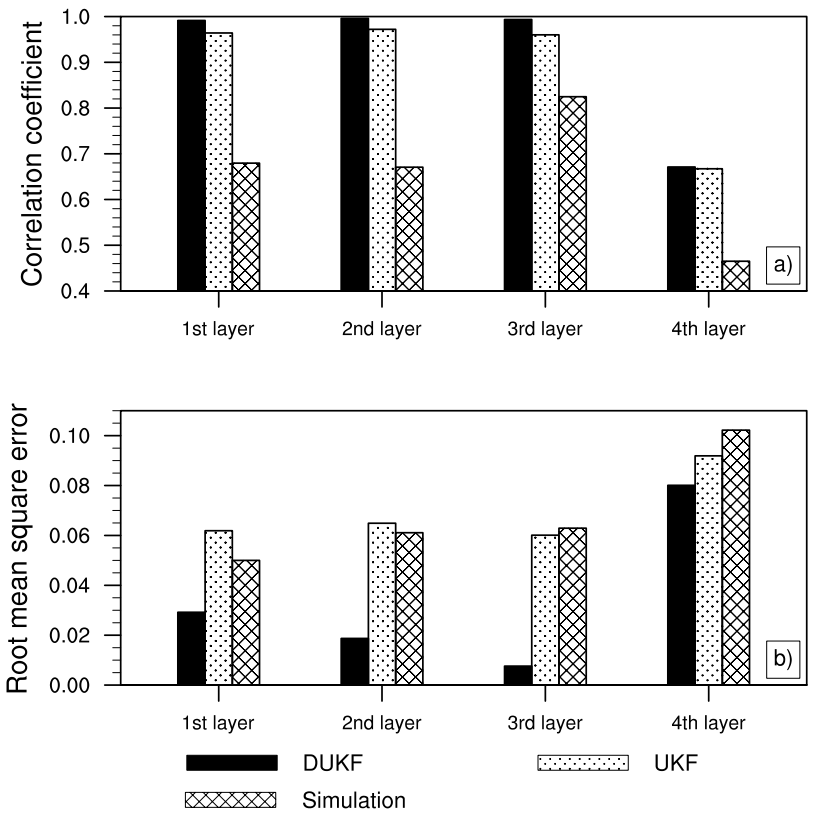

Figure 7. (a) Correlation coefficients between the observed and assimilated or simulated daily soil moisture content shown in Figure 6. (b) Same as Figure 7a but for the RMS errors (m^3/m^3) for the assimilated or simulated daily soil moisture content shown in Figure 6.

3.2. Soil Temperature and Surface Fluxes

[31] Figure 5a shows that the UKF minus DUKF difference in the simulated soil temperature for the top four layers by the two assimilation systems. The magnitude of the temperature differences varies from about −2.8 to 1.8 K, slightly larger in the top layer. These temperature differences are substantial and nonnegligible, suggesting that the two assimilation systems produce different results not only for soil moisture but also for temperature.

[32] The differences in soil moisture and temperature can change surface latent and sensible fluxes. Sensible heat flux simulation then becomes different because of simulated soil temperature and latent heat flux varying with the variations of the soil moisture. Figure 5b shows the UKF minus DUKF differences in surface simulated heat flux (sensible heat flux and latent heat flux) differences simulated by between the two assimilation systems. The difference can be as large as −18 to 8 W/m^2 for the latent heat flux and the magnitude of the simulated sensible heat flux difference is about from −11 to 15 W/m^2. These numbers are large and could have significant effects on the interactions between the atmosphere and land surface conditions.

4. More Investigation Experiments at Qianyanzhou Site

[33] Because satellite observations of soil moisture are available only for the several topsoil layers, we want to know whether the DUKF-based assimilation system can improve the soil moisture in the deep layers when observations for these layers are unavailable. In this section, we use soil moisture data from a site (Qianyanzhou, 26°44′N, 115°04′E) in south China to investigate this issue. Soil moisture measured once an hour from the skin downward to about 18 cm from 1 January 2003 to 31 December 2003 was converted into moisture content for the top four model layers (i.e., 0–1.75 cm, 1.75–4.51 cm, 4.51–9.05 cm, and 9.05–16.54 cm), but only the data for the top three layers were assimilated. The observed soil moisture for the fourth layer was used to evaluate the performance of the two assimilation systems. The atmospheric forcing data for driving the CLM2 were derived from hourly meteorological measurements at the site.

[34] Figure 6 shows the time series of daily soil moisture from the CLM2-based assimilations compared with the observed and CLM2-simulated soil moisture for the top four layers at the Qianyanzhou site from 1 January 2003 to 31 December 2003. For the top three layers (observed soil moisture is assimilated in these layers), the DUKF-based system reproduces the observed soil moisture evolution very well, with near-perfect correlation (Figure 7a) and small RMS errors (Figure 7b), Since no observations for the fourth layer were assimilated, the observational information is transferred from the top three layers to the fourth layer only through the soil water hydrology dynamics and the covariance information. Figures 6 and 7 show that the DUKF method slightly outperforms the UKF method for the fourth layer with a smaller RMS error. Thus, the DUKF

method not only greatly improves the assimilation in layers with observations, but also, albeit to a lesser extent, outperforms the usual UKF method in layers without observations.

5. Summary and Concluding Remarks

[35] In this study, we have implemented the dual-UKF method into the CLM2 to account for the effects of subgrid-scale heterogeneity and soil water thawing and freezing and developed a new data assimilation system for assimilating satellite observations of soil moisture. The dual-UKF method differs from other conventional data assimilation approaches such as the usual UKF and EKF in that it uses two filters (one for states and one for parameters), in contrast to only one filter in the other methods, to simultaneously optimized the states and parameters of a nonlinear systems. We have applied this dual-UKF-based system to two sites in north and south China to evaluate its performance in comparison with the usual UKF and EKF methods that do not consider the subgrid heterogeneity and soil water phase changes. The results show that the dual-UKF method assimilates the observed variations in daily soil moisture content much better than the usual UKF and EKF methods, with improved correlations and reduced RMS errors. To a lesser extent, this improvement also propagates to lower layers where observations are unavailable.

[36] The evaluation experiments carried out here are only for two sites in East China. More comprehensive experiments using in situ observations from other locations [e.g., *Robock et al.*, 2003; *Dai et al.*, 2004] and using satellite observations are necessary before the dual-UKF-based system can be applied for large-scale data assimilation of soil moisture. Furthermore, the dual-UKF method can also be implemented into other land surface models such as the newer versions of the CLM.

[37] **Acknowledgments.** We thank three anonymous reviewers for helpful comments and Feng Chen for constructive discussions on figure drawing. This work was supported by the National Natural Science Foundation of China under grant 40705035, the Knowledge Innovation Project of Chinese Academy of Sciences under grants KZCX2-YW-217 and KZCX2-YW-126-2, the National Basic Research Program under grant 2005CB321704, and the Chinese COPES project (GYHY200706005). The National Center for Atmospheric Research is sponsored by the U.S. National Science Foundation.

References

Bonan, G. B., K. W. Oleson, M. Vertenstein, S. Levis, X. Zeng, Y. Dao, R. E. Dickinson, and Z.-L. Yang (2002), The land surface climatology of the Community Land Model coupled to the NCAR Community Climate Model, *J. Clim.*, *15*, 3123–3139, doi:10.1175/1520-0442(2002)015<3123:TLSCOT>2.0.CO;2.

Chahine, T. M. (1992), The hydrological cycle and its influence on climate, *Nature*, *359*, 373–380, doi:10.1038/359373a0.

Chen, Y., T. Huang, and Y. Rui (2002), Parametric contour tracking using unscented Kalman filter, paper presented at International Conference on Image Processing, Inst. of Electr. and Electron. Eng., Rochester, N. Y.

Crassidis, J. L., and F. L. Markley (2003), Unscented filtering for spacecraft attitude estimation, *J. Guidance Control Dyn.*, *26*(4), 536–542.

Dai, A., K. E. Trenberth, and T. T. Qian (2004), A global dataset of Palmer Drought Severity Index for 1870–2002: Relationship with soil moisture and effects of surface warming, *J. Hydrometeorol.*, *5*, 1117–1130, doi:10.1175/JHM-386.1.

Dai, Y., et al. (2003), The Common Land Model, *Bull. Am. Meteorol. Soc.*, *84*(8), 1013–1023, doi:10.1175/BAMS-84-8-1013.

Darcy, H. (1856), *Les Fontaines Publiques de la Ville de Dijon*, Dalmont, Paris.

De Vries, D. A. (1963), Thermal properties of soils, in *Physics of the Plant Environment*, edited by W. R. van Wijk, pp. 210–235, Elsevier, New York.

Dirmeyer, P. A., A. J. Dolman, and N. Sato (1999), The Global Soil Wetness Project: A pilot project for global land surface modeling and validation, *Bull. Am. Meteorol. Soc.*, *80*(5), 851–878, doi:10.1175/1520-0477(1999)080<0851:TPPOTG>2.0.CO;2.

Entekhabi, D., J. F. Galantowicz, and E. G. Njoku (1994), Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observation, *IEEE Trans. Geosci. Remote Sens.*, *32*, 438–448, doi:10.1109/36.295058.

Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, *53*, 343–367, doi:10.1007/s10236-003-0036-9.

Gove, J. H., and D. Y. Hollinger (2006), Application of a dual unscented Kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange, *J. Geophys. Res.*, *111*, D08S07, doi:10.1029/2005JD006021.

Julier, S. J., and J. K. Uhlmann (2004), Unscented filtering and nonlinear estimation, *Proc. IEEE*, *92*(3), 401–422, doi:10.1109/JPROC.2003.823141.

Li, H., A. Robock, S. Liu, X. Mo, and P. Viterbo (2005), Evaluation of reanalysis soil moisture simulation using updated Chinese soil moisture observations, *J. Hydrometeorol.*, *6*, 180–193, doi:10.1175/JHM416.1.

Mitchell, K. E., et al. (2004), The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *J. Geophys. Res.*, *109*, D07S90, doi:10.1029/2003JD003823.

Nijssen, B., R. Schnur, and D. P. Lettenmaier (2001), Global retrospective estimation of soil moisture using the variable infiltration capacity land surface model, 1980–93, *J. Clim.*, *14*, 1790–1808, doi:10.1175/1520-0442(2001)014<1790:GREOSM>2.0.CO;2.

Oleson, K. W., et al. (2004), Technical description of the community land model (CLM), *Tech. Note NCAR/TN-461+STR*, 186 pp., Natl. Cent. for Atmos. Res., Boulder, Colo.

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, doi:10.1175/JHM540.1.

Robock, A., K. Y. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai (2000), The Global Soil Moisture Data Bank, *Bull. Am. Meteorol. Soc.*, *81*(6), 1281–1299, doi:10.1175/1520-0477(2000)081<1281:TGSMDB>2.3.CO;2.

Robock, A., et al. (2003), Evaluation of the North American Land Data Assimilation System over the southern Great Plains during the warm season, *J. Geophys. Res.*, *108*(D22), 8846, doi:10.1029/2002JD003245.

Saulson, B., and K. C. Chang (2004), Comparison of nonlinear estimation for ballistic missile tracking, *Opt. Eng.*, *43*(6), 1424–1438, doi:10.1117/1.1738431.

Sheffield, J., and E. F. Wood (2008), Global trends and variability in soil moisture and drought characteristics, 1950–2000, from observation-driven simulations of the terrestrial hydrologic cycle, *J. Clim.*, *21*, 432–458, doi:10.1175/2007JCLI1822.1.

Tian, X., A. Dai, D. Yang, and Z. Xie (2007), Effects of precipitation-bias corrections on surface hydrology over northern latitudes, *J. Geophys. Res.*, *112*, D14101, doi:10.1029/2007JD008420.

U.S. National Research Council (1994), *GOALS (Global Ocean-Atmosphere-Land System) for Predicting Seasonal-to-International Climate*, 103 pp., Natl. Acad. Press, Washington, D. C.

van der Merwe, R. (2004), Sigma-point Kalman filters for probabilistic inference in dynamic state-space models, Ph.D. thesis, Oreg. Grad. Inst. Sch. of Sci. and Eng., Oreg. Health and Sci. Univ., Portland.

van der Merwe, R., and E. A. Wan (2004), Sigma-point filters for integrated navigation, paper presented at 60th Annual Meeting, Inst. of Navig., Dayton, Ohio.

Wan, E. A., and A. T. Nelson (2001), Dual extended Kalman filter methods, in *Kalman Filtering and Neural Networks*, edited by S. Haykin, pp. 123–280, John Wiley, Hoboken, N. J.

Wan, E. A., and R. van der Merwe (2001), The unscented Kalman filter, in *Kalman Filtering and Neural Networks*, edited by S. Haykin, pp. 221–280, John Wiley, Hoboken, N. J.

Zhang, S., Z. Xie, X. Tian, S. Chunxiang, and F. Chen (2006), A soil moisture assimilation scheme with an unsaturated soil water flow model and in-site observation (in Chinese), *Adv. Earth Sci.*, 1350–1362.

A. Dai, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000, USA. (adai@ucar.edu)

X. Tian and Z. Xie, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China. (tianxj@mail.iap.ac.cn; zxie@lasg.iap.ac.cn)