Simultaneous estimation of both soil moisture and model parameters using particle filtering method through the assimilation of microwave signal

Jun Qin,¹ Shunlin Liang,² Kun Yang,¹ Ichiro Kaihotsu,³ Ronggao Liu,⁴ and Toshio Koike⁵

Received 25 October 2008; revised 10 March 2009; accepted 3 June 2009; published 5 August 2009.

[1] Soil moisture is a very important variable in land surface processes. Both field moisture measurements and estimates from modeling have their limitations when being used to estimate soil moisture on a large spatial scale. Remote sensing is becoming a practical method to estimate soil moisture globally; however, the quality of current soil surface moisture products needs to be improved in order to meet practical requirements. Data assimilation (DA) is a promising approach to merge model dynamics and remote sensing observations, thus having the potential to estimate soil moisture more accurately. In this study, a data assimilation algorithm, which couples the particle filter and the kernel smoothing technique, is presented to estimate soil moisture and soil parameters from microwave signals. A simple hydrological model with a daily time step is utilized to reduce the computational burden in the process of data assimilation. An observation operator based on the ratio of two microwave brightness temperatures at different frequencies is designed to link surface soil moisture with remote sensing measurements, and a sensitivity analysis of this operator is also conducted. Additionally, a variant of particle filtering method is developed for the joint estimation of soil moisture and soil parameters such as texture and porosity. This assimilation scheme is validated against field moisture measurements at the CEOP/Mongolia experiment site and is found to estimate near-surface soil moisture very well. The retrieved soil texture still contains large uncertainties as the retrieved values cannot converge to fixed points or narrow ranges when using different initial soil texture values, but the retrieved soil porosity has relatively small uncertainties.

Citation: Qin, J., S. Liang, K. Yang, I. Kaihotsu, R. Liu, and T. Koike (2009), Simultaneous estimation of both soil moisture and model parameters using particle filtering method through the assimilation of microwave signal, J. Geophys. Res., 114, D15103, doi:10.1029/2008JD011358.

1. Introduction

[2] Soil moisture plays a significant role in the terrestrial water cycle [Daly and Porporato, 2005; Hirabayashi et al., 2005; Reichle et al., 2007; Sheffield and Wood, 2007]. It is very important to obtain information about soil moisture due to its profound impacts on practical water resource applications such as flood forecasting, weather and climate prediction, crop growth monitoring, and water resource management [Claussen, 1998; Davies and Allen, 1973; Drusch, 2007; Drusch and Viterbo, 2007; Foley, 1994; Schmugge et al., 2002; Teixeir et al., 1997]. There are two common methods to obtain the soil moisture status [Moradkhani, 2008]. One is to measure it in the field with instruments. These measurements are merely representative over a small spatial scale since the soil moisture has large spatial heterogeneity. It is not practical to densely install many instruments on a large scale. The other is to simulate soil moisture by running land surface models (LSMs) with meteorological data and other parameters as inputs. The simulated soil moisture performs well when both the model parameters and meteorological forcing are known with a high degree of precision and accuracy. This can be realized at only a very limited number of sites, where a variety of measurement instruments are installed. When running the model on a large scale, it is very difficult to accurately obtain model inputs and parameter values.

[3] Microwave remote sensing data has offered another means to map land surface soil moisture on a large scale [Kerr et al., 2001; Njoku et al., 2003; Wagner et al., 2003]. However, it also has many limitations and thus mapping
results cannot satisfy the practical requirements. Data assimilation (DA) methods can consistently couple both modeling and observations and thus yield superior soil moisture retrievals [Entekhabi et al., 1994; Galantowicz et al., 1999; Houser et al., 1998; Margulis et al., 2002; McLaughlin, 2002; Reichle et al., 2001; Walker and Houser, 2001]. Thus it has attracted much attention from researchers in many fields.

[4] Data assimilation techniques can generally be divided into two categories: sequential-based and cost-function-based methods. Sequentially based methods [Bertino et al., 2003], especially those based on a Monte Carlo approach such as Ensemble Kalman Filtering (EnKF [Evensen, 2003]), are currently popular in land data assimilation research and applications since they can be applied to nonlinear and discontinuous models and be realized easily [Huang et al., 2008a, 2008b]. This is especially important in land surface data assimilation because there are many land surface parameterizations which are not continuous or not differentiable and this makes it difficult or inefficient to use cost-function-based methods as assimilation algorithms. In addition, the cost-function-based method cannot directly consider uncertainties in atmospheric inputs which are used to drive a LSM. It just treats uncertainties included in inputs as one part of model noise. These uncertainties can, however, be handled readily in the sequential methods [Liang and Qin, 2008].

[5] Reichle et al. [2002] used EnKF to assimilate L-band (1.4 GHz) microwave brightness temperature observations into a LSM. Their research indicated that the EnKF is a flexible and robust DA option that gives satisfactory estimates even for moderate ensemble sizes although the updating process is suboptimal. Crow [2003] and Crow and Wood [2003] applied EnKF to assimilate L-band microwave data to correct for the impact of poorly sampled rainfall on land surface modeling of root-zone soil moisture and surface energy fluxes. The results suggested that the EnKF-based assimilation system is capable of correcting a substantial fraction of model errors in root-zone soil moisture and latent heat flux predictions associated with the use of temporally sparse rainfall measurements as the forcing data. Ni-Meister et al. [2006] assimilated retrieved soil surface moisture from Scanning Multichannel Microwave Radiometer (SMMR) data using EnKF. Reichle et al. [2007] applied EnKF to assimilate retrieved soil surface moisture from the Advanced Microwave Scanning Radiometer-Earth Observing System (EOS) (AMSR-E) as observations into a LSM. Comparisons were also performed between EnKF and other Monte Carlo-based filtering methods [Zhou et al., 2006].

[6] Most of the studies mentioned above assimilated microwave brightness into a LSM with an hourly or sub-hourly time step rather than with a daily time step. It is because the microwave radiative transfer equation (RTE) as the observation operator requires instantaneous soil surface and canopy temperatures as inputs, which have the apparent diurnal variations, but the daily-based model lacks such a temporal resolution. It is obvious that significant computational cost could be saved if a daily-based model is used in the process of assimilation. Furthermore, most of previous land surface assimilation studies focus on either retrieving state variables such as soil moisture or estimating some model parameters independently. Moreover, some aforementioned studies assimilated L-band brightness temperature, which has not been available for large spatial regions. Other studies assimilated retrieved soil surface moisture from AMSR-E and/or SMMR data at a continental scale. In addition, all these studies assumed that soil texture data or hydraulic properties are available, although it is rather difficult to obtain their accurate values at a large scale. Few investigations [Moradkhani et al., 2005a; Yang et al., 2007] put forward the idea of jointly retrieving state variables and model parameters, and perform assimilation experiments using a conceptual rainfall-runoff model and a complex LSM, respectively.

[7] In this study, a simple model is used to characterize the water movement in soils with a daily-based time step, an observation operator is designed to link the AMSR-E microwave signal and soil surface moisture, and a variant of particle filtering method is used to simultaneously estimate soil surface moisture and soil parameters such as texture, and porosity, and surface parameters. Then, the whole DA scheme is validated against the field measurements. In this paper, the DA scheme is first described. Validation results are then presented and finally followed by discussions and conclusions.

2. Data Assimilation Scheme

[8] A DA system consists of four parts: model dynamics, observation operator, assimilation algorithm, and error models [Lermusiaux and Robinson, 2001]. In the following subsections, details of these four parts are presented.

2.1. Land Surface Water Balance Model

[9] A land surface scheme to model the water balance on a daily basis is simplified from the Simple Biosphere Model 2 (SiB2 [Sellers et al., 1996]). We aim to develop the data assimilation system mainly for arid or semiarid areas; thus the interception storage of the canopy can be ignored, since the leaf area index (LAI) normally peaks around 1.5. The soil column is vertically divided into three layers: surface layer, root zone layer, and recharge zone. The governing equations characterizing the water movement in the soil are as follows:

\[ \frac{\partial \theta_1}{\partial t} = \frac{1}{D_1} \left[ P_t - Q_{1,2} - \frac{1}{\rho_w} E_g \right], \]

\[ \frac{\partial \theta_2}{\partial t} = \frac{1}{D_2} \left[ Q_{1,2} - Q_{2,3} - \frac{1}{\rho_w} E_{gr} \right], \]

\[ \frac{\partial \theta_3}{\partial t} = \frac{1}{D_3} \left[ Q_{2,3} - Q_3 \right], \]

where \( \theta_i \) is volumetric soil moisture content of each layer, \( D_i \) the soil thickness of each layer, \( P_t \) is the precipitation, \( Q_{1,2}, Q_{2,3}, \) and \( Q_3 \) are soil water fluxes between layers and out of the bottom layer, \( E_g \) and \( E_{gr} \) are evaporation from the soil surface and transpiration from the vegetation canopy, respectively, and \( \rho_w \) is the water density. Equations (1)–(3)
are discretized on a daily basis using an implicit difference scheme. The formulas for soil water flux are as follows:

\[
Q_{i,j+1} = \left( \frac{\psi_i - \psi_{i+1}}{0.5(D_i + D_{i+1})} + 1 \right) \left( \frac{K_i \psi_i - K_{i+1} \psi_{i+1}}{\psi_{i+1} - \psi_i} \right) \left( \frac{B}{B + 3} \right),
\]

(4)

\[
\psi_i = \psi_{s,i} \left( \frac{\theta_i}{\theta_{sat}} \right)^{-B},
\]

(5)

\[
K_i = K_{s,i} \left( \frac{\theta_i}{\theta_{sat}} \right)^{2B+3},
\]

(6)

\[
K_{sat} = 7.0556 \cdot 10^{-6.884+0.0153 \%clay}
\]

(7)

\[
B = 2.91 + 0.159 \cdot \%clay
\]

(8)

where \( K_i \) is the hydraulic conductivity of each layer, \( \psi_i \) the matrix potential of each layer, \( \theta_{sat} \) the soil porosity, \( K_{sat} \) the hydraulic conductivity at saturation, and \( B \) the empirical parameter related to soil texture, \%sand the sand content, and \%clay the clay content. The drainage out of the bottom layer is assumed to be \( K_3 \) and the surface runoff occurs when surface soil water content \( \theta_1 \) exceeds the porosity \( \theta_{sat} \).

[10] Both the evaporation \( E_g \) from the soil surface and transpiration \( E_t \) from the vegetation canopy are the important components in equations (1)–(3). There exist many methods to compute the potential evapotranspiration on a daily basis, including Penman-Monteith, Priestley-Taylor, and so on. The actual evapotranspiration and its partition into evaporation and transpiration are needed in the calculation of the water balance. In this study, a variant of the Priestley-Taylor equation [Davies and Allen, 1973] is taken to estimate the daily actual evapotranspiration and the vegetation coverage is used to separate it. The daily evapotranspiration process is parameterized as follows [Sau et al., 2004]:

\[
ET_a = \alpha \frac{\Delta + \gamma}{\Delta + \gamma + R_0} \cdot \left\{ 1 - \exp \left[ \beta \left( \frac{\theta_1}{\theta_{sat}} \right)^3 \right] \right\},
\]

(9)

\[
E_g = ET_a \cdot (1 - f_v),
\]

(10)

\[
E_t = ET_a \cdot f_v,
\]

(11)

\[
f_v = 1 - \exp(-0.5LAI),
\]

(12)

where \( ET_a \) denotes the actual evapotranspiration, \( \alpha = 1.26 \), the Priestly-Taylor constant, \( \beta \) the constraint coefficient, \( \Delta \) the slope of the saturated vapor pressure with respect to the air temperature, \( \gamma \) the psychometric constant, and \( f_v \) the vegetation coverage. The power 3 does not exist in its original form of equation (9) and it is found that equation (9) performs better after adding an exponent of 3 [Nakayama et al., 1993].

[11] An implicit scheme is used for the computation with daily time step and is stable, but big errors may occur immediately after a rainfall event. Nevertheless, such errors may be compensated to some degree by information from satellite signals.

2.2. Observation Operator and Microwave Data

[12] In this study, a microwave RTE is implemented to link the surface soil moisture to satellite measurements. It is a Q-h model with minor revisions to include vegetation effect. The concrete form of this RTE is as follows:

\[
T_{bp} = T_g(1 - \Gamma_p) \exp(-\tau_v) + T_q[1 - \exp(-\tau_v)] [1 + \Gamma_p \exp(-\tau_v)],
\]

(13)

where the subscript \( p \) represents the vertical or horizontal polarization, \( \Gamma_p \) the soil reflectivity, \( \tau_v \) the vegetation optical depth, and \( \omega \) the vegetation single scattering albedo. A Q-h model is used to calculate the soil reflectivity as follows:

\[
\Gamma_p = [(1 - Q)R_p + QR_q] \exp(-h),
\]

(14)

where the subscripts \( p \) denotes the vertical or horizontal polarization, respectively, \( Q \) and \( h \) are empirical surface roughness parameters, and \( R \) the Fresnel power reflectivity with a smooth soil surface. The \( R \) is determined using the following equations:

\[
R_p = \frac{\cos \gamma - \sqrt{\varepsilon - \sin^2 \gamma}}{\cos \gamma + \sqrt{\varepsilon - \sin^2 \gamma}}^2,
\]

(15)

\[
R_q = \frac{\varepsilon \cos \gamma - \sqrt{\varepsilon - \sin^2 \gamma}}{\varepsilon \cos \gamma + \sqrt{\varepsilon - \sin^2 \gamma}}^2,
\]

(16)

where \( \varepsilon \) and \( \gamma \) denote the horizontal and vertical polarization, \( \gamma \) the incident angle, and \( \varepsilon \) the soil dielectric constant. The soil dielectric constant is computed as follows:

\[
\varepsilon = \left[ 1 + (1 - \theta_{sat})(\varepsilon_s^a - 1) + \theta_{sat}^a \varepsilon_s^a - \theta_1 \right]^{1/\alpha},
\]

(17)

where \( \varepsilon_s = 4.7 + 0.0j \) denotes the dielectric constant for mineral soil, \( \varepsilon_{fw} \) the dielectric constant of free water, \( a = 0.65 \), and \( b \) the coefficient dependent upon the soil texture. The parameters in equations (13) and (14) are dependent on wave frequency and can be parameterized as follows:

\[
h = (k \cdot s)^{0.1 \cos \gamma},
\]

(18)

\[
Q = Q_0 (k \cdot s)^{0.795},
\]

(19)
\[
\tau_c = \frac{b'(100\lambda)^3w_c}{\cos \theta},
\]
(20)

\[
w_c = \exp(LAI/3.3) - 1
\]
(21)

\[
\omega = \frac{0.00083}{\lambda},
\]
(22)

where \(\lambda\) [m] is the wavelength, \(k\) the wave number defined as \(2\pi/\lambda\), \(s\) the standard deviation of surface roughness, \(w_c\) [kg m\(^{-2}\)] the vegetation water content, and \(Q_o, b',\) and \(\chi\) the empirical coefficients. Equations (19) and (22) are empirical formulas fitted from limited microwave experimental data [Fujii, 2005] and were firstly used by Yang et al. [2007] to reduce the parameter number of the radiative transfer equation.

As shown in equation (13), both the soil surface temperature and the canopy temperature are important factors to determine the value of the simulated microwave brightness temperature. Since the time step is one day in this model, the brightness temperature cannot be simulated at the satellite overpassing times. So it is challenging to directly assimilate the satellite-observed brightness temperature into the dynamics. However, this dilemma can be removed by defining a ratio of two brightness temperatures at different frequencies, which can reflect the soil wetness, based on the assumption that the soil surface temperature and the canopy temperature have equal values. When AMSR-E data are used as the information source to be assimilated, a new index called the soil water ratio (SWR) is defined in this study in accordance with equation (13) by canceling out temperatures as follows:

\[
\text{SWR} = \frac{T_{\text{bv}}^{18.7}}{T_{\text{bv}}^{6.9}},
\]

where the superscript denotes the frequency and \(T_{\text{bv}}^{18.7}\) and \(T_{\text{bv}}^{6.9}\) the vertical polarization brightness temperatures at 18.7 GHz and 6.9 GHz. The reason for choosing the vertical polarization temperatures in (23) is that they are less sensitive to vegetation heterogeneity than the horizontal polarization temperatures. Up to now, there are still two problems not to be resolved. One is that SWR does not completely eliminate the influence of temperature since \(\Gamma_q\) in equation (23) is a function of the dielectric constant of free water which is in turn dependent on the soil temperature. It is, however, found that SWR is not sensitive to the soil surface temperature and thus this temperature can be replaced with the daily averaged air temperature. The other is the issue of whether the defined SWR really has the capacity to reflect the soil wetness. The solutions to these two problems will be shown in section 4.

2.3. Assimilation Algorithm

Mainstream sequential-based methods include Kalman Filtering (KF) and Particle Filtering (PF), and their variants. PF is also called sequential Monte Carlo filtering.

It has been applied in many engineering fields and attracted some data assimilation practitioners since the posterior distribution of the state vector can be represented with Monte Carlo samples and the Gaussian assumption can be avoided. KF and its variants, however, just evaluate the mean and covariance of the posterior distribution. Thus PF can better grasp the filtering density evolution of the nonlinear system in time than KF and its variants do. PF itself also has many variants such as the sampling importance resampling filter (SIR). Han and Li [2008] make a detailed evaluation of PF, KF, and their variants and conclude that PF is suited to applications in land surface data assimilation according to both effectiveness and efficiency. Before moving on, some notations are introduced to facilitate the following discussions.

\[
x_{t+1} = f(x_t, u_t, \xi_t) + v_t
\]
(24)

where \(x\) denotes the model state vector, \(u\) the external forcing data, \(\xi\) the model parameter vector, \(v\) the model noise, \(t\) the subscript for time step, and \(f()\) the model operator mapping the previous state \(x_t\) to the next state \(x_{t+1}\).

\[
y_t = h(x_t, \xi_t) + e_t
\]
(25)

where \(y\) denotes the observation vector, \(e\) the observation noise, and \(h()\) the observation operator. In this study, the state vector \(x = [\theta_1, \theta_2, \theta_3]^T\), external forcing \(u = [P_t, R_n, T_a, LAI]^T\), \(\xi = [\%\text{sand}, \%\text{clay}, \theta_{\text{sat}}, \beta, s, Q_o, b', \chi]^T\), \(f()\) is the discrete form from equations (1)–(3), and \(h()\) is the equation (23).

In the framework of sequential filtering techniques, the joint estimation of state variables and model parameters can be performed through the state augmentation method [Chen et al., 2005]. This approach regards model parameters to be estimated as part of the state vector. The new augmented state vector becomes \([x', \xi']^T\). Conventionally, the random walk model is assumed for the time evolution of \(\xi\) [Moradkhani et al., 2005b]. However, this model results in much larger variances of \(\xi\) than actual ones in the estimation process. The kernel smoothing technique is currently introduced to remove this feature.

In this study, both the SIR method and kernel smoothing technique are combined [Chen et al., 2005] to merge the dynamics and the observations for simultaneous estimation of the model states and parameters. Main steps for the entire assimilation algorithm [Thomas, 2006] are as follows:

Step 1: for \(t = 0\), sample \([\xi_0^{(i)}, \xi_0^{(i)}]_{i=1}^N : p(\xi_0)\) and \([(x_0^{(i)})]_{i=1}^N : p(x_0);\)

Step 2: draw \([\xi_{t+1}^{(i)}]_{i=1}^N \sim (N(\xi_{t+1}^{(i)}))_{i=1}^N : h^2 \cdot V_t\), where \(m_t^{(i)} = (1 - F_t)^{1/2})_{i=1}^N\), \(V_t\) denotes the covariance matrix of \(\xi_t\), and \(s\) an adjustable parameter;
Step 3: draw \{x^{(0)}_{r+1}, y^{(0)}_{r+1}\} \sim p(x_r + u^{(0)}_r, \zeta^{(0)}_{r+1})\), where \(x^{(0)}_{r+1} = f(x^{(0)}_r, u^{(0)}_r, \zeta^{(0)}_{r+1})\) + \(v^{(0)}_r\), \(v^{(0)}_r\) : \(p(v_r)\);

Step 4: compute weights \(w^{(0)}_{r+1} = \frac{p(y_{r+1} | x^{(0)}_{r+1}, \zeta^{(0)}_{r+1})}{\sum \frac{p(y_{r+1} | x^{(0)}_r, \zeta^{(0)}_{r+1})}{p(y_{r+1} | x^{(0)}_r, \zeta^{(0)}_{r+1})}}\), where \(p(y_{r+1} | x^{(0)}_{r+1}, \zeta^{(0)}_{r+1})\) denotes the value of \(p(y_{r+1} | (x^{(0)}_{r+1}, \zeta^{(0)}_{r+1}))\);

Step 5: resample \(\{x^{(0)}_{r+1}, \zeta^{(0)}_{r+1}\}_{i=1}^N\) with replacement according to weights \(w^{(0)}_{r+1}\) in order to get \(\{x^{(0)}_{r+1}, \zeta^{(0)}_{r+1}\}_{i=1}^N\) with weights \(\{1/N\}_{i=1}^N\);

Step 6: set \(r = r + 1\) and go to step 2.

[17] The parameter \(l\) is an adjustable constant for considering that the model parameters to be estimated are to change quickly or slowly. \(l\) is set to \(0 < l \leq 0.2\) for slowly varying parameters and \(0.8 < l \leq 1.0\) for rapidly varying parameters. In this work, only variances of each model parameter are computed and nondiagonal elements are set to zero for the covariance matrix \(V_r\) in Step 2 above. Thus different values of \(l\) can be easily set independently for determination of different parameters.

2.4. Error Models

[18] One significant advantage of sequential-based data assimilation methods is that uncertainties, which are from model structure, model parameters, and inputs, etc., can be handled explicitly in their own framework. As shown in above subsections, there are four error sources, including errors in the model dynamics, observations, input forcing, and model parameters, respectively. All of these errors can be divided into biased and unbiased noises. Only the unbiased part in these errors is taken into account in this study. The error model must be specified for each error source.

[19] As summarized by Hamill [2006], four methods can be applied to parameterize uncertainties in model dynamics. The so-called covariance inflation approach is used and coupled with the assimilation algorithm. Before assimilating observation information into dynamics, deviations of particles around their mean are inflated by a factor \(r\), which is a little bit greater than 1.0, as follows:

\[
x^{(0)}_i' \leftarrow r \left( x^{(0)}_i - \mu^{(0)}_i \right) + \mu^{(0)}_i
\]

where the operation \(\leftarrow\) denotes the replacement of the previous value of \(x^{(0)}_i\). It is found that a moderate inflation improves the assimilation accuracy. More details can be obtained in the work of Hamill [2006]. One advantage of sequential data assimilation methods over variational ones is that errors in input forcing can be explicitly considered by adding perturbations to them according to some error parameterizations. In this work, error models for inputs are taken as the following uniform probability distribution function (PDF):

\[
u^{(a)}_i = u_i(\zeta) \sim U[-\delta, +\delta]
\]

where \(\zeta\) obeys the uniform distribution and \(\delta\) reflects the knowledge of inputs. Different values of \(\delta\) can be assigned to each component in \(u = \{Pt, Rn, Ta, LAI\}^{T}\). Similarly, uniform distributions are also assumed for the parameter error models. The observation error model is assumed to be Gaussian.

3. Sensitivity Study of Soil Water Ratio

[20] As mentioned in section 2.2, two issues with SWR need to be addressed. As for the first problem, the variance-based sensitivity analysis method is applied to determine the global sensitivity of SWR to each input parameter in equation (23). The variance-based sensitivity method is briefly introduced below [Helton et al., 2006].

[21] The entire SWR formula can be denoted by \(Y = h(X)\) where \(Y\) means SWR, \(X\) the input parameters, and \(h(\cdot)\) the equation in the previous section. If each component of vector \(X = [x_1, x_2, \ldots, x_N]^T\) is considered to be an independent random variable, then the variance \(V_Y\) of \(Y\) can be decomposed and expressed as:

\[
V_Y = \sum_i V_i + \sum_{i < j} V_{ij} + \sum_{i < j < k} V_{ijk} + \cdots + V_{12\ldots N} \tag{28}
\]

where \(V_i\) is the contribution of \(x_i\) to \(V_Y\), \(V_{ij}\) the contribution of the interaction of \(x_i\) and \(x_j\) to \(V_Y\) and so on up to \(V_{12\ldots N}\) which is the contribution of the interaction of \(x_1, x_2, \ldots, x_N\) to \(V_Y\). Two types of sensitivity indices can be defined as:

\[
S_i = \frac{V_i}{V_Y}\tag{29}
\]

and

\[
S_{ij} = \frac{V_{ij}}{V_Y}
\]

where \(V_{ij}\) is the sum of all variance terms which do not include the index \(i\). \(S_i\) is the first-order sensitivity index for the \(i\)th parameter. This index characterizes the main influence of parameter \(x_i\) on the output variable \(Y\) and measures the variance reduction that would be achieved by fixing that parameter. \(S_{ij}\) is the total sensitivity index for the \(i\)th parameter and measures the sum of all effects related to this parameter, considering the interaction between the \(i\)th parameter and other ones.

[22] The Monte Carlo based method can be used to evaluate these indices. The computation steps are as follows:

Step 1: generate two sets of random samples according to given distributions to inputs \(X\)

\[
X^{(a)}_i = \left[ x^{(a)}_1, x^{(a)}_2, \ldots, x^{(a)}_N \right], \quad i = 1, 2, \ldots, nS
\]

and

\[
X^{(b)}_i = \left[ x^{(b)}_1, x^{(b)}_2, \ldots, x^{(b)}_N \right], \quad i = 1, 2, \ldots, nS
\]

\[\text{5 of 13}\]
in which \( n_S \) is the number of samples.

Step 2: estimate the sample mean and variance of \( Y \) as

\[
\hat{E}_Y = \frac{1}{n_S} \sum_{i=1}^{n_S} \bar{h}(\bar{x}^{(a)}_i)
\]

and

\[
\hat{V}_Y = \frac{1}{n_S} \sum_{i=1}^{n_S} \bar{h}^2(\bar{x}^{(a)}_i) - \hat{E}_Y^2
\]

Step 3: calculate some intermediate parameters as

\[
\hat{V}_i = \frac{1}{n_S} \sum_{p=1}^{n_S} \bar{h}(\bar{x}^{(a)}_p) \bar{h}(\bar{x}^{(b)}_p) - \hat{E}_Y^2
\]

and

\[
\hat{V}_{\sim i} = \frac{1}{n_S} \sum_{p=1}^{n_S} \bar{h}(\bar{x}^{(a)}_p) \hat{h}(\bar{x}^{(b)}_p) - \hat{E}_Y^2
\]

Step 4: evaluate sensitivity indices as follows:

\[
S_i = \frac{\hat{V}_i}{\hat{V}_Y}
\]

and

\[
S_{T_i} = (1 - \hat{V}_{\sim i}/\hat{V}_Y) \hat{V}_Y
\]

The sensitivity of all input parameters in equation (23) to SWR can be evaluated according to the above global sensitivity analysis algorithm. The distributions of these input parameters are listed in Table 1. Since there is no information on these parameters, the uniform distribution is assumed for them. The sensitivity analysis results are shown in Table 1. This sensitivity analysis answers two questions raised in section 2.2. It is found that three of the most sensitive input parameters in the calculation of SWI are leaf area index, soil surface moisture, and \( \chi \) in equation (20), respectively. Other parameters merely have very small effects on SWI, including the soil surface temperature. Thus it is reasonable to construct SWI as the assimilated data.

### 4. Determination of Model Parameters

[23] There are a total of eight model parameters \( \xi = [\%\text{sand}, \%\text{clay}, \theta_{\text{sat}}, \beta, s, Q_0, b', \chi] \) in the model operator and observation operator. Not all of the parameters can be estimated in terms of the sensitivity analysis performed above through the data assimilation algorithm presented in section 2.3. A careful analysis is needed to determine which parameters should be estimated together with model states.

[24] Soil parameters \( \%\text{sand}, \%\text{clay}, \) and \( \theta_{\text{sat}} \) have very small influences on SWI, but they highly affect the soil water movement as shown in equations (1)–(8) and then influence the value of SWI at the subsequent instant through surface soil moisture \( \theta_t \). Thus it is possible to retrieve \( \%\text{sand}, \%\text{clay}, \) and \( \theta_{\text{sat}} \) step-by-step in time. In fact, there exist parameterization schemes to estimate \( \theta_{\text{sat}} \) from the soil texture. However, they are not used in this study since these schemes lack sufficient accuracy and therefore \( \theta_{\text{sat}} \) is independently estimated.

[25] Parameters \( s, Q_0, \) and \( b' \) have no apparent impacts on SWI in accordance with sensitivity analysis results. At the same time, they are also not similar to \( \%\text{sand}, \%\text{clay}, \) and \( \theta_{\text{sat}} \) which can influence the subsequent soil surface moisture and in turn the SWI. So their values can be fixed to median ones between maximum and minimum values as listed in Table 1. Since \( \chi \) can affect the SWI to some degree, it needs to be estimated.

[26] Many investigations [Castellvi et al., 2001; Kustas et al., 1996] indicate that the Priestly-Taylor constant varies in different situations and does not always keep the value of 1.26. Some researchers present modifications to the original Priestly-Taylor formula. Equation (9) is one of them, which introduces the soil water content as a limiting factor. The original value of the parameter \( \beta \) is estimated to be \(-10.563\) by fitting equation (9) to some data sets. However, it is not guaranteed that this value of \( \beta \) can be applicable to other cases. Thus it may improve the application scope by taking \( \beta \) as a parameter to be estimated.

### 5. Experiment Site and Data

[27] The CEOP/Mongolia experiment site located at Mandalgobi of Mongolia covers a flat area of 120 km × 160 km, where 12 long-term Automatic Stations for Soil Hydrology (ASSH) and 6 Automatic Weather Stations (AWS) are deployed. Their geographic locations are shown in Figure 1. The detailed description of this experiment can be found in the work of Kaihotsu [2005]. In this study, the experimental period (a total of 153 days), was chosen from 1 May 2003 to 30 September 2003 because the soil freezing and thawing processes are not parameterized in the system.

[28] At ASSH, soil temperature and water content were measured at 3 cm and 10 cm depth. Meteorological parameters were observed at AWS, including wind, temperature, humidity, pressure, precipitation, net radiation, and soil temperature and moisture profiles. However, two AWS (TDS and CRS) data sets are not archived in the CEOP project. In this article, near-surface water content data comes from 12 ASSH and 4 AWS (BTS, DGS, DRS, and MGS). The meteorological forcing data with a temporal resolution of 3 hours, including precipitation, net radiation,
and near-surface air temperature, were extracted from the Global Land Data Assimilation System (GLDAS Rodell et al., 2004)). The microwave data used in this study are the brightness temperatures of AMSR-E 6.9 GHz and 18.7 GHz vertical polarization. The MODIS LAI product was chosen as input to the dynamic model in this paper.

Since GLDAS meteorological data has a spatial resolution of \( \frac{1}{176} \), roughly matching the experiment site area, the mean of 16 soil moisture measurements was used to represent the integral soil water status and compared to the retrieved values through the data assimilation system built in section 3. In addition, the area-average values of brightness temperatures and LAIs are also calculated for the same region. As discussed above, the time step of the dynamics is daily. Consequently, daily averaging is performed on the in situ soil moisture observations and GLDAS meteorological data.

### 6. Results and Discussions

Numerical experiments are performed to validate results from this assimilation system against moisture estimates from a reference run and from other methods. Furthermore, issues about the parameter estimation are also addressed. Before moving on, concrete settings for uncertainties used in the following numerical experiments are presented, although their parameterization forms are introduced in the previous sections. PDFs for initial moistures \([\theta_1, \theta_2, \theta_3] \) in three soil layers are all subject to the same uniform distribution \([0.05, 0.55] \). Initial PDFs for parameters to be estimated are listed in Table 1. The parameter \( \delta \) is set to 0.4 in this study. Controlling constants \( l \) for slowing varying parameter and rapidly one are set to 0.12 and 0.9, respectively. The parameter \( r \) controlling inflation of particles is set to 1.01. The observation noise \( e \) conforms to the normal distribution \( N(0, 0.008) \). In the reference run, all parameters are set as median values in their respective ranges.

#### 6.1. Comparison With Other Moisture Retrievals

As many investigations have shown, the size of ensembles or particles has a significant impact on the final assimilation results when Monte Carlo-based assimilation methods are applied. When relatively small particle size is used in the assimilation computation, the results display instability and contain large sampling errors although the computation burden can be reduced. Conversely, the assimilation results keep stable and their errors decline when large particle size is used. The computational intensity is acceptable for assimilation experiments at some sites when large particle size is selected, but this can be unaffordable for applications at a large scale. Thus the balance point needs to

---

There were 6 Automatic Weather Stations (AWS), 12 Automatic Stations for Soil Hydrology (ASSH). Data at two AWS (TDS, CRS) were not archived in CEOP.

![Figure 1. Schematic of the experiment site and locations of 6 AWS and 12 ASSH at CEOP/Mongolia.](image)
be found for the particle size in the assimilation practice, around which both the computational cost and assimilation precision are all acceptable. For this purpose, a series of trials are performed using different particle sizes in accordance with assimilation settings in preceding sections. The analysis indicates that a particle size of 500 is large enough to achieve stable results. In the following, all experiments and analyses are conducted based upon a 500-particle size.

In order to verify the effectiveness of the data assimilation algorithm presented above, the reference run is performed and then resultant soil surface moistures are compared with ones from the assimilation system in Figure 2. GLDAS precipitation data is also depicted and represented by black bars in Figure 2. As have been shown, the assimilation algorithm obviously improves the simulation results in comparison with ones from the reference run. RMSE drops from 0.057 to 0.033 and correlation coefficient R rises from 0.57 to 0.72, respectively. The assimilation results do better agree with field measurements than the reference ones do. This can be explained in accordance with the following uncertainty sources. The first is the uncertainty in the meteorological forcing, especially in precipitation as shown in Figure 3. GLDAS data grasps main precipitation events during the data assimilation period through the analysis of observed soil surface water content denoted by blank circles in Figure 2. However, the amount of precipitation in each precipitation event could contain large errors, causing the reference run to poorly model the surface soil moisture, but the response of the soil moisture to the precipitation event looks reasonable. Other forcing data also contain more or less uncertainties. The second source is the model structure itself and the third is uncertainties included in model parameters. Effects of these three sources can be alleviated in the framework of the current assimilation algorithm due to the constraint of microwave remote sensing information. Thus assimilation results are superior to those in the reference run.

Figure 4 shows the assimilation results compared with NASA AMSR-E standard surface soil moisture products. Both NASA AMSR-E daytime and nighttime products
are shown in Figure 4 and they have almost the same values. Surface soil moisture from the assimilation algorithm clearly outperforms NASA AMSR-E moisture products according to two error metrics, RMSE (Root Mean Square Error) and R (Correlation Coefficient). It is also found that NASA AMSR-E soil moisture products cannot reflect the absolute soil moisture values over the temperate and semiarid regions [Gruhier et al., 2008]. Some studies [Draper et al., 2009] indicate that the poor performance of NASA AMSR-E soil moisture is due to its retrieval algorithm. In addition, inversion algorithms retrieves soil surface moisture merely according to the instantaneous information recorded by remote sensors and other ancillary ground data, which are hard to be acquired accurately at a regional or global scale. Here the merit of the data assimilation algorithm presented in this study is obviously exhibited, though the hydrological model and the water index (SWI) are simple.

[34] The dynamic model used in the assimilation algorithm is constructed at the daily time step and the thermal process is not taken into account. This differs from the assimilation work by Yang et al. [2007]. In their work, the model SiB2 is applied as dynamics, which characterizes both hydrological and thermal processes of the land surface and integrates forward with hourly meteorological forcing data. Consequently, microwave brightness temperature is directly assimilated through a dual-pass method that first estimates optimal model parameters and then estimates soil moisture. Figure 5 shows comparisons of soil moisture between Yang’s and our algorithms at the same site. The algorithm presented in this work slightly overestimates the water content relative to Yang’s method during conditions of intensive precipitation, but the two estimates are quite comparable according to the error metrics or visually.

6.2. Retrievals of Model Parameters
[35] The retrieval of soil texture properties from soil surface information, especially the remote sensing signal, attract much attention from researchers in different fields since they play an important role in the determination of soil

Figure 4. Comparison of soil moisture between the assimilation and NASA AMSR-E retrievals at ascending pass and descending pass.

Figure 5. Comparison of daily mean soil moisture between the algorithm presented in this study and the dual-pass assimilation by Yang et al. [2007]. Note that this study uses a daily time step for integration of the LSM while et al. uses hourly time step.
hydraulic and thermal properties, which greatly affect the soil water and heat movement. However, barely effective values could be retrieved for each pixel with remote sensing information as constraints, because the variability of soil properties in both horizontal and vertical directions is rather large and there is not enough information to retrieve these heterogeneous properties.

In this work, model parameters are also estimated in addition to soil moisture, which are [%sand, %clay, \( \theta_{\text{sat}} \), \( \beta \), \( \chi \)]. In the following, the main focus is on three soil parameters %sand, %clay, and \( \theta_{\text{sat}} \) since they are physical properties of soil, from which soil parameters that control water and heat movement are estimated. These three parameters should keep stable or change slightly around some values after going through an intense adjustment at the initial stage of the whole assimilation. Moreover, twenty trials are performed in order to investigate the sensitivity of both the retrieved soil moisture and the retrieved values of %sand, %clay, and \( \theta_{\text{sat}} \) to initial values of state variables and all parameters. As shown in Figure 6, initial settings almost have no influence on retrieved surface soil moisture in accordance with two error metrics. However, the values of retrieved parameters %sand, %clay, and \( \theta_{\text{sat}} \) do not converge to certain fixed points although they are stable after the initial stage, as indicated in Figure 7. The retrieved %sand and %clay have a large variability, but retrieved soil saturation \( \theta_{\text{sat}} \) shows a small variability and stays within a reasonable range according to measured soil saturation at three different sites, MGS, DRS, and BTS. There are no soil texture data available. Thus some quantitative comparisons cannot be given to %sand and %clay retrievals. Many investigators have found that it is very difficult to retrieve soil properties by using soil surface moisture information. This could be explained by the concept of equifinality [Beven and Freer, 2001] which essentially recognizes that different initial states can lead to similar end states. According to Beven’s point of view, only a certain \textit{a posteriori} joint probability distribution could be obtained for parameters to be estimated. The more information, which is closely related to variables and parameters, is included in observations, the more uncertainties are removed in retrieved results. Different combinations of values of parameters could lead to similar observations. Each line in Figure 7 can be regarded as a realization of the posterior distribution.

7. Conclusions

Accurate estimation of soil moisture is very important since it is a key parameter in the terrestrial water cycle. Traditional methods for collecting soil moisture information cannot meet the requirements in many applications. Remote sensing has become a feasible approach to map surface soil moisture on a global scale. However, it cannot obtain soil moisture status in the root zone since the remote sensing signal generally reflects superficial land surface information. The data assimilation method can couple land surface models and remote sensing observations so that it opens up prospects for accurate estimation of the soil moisture. Research on this topic has been widely performed in a range of fields such as hydrology, agriculture, and meteorology.

In this study, a relatively simple data assimilation system is developed. This DA scheme takes the daily-based model as dynamic constraints, a new water index derived from the microwave radiative transfer as the observation operator, and a particle filter as merging scheme, and then assimilates microwave AMSR-E data to jointly estimate soil moisture and model parameters. Retrieval results from the assimilation algorithm presented in this study are compared with field measurements, retrieved soil moisture from the standard AMSR-E inversion algorithm, and a dual-pass assimilation scheme with an hourly-based LSM. Comparisons indicate that this assimilation algorithm can estimate...
soil surface moisture with satisfactory speed and precision at a daily time resolution. Particularly, the temporal variability of AMR-E product is too small compared to the observed one in our studied area. At the same time, soil hydraulic properties are also estimated. Results show that the retrieved soil texture converges to a fixed value or a narrow range in our experiments while the retrieved soil porosity is confined to a relatively narrow range. However, retrieved soil surface moisture agrees well with station-averaged moisture measurements. Soil moisture in the root zone or deep zone could not be retrieved with a high accuracy since there is no sufficient information to derive

Figure 7. Variation of retrieved soil texture and soil porosity with time, given different initial values in 20 trials.
Acknowledgments. This work was supported by the Chinese Natural Science Foundation under contract 40771129, the “100-Talent” Project of the Chinese Academy of Sciences, Chinese 863 project under contract 2006AA06A102, and Data Sharing Network of Earth System Science under contract 2005DKA32300. We thank Dr. Carol Russell for English editing. AMPEX was implemented in the framework of NASA-JRA “Ground Truth for Evaluation of Soil Moisture and Geophysical/ Vegetation parameters Related to Ground Surface Conditions with AMSR and GLI in the Mongolian Plateau” (PI: Prof. Ichiro Kaitoh, University of Hiroshima). The Mongolian partnership is among the Institute of Meteorology and Hydrology, the National Agency for Meteorology, Hydrology, and Environment Monitoring of Mongolia.

References
Bertino, L., et al. (2003), Sequential data assimilation techniques in oceano-
Beven, K. J. (2001), Equifinality, data assimilation, and uncer-
tainty estimation in mechanistic modelling of complex environmental sys-
tems using the GLUE methodology, J. Hydrol., 249, 11–29.
Castelli, F., et al. (2001), Comparison of methods for applying the Priestley-
Taylor equation at a regional scale, Hydrolo. Processes, 15, 1609–1620.
Chu, T., et al. (2005), Particle filters for state and parameter estimation in
Crow, W. T. (2003), Correcting land surface model predictions for the im-
portance of temporally sparse rainfall rate measurements using an ensemble
Kalman filter and surface brightness temperature observations, J. Hydromet-
ology, 4, 960–973.
Crow, W. T., and E. F. Wood (2003), The assimilation of remotely sensed
soil brightness temperature imagery into a land surface model using Ensemble
Daly, E., and A. Porporato (2005), A review of soil moisture dynamics:
evaporation from cropped surfaces in southern Ontario, J. Appl. Meteorol.,
12, 649–657.
Draper, C. S., et al. (2009), An evaluation of AMSR-E derived soil moisture
Drusch, M. (2007), Initializing numerical weather prediction models with
satellite derived surface soil moisture: Data assimilation experiments with
ECMWF’s Integrated Forecast System and the TMI soil moisture data set,
Drusch, M., and R. Ferraro (2007), Assimilation of screen-level variables in
ECMWF’s integrated forecast system: A study on the impact on the
forecast quality and analyzed soil moisture, Mon. Weather Rev., 135,
300–314.
Entekhabi, D., et al. (1994), Solving the inverse problem for soil moisture
and temperature anomalies by sequential assimilation of multiple, remotely
Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation
Foley, J. A. (1994), The sensitivity of the terrestrial biosphere to climatic
drought, 1950–2000: Analysis of soil moisture data from off-line simu-
ation of the terrestrial hydrological cycle, J. Geophys. Res., 110, D14101,
Houser, P. R., et al. (1998), Integration of soil moisture remote sensing and
hydrologic modeling using data assimilation, Water Resour. Res., 34,
3401–3410.
Huang, C., et al. (2008a), Retrieving soil temperature profile by assimilat-
ing MODIS LST products with ensemble Kalman filter, Remote Sens.
Environ., 112, 1320–1336.
Huang, C., et al. (2008b), Experiments of one-dimensional soil moisture
assimilation system based on ensemble Kalman filter, Remote Sens.
Kaitoh, I., T. Yamanaka, T. Koike, D. Oyanaabt, and G. Davaa (Eds.)
(2005), Ground truth for evaluation of soil moisture and geophysical/vegetation
parameters related to ground surface conditions with AMSR and GLI
Liang, S., and J. Qin (2008), Data assimilation methods for land surface
variable estimation, in Advances in Land Remote Sensing, edited by S. Liang,
Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne (2002), Land
data assimilation and estimation of soil moisture measurements from the
McLaughlin, D. (2002), An integrated approach to hydrologic data assim-
ilation: Interpolation, smoothing, and filtering, Adv. Water Resour., 25,
1275–1286.
Moradkhani, H. (2008), Hydrologic remote sensing and land surface data
assimilation, Sensors, 8, 2986–3004.
Moradkhani, H., et al. (2005a), Uncertainty assessment of hydrologic model
states and parameters: Sequential data assimilation using the particle filter,
Moradkhani, H., et al. (2005b), Dual state-parameter estimation of hydro-
logic models using ensemble Kalman filter, Adv. Water Resour., 28,
135–147.
Nakayama, K., et al. (1993), Estimation of soil moisture in the shallow root
Ni-Meister, W., et al. (2006), Soil moisture initialization for climate predic-
tion: Assimilation of scanning microwave radiometer soil moisture data into a
Njoku, E. G., et al. (2003), Soil moisture retrieval from AMSR-E, Geosci.
Reichle, R. H., et al. (2001), Downscaling of radio brightness measurements
for soil moisture estimation: A four-dimensional variational data assimilation
Reichle, R. H., et al. (2002), Hydrologic data assimilation with the ensemble
Reichle, R. H., et al. (2007), Comparison and assimilation of global soil
moisture data retrievals from the Advanced Microwave Scanning Radiometer
for the Earth Observing System (AMSR-E) and the Scanning Multichannel
Microwave Radiometer (SMMR), J. Geophys. Res., 112, D09108,
Sau, F., et al. (2004), Testing and improving evapotranspiration and soil
Resour., 25, 1367–1375.
Sellers, P. J., et al. (1996), A revised land surface parameterization (SiB2)
drought, 1950–2000: Analysis of soil moisture data from offline simu-
ation of the terrestrial hydrological cycle, J. Geophys. Res., 112, D17115,


I. Kaihotsu, Department of Natural and Environmental Sciences, Faculty of Integrated Arts and Sciences, Hiroshima University, Kagamiyama 1-7-1, Higashi-Hiroshima 739-8521, Japan.

T. Koike, Department of Civil Engineering, School of Engineering, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.

S. Liang, Department of Geography, University of Maryland, 2181 LeFrak Hall, College Park, MD 20742, USA.

R. Liu, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, No. 11A, Datun Road, Chaoyang District, Beijing 100101, China.

J. Qin and K. Yang, Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, P.O. Box 2871, Beijing 100085, China. (shuairenqin@gmail.com)