Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter

Chunlin Huang *, Xin Li, Ling Lu, Juan Gu

Cold and Arid Regions Environmental and Engineering Research Institute, CAS, Lanzhou, 730000, China

Received 30 November 2005; received in revised form 20 June 2007; accepted 30 June 2007

Abstract

Ensemble Kalman filter is a new sequential data assimilation algorithm which was originally developed for atmospheric and oceanographic data assimilation. It can be applied to calculate error covariance matrix through Monte-Carlo simulation. This approach is able to resolve the nonlinearity and discontinuity existed within model operator and observation operator. When observation data are assimilated at each time step, error covariances are estimated from the phase-space distribution of an ensemble of model states. The error statistics is then used to calculate Kalman gain matrix and analysis increments. In this study, we develop a one-dimensional soil moisture data assimilation system based on ensemble Kalman filter, the Simple Biosphere Model (SiB2) and microwave radiation transfer model (AIEM, advanced integration equation model). We conduct numerical experiments to assimilate in situ soil surface moisture measurements and low-frequency passive microwave remote sensing data into a land surface model, respectively. The results indicate that data assimilation can significantly improve the soil surface moisture estimation. The improvement in rootzone is related to the model bias errors at surface layer and root zone. The soil moisture does not vary significantly in deep layer. Additionally, the ensemble Kalman filter is predominant in dealing with the nonlinearity of model operator and observation operator. It is practical and effective for assimilating observations in situ and remotely sensed data into land surface models.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Ensemble Kalman filter; Land data assimilation; AIEM; SiB2; Soil moisture

1. Introduction

Soil moisture is an important quantity required to study partitioning of energy, runoff, radiance balance, and moisture movement in a watershed. The accurate estimation of soil moisture provides the fundament to study bio-geophysical processes in land surface. Owing to the current technological and economic limitation, soil moisture can be measured only at limited locations and temporal periods. However, soil moisture in a field generally varies dramatically in space and time. Although the land process model or hydrology model can simulate the continuous variation of soil moisture, the prediction errors caused by the uncertainties of model and parameters would accumulate gradually so that simulation will diverge from the true state. Four-dimensional data assimilation method, which originates from meteorology and oceanography (Daley, 1991), has been applied in land surface science and hydrology (McLaughlin, 1995). The land data assimilation system can assimilate available information such as observation from ground, satellite, and radar to obtain more accurate descriptions of the system state. In land surface and hydrological sciences, some tentative and innovated studies have been conducted. The methods of assimilating microwave data into hydrological models have been explored in many literatures (Entekhabi et al., 1994; Galantowicz et al., 1999; Hoeben & Troch, 2000; Houser et al., 1998; Schuurmans & Troch, 2003). The performance of different data assimilation algorithms coupled with land surface model or hydrological model have been tested and evaluated (Crow & Wood, 2003; Kumar, 2003; Li et al., 2004; McLaughlin, 2002; Pathmathavan et al., 2003a,b; Reichle & Entekhabi, 2001; Reichle et al., 2002a,b; Walker & Willgoose, 2001).

The variational and Kalman filter methods are currently most popular algorithms in data assimilation. However, the two methods can hardly deal with the high nonlinearity and discontinuity of a model operator and an observation operator. The variational approach generally requires an adjoint model, which is
difficult to derive from a land surface model (McLaughlin, 1995). The traditional Kalman filter (Kalman, 1960) is an efficient sequential data assimilation method for linear dynamics and measurement processes with Gaussian error statistics. To assimilate data for nonlinear dynamics and measurement processes, the extended Kalman filter (EKF) was developed (Miller et al., 1994, 1999). However, the EKF is very unstable if the nonlinearities are strong. Furthermore, this method is not computationally feasible for large-scale environmental systems. To overcome these limitations, the ensemble Kalman filter (EnKF) was proposed by Evensen (1994). The EnKF is a sequential data assimilation method, which applies an ensemble of model states to represent the error statistics of the model estimation and to predict the error statistics continuously updated in time. In this method, an analysis scheme is used to operate directly on the ensemble of model states when observations are assimilated. The EnKF has been proven to be an efficient approach to handle strongly nonlinear dynamics and large state spaces. It has been adopted in the atmospheric and oceanic models for numerical prediction (Anderson, 2001; Burgers et al., 1998; Evensen, 1994; Evensen & van Leeuwen, 1996; Houtekamer & Mitchell, 1998, 2001; Keppenne, 2000; Mitchell & Houtekamer, 2000; Whitaker & Hamill, 2002). In the field of hydrology, Reichle et al. (2002a) applied the EnKF to estimate soil moisture profile and found that the method makes a better prediction than the variational assimilation method. Furthermore, after assessing the performance of the EKF and EnKF for soil moisture estimation, Reichle et al. (2002b) concluded that the flexibility and performance of EnKF are also better than those of EKF.

In this study, we develop a one-dimensional soil moisture assimilation system based on ensemble Kalman filter algorithm and SiB2. Advanced integral equation method (AIEM) is adopted as observation operator. Our system was tested with the Tropical Rainfall Measuring Mission Microwave Imager (TMI) brightness temperature and GAME—Tibet soil moisture observations. This paper is organized as follows. In Section 2 the framework of soil moisture assimilation scheme is introduced. The data sets used in this paper are described in Section 3. The assimilation results are analyzed in Section 4 and conclusions and discussions are given in Section 5.

2. Data assimilation methodology

In general, the data assimilation system is composed of a model operator, an observation operator, data assimilation algorithm and data sets. In land data assimilation system (LDAS), the model operator is usually a land surface model to simulate energy and mass transfer between the soil, vegetation and the atmosphere. In this study, SiB2 model is adopted as model operator (Sellers et al., 1996a). The observation operator is used to build the relationship between simulated state variables and observations. In our scheme, the simulated variables are soil moisture in three soil layer but the observed variables are

![Fig. 1. Framework of one-dimensional soil moisture assimilation scheme.](image-url)
brightness temperature from satellite sensor, so the observation operator is a forward microwave radiation transfer model, which is applied to convert the simulated soil moisture at the surface layer into brightness temperature. Data assimilation algorithm is used to integrate simulation and observation, which utilizes observation information to update the state variables produced by model operator. Fig. 1 illustrates the flowchart of assimilating TMI 10.7 GHz brightness temperature with EnKF.

(1) Generate the ensemble of initial soil moisture profiles by adding pseudorandom noise with prescribed statistics to the first guess initial soil moisture in surface layer, root zone, and deep layer.

(2) Drive SiB2 by every ensemble member of initial soil moisture profile, model parameters and atmospheric forcing data, independently. The ensemble of forecast soil moisture profile is computed.

(3) At times when TMI observations are available, the ensemble of observed brightness temperatures are generated by adding observation noise with prescribed statistics to TMI brightness temperature of horizontal and vertical polarization at 10.7 GHz (10.7 GHz H&V). Furthermore, based on the ensemble of forecasted soil moisture profiles, ground temperatures, and vegetation temperatures propagated by model operator, the ensemble of simulated brightness temperatures (10.7 GHz H&V) are calculated by the forward radiation transfer model (AIEM).

(4) The ensemble of updated soil moisture profiles are calculated by EnKF with the ensemble of forecasted soil moisture profiles, the ensemble of simulated brightness temperatures, and the ensemble of observed brightness temperatures. The resulting ensemble of updated soil moisture profiles are then used to reinitialize the model at next times, which is run until TMI observations are available again.

2.1. Model operator

The Simple Biosphere (SiB2) Model, originally developed by Sellers and Mintz (1986), was substantially modified (Sellers et al., 1996a), and has since been referred to as SiB2. The number of biome-specific parameters has been reduced, and most can be derived directly from processed satellite data (Sellers et al., 1996b). SiB2 model includes three soil layers: a surface soil layer of a few centimeters (0–5 cm), which act as a significant source of direct evaporation when moist; a root zone (5–20 cm), which is the supplier of soil moisture to the roots and accounts for transpiration; and a deep soil layer, which acts as a source for hydrological base flow and upward recharge of the root zone. Furthermore, SiB2 defines eleven prognostic physical state variables: three temperatures (canopy temperature, soil surface temperature, and deep soil temperature); two interception water store (canopy water, soil surface); two interception snow/ice stores; three soil moisture wetness values (surface layer, root zone, and deep layer); canopy conductance (Sellers et al., 1996a). In this study, we only consider three prognostic state variables (soil moisture in surface layer, root zone, and deep layer). The three governing equations of water balance are summarized as follows:

Surface layer
\[
\frac{\partial w_1}{\partial t} = \frac{1}{D_1} \left[ (D_e + D_d Q_1 - R_{o1}) - \frac{E_w}{\rho_w} \right]
\]

Root zone
\[
\frac{\partial w_2}{\partial t} = \frac{1}{D_2} \left[ Q_{12} - Q_{23} - \frac{E_{ct}}{\rho_w} \right]
\]

Deep layer
\[
\frac{\partial w_3}{\partial t} = \frac{1}{D_3} [Q_{23} - Q_3].
\]

Where \( w_i \) (\( i = 1, 2, 3 \)) are the liquid water content (\( m^3 m^{-3} \)) of each soil layer; \( D_i \) (\( i = 1, 2, 3 \)) is the thickness of each soil layer (m); \( Q_{ij} \) (\( j = 1, 2 \)) is the water flow between layers \( j \) and \( j+1 \) (m s\(^{-1} \)); \( Q_3 \) is the gravitational drainage from recharge soil moisture store (m s\(^{-1} \)); \( D_e \) is the canopy drainage rate (m s\(^{-1} \)); \( R_{o1} \) is the canopy throughfall rate (m s\(^{-1} \)); \( E_w \) and \( E_{ct} \) are the evaporation rate from soil surface layer (m s\(^{-1} \)) and the canopy transpiration rate through stomata (m s\(^{-1} \)), respectively.

2.2. Ensemble Kalman filter

In this study, the EnKF with perturbed observation is adopted as assimilation algorithm. This algorithm was, proposed by Burgers et al. (1998). It adds random perturbations with the correct statistics to the observations and generates an ensemble of observations which then is used in updating the ensemble of model states. Experiment has shown that this method can result in an updated ensemble with a low variance. The EnKF algorithm includes two steps: forecast and analysis. The overview of procedure of EnKF is as follows.

Considering \( X = [w_1, w_2, w_3]^T \) as a state variable, where \( w_1, w_2 \) and \( w_3 \) are soil moistures in the surface layer, root zone and deep layer, respectively. At time \( t_0 \), a first guess state is used to create an ensemble of size \( N \) by adding pseudorandom noise with prescribed statistics to the first guess initial state.

\[
X_{h,i} = \bar{X}_h + \epsilon_i \quad \epsilon_i \sim N(0, P).
\]

Where, \( X_{h,i} \) is state variable of each member (\( i = 1, 2, ..., N \)) at time \( t_0; \bar{X}_h \) is the expectation of the background field at time \( t_0; P \) is error covariance matrix of background field at time \( t_0; \epsilon_i \) is the random error vector of each member, which is obtained from a multivariate Gaussian distribution with zero mean and error covariance matrix \( P \).

In the forecast step, the forecasted state variable of each member is updated according to

\[
X_{f,i+1} = M(X_{h,i}, a_{i+1}, b) + u_i \quad u_i \sim N(0, Q).
\]
Here, the superscripts ‘a’ and ‘f’ refer to state variables of analysis and forecast, respectively. $X_{i+1}^{a}$ is the forecasted state variable of each member at the time $t+1$; $X_{i+1}^{f}$ is the analyzed state variable of each member at the time $t$; $a_{i+1}$ and $b$ represents atmospheric forcing data and model parameters, respectively; $M(.)$ is model operator; $Q$ is model error covariance matrix $u_{i}$ is model error vector, which conforms to Gaussian distribution with zero mean and covariance matrix $Q$.

In the analysis step, the observation data are perturbed by adding a random observation error and the analyzed state variable of each member is updated as follows:

$$X_{i+1}^{a} = X_{i+1}^{f} + K_{t+1}[(Y_{t+1} + v_{i}) - H(X_{i+1}^{a})] \quad \text{where} \quad v_{i} \sim N(0, R)$$ (6)

Here

$$K_{t+1} = P_{t+1}^{f}H^{T}(HP_{t+1}^{f}HT + R)^{-1}$$ (7)

$$P_{t+1}^{a} = \frac{1}{N-1} \sum_{i=1}^{N} (X_{i+1}^{a} - \overline{X}_{t+1}^{a})(X_{i+1}^{a} - \overline{X}_{t+1}^{a})^{T}$$ (8)

$$P_{t+1}^{f}H^{T} = \frac{1}{N-1} \sum_{i=1}^{N} [X_{i+1}^{f} - \overline{X}_{t+1}^{f}][H(X_{i+1}^{f}) - H(\overline{X}_{t+1}^{f})]^{T}$$ (9)

$$HP_{t+1}^{f}H^{T} = \frac{1}{N-1} \sum_{i=1}^{N} [H(X_{i+1}^{f}) - H(\overline{X}_{t+1}^{f})]^{T}$$ (10)

$$\overline{X}_{t+1}^{a} = \frac{1}{N} \sum_{i=1}^{N} X_{i+1}^{a}$$ (11)

Finally, the analysis state estimate at time $t+1$ is given by the mean of the ensemble members. The analyzed ensemble is then integrated forward until the next observation is available and the process is repeated.

2.3. Observation operator

Without considering multiscattering, the brightness temperature of an emitting rough surface with vegetation can be expressed by

$$T_{B} = e \times T_{s} \times \exp(-\tau_{c}) + T_{c} \times (1 - \omega_{c}) \times [1 - \exp(-\tau_{c})]$$ (12)

Where $T_{B}$ is the brightness temperature; $e$ is the emissivity of the soil surface; $T_{s}$ and $T_{c}$ are the soil surface temperature and canopy temperature, respectively. They are directly obtained from the outputs of SiB2 model. $\tau_{c}$ is the vegetation optical depth; $\omega_{c}$ is the vegetation single scattering albedo. In this model, $\tau_{c}$ is a function of vegetation water content $w_{v}$ (kg m$^{-2}$) (Jackson & Schmugge, 1991)

$$\tau_{c} = b_{v}w_{v}/\cos \theta$$ (13)

Where $\theta$ is the incident angle; the coefficient $b_{v}$ depends on canopy structure and frequency, which can be expressed by

$$b_{v} = b_{1}\lambda^{2}$$ (14)

Where $\lambda$ is the wavelength (cm).

According to Kirchhoff law, the emissivity of a surface represents the fraction of the incident energy absorbed by the surface under thermal equilibrium (Shi et al., 2002; Ulaby et al., 1982b). So, the emissivity of the soil surface can be derived as one minus the reflectivity $\Gamma(\theta, \phi)$

$$e_{p}(\theta, \phi) = 1 - \Gamma(\theta, \phi)$$ (15)

Where $\theta$ and $\phi$ are zenith and azimuth angle of the sensor, respectively. Then, the reflectivity $\Gamma(\theta, \phi)$ can be expressed as

$$\Gamma(\theta, \phi) = \frac{1}{4\pi \cos \theta} \int [\sigma_{p}^{s}(\theta, \phi; \theta_{i}, \phi_{i}) + \sigma_{p}^{s}(\theta, \phi; \theta_{o}, \phi_{o})] d\Omega_{s}$$ (16)

Where $\sigma_{p}^{s}$ and $\sigma_{p}^{c}$ are like-polarized and copolarized bistatic single scattering coefficient, respectively. The subscript $p$ indicates polarization state. $\theta_{i}$ and $\phi_{i}$ are zenith and azimuth of scattering angle, respectively.

The integral equation model (IEM) was proposed by Fung (1994). The IEM was verified by laboratory measurements of bistatic scattering from surfaces with small, intermediate and large-scale roughness. The advanced IEM (AIEM) improves the calculation accuracy of scattering coefficient by keeping the absolute phase term in Greens function which was neglected by IEM (Chen et al., 2003). Therefore, we applied AIEM to calculate the soil single scattering coefficient in our study. In AIEM, the single scattering term is given by

$$\sigma_{p}^{s} = \frac{k^{2}}{2} \exp[-\sigma^{2}(k_{x}^{2} + k_{y}^{2})] \sum_{n=1}^{\infty} \sigma^{2n} \left| f_{p}^{n} \right| \frac{W^{(n)}(k_{x}, k_{z}, k_{y} - k_{s})}{n!}$$ (17)

$$f_{p}^{n} = (k_{z} + k_{x})^{n}f_{p}^{n} \exp(-\sigma^{2}k_{z}k_{x}) + \left( k_{z} + k_{y} \right)^{n}f_{p}^{n} \left( -k_{x}, k_{z}, -k_{y} \right) \left( k_{z} + k_{x} ight)^{n}f_{p}^{n} \left( -k_{x}, -k_{y} \right)$$ (18)

$$k_{s} = k \sin \theta \cos \phi$$ (19)

$$k_{v} = k \sin \theta \sin \phi$$ (20)

$$k_{z} = k \cos \theta$$ (21)

$$k_{x} = k \sin \theta \cos \phi$$ (22)

$$k_{y} = k \sin \theta \sin \phi$$ (23)

$$k_{s} = k \cos \theta$$ (24)

where $k$ is the wave number, $\sigma$ is the root-mean-square height of surface. $f_{p}^{n}$ is a function of $\theta, \phi, \sigma$ and $\epsilon_{r}$ (soil dielectric constant).
$f_{pq}$ and $F_{pq}$ are the Kirchhoff coefficient and the complementary field coefficient, respectively. $W^{(n)}$ is the Fourier transform of the $n$th power of the surface correlation function.

The dielectric constant of the soil ($\varepsilon_r$) depends mainly on our assimilation variable, soil moisture, $m_v$ (Dobson et al., 1985)

$$\varepsilon_r^2 = 1 + \frac{\rho_b}{\rho_s} (\varepsilon_s^2 - 1) + m_v^\beta \varepsilon_{fw}^2 - m_v$$

where, $\rho_b$ is the soil bulk density (g cm$^{-3}$); $\rho_s$ is the soil specific density (g cm$^{-3}$); $\varepsilon_s$ is the dielectric constant of soil having extremely low moisture contents, and $\varepsilon_{fw} \approx (4.7, 0)$; $\varepsilon_{fw}$ is the dielectric constant of free water (Ulaby et al., 1982a). In our study, $\alpha=0.65$ and $\beta$ are determined by soil texture and expressed by the formula (Ulaby et al., 1982a)

$$\beta = 1.09 - 0.11 \times S/100 + 0.18 \times C/100$$

Where $S$ and $C$ are percentage of sand and clay, respectively.

### 3. Experiments sites and data

#### 3.1. The description of experiment sites

The Tibetan Plateau, which is considered as a heat source of atmosphere in summer and to have an important impact on the Asian monsoon system (Ye & Gao, 1979), is one of studying regions of the GEWEX (Global Energy and Water Cycle Experiment), Asian Monsoon Experiment (GAME). To understand the interactions between the land surface and the atmosphere over the plateau in the context of the Asian monsoon system, a plateau-scale experiment and a meso-scale experiment were carried out by the prophase observation period (POP) field work in August–September 1997 and the intensive observation period (IOP) field work in May–September 1998 as well. A much dense automatic observation network of meteorology and land surface hydrology has been established for long-period observation (Fig. 2). Numerous high quality data are being
obtained and a database of GAME–Tibet IOP and POP has been developed. The web page [http://monsoon.t.u-tokyo.ac.jp/tibet/](http://monsoon.t.u-tokyo.ac.jp/tibet/) gives more details about the instrumentation and data collection methodology.

### 3.2. Observations in situ

In order to perform our experiments, the observation site for candidate should have automated weather station (AWS), soil moisture and temperature system (SMTMS) and precipitation gauge. Unreliable data due to instrument malfunction have been removed, so only observation data at the sites of MS3478, MS3608, and MS 3637 are suitable for our experiments from July 9 to August 7, in 1998. Among the six forcing variables required to run SiB2, the shortwave downward radiation, air temperature and wind speed are from AWS (Automatic weather station) measurement directly. The longwave downward radiation is calculated by Brunt’s equation, which is a standard method in SiB2. The vapor pressure is transformed from humidity. The precipitation in summer is measured during the IOP using a rain gauge with digital recorder. At these three sites, SMTMS had been working continuously to collect soil data at the temporal resolution of 1 h. The time domain reflectometry (TDR) was employed to measure the volumetric water content at depths of 4, 20, 60, 100, 160, 196 cm. In this study, the soil volumetric water content at depths of 4, 20 and 100 cm is applied in our assimilation experiments.

### 3.3. Model parameters

In SiB2 model, many parameters are defined, including land cover, soil types, spatial–temporal varied vegetation parameters

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
<th>Value</th>
<th>Data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>Frequency (GHz)</td>
<td>10.7</td>
<td>GAME–Tibet/TRMM</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Incident angle (°)</td>
<td>52.8</td>
<td>GAME–Tibet/TRMM</td>
</tr>
<tr>
<td>$x$</td>
<td>Vegetation canopy parameter</td>
<td>-1.08</td>
<td>Pathmathevan et al. (2003a)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>Vegetation canopy parameter</td>
<td>0.62</td>
<td>Pathmathevan et al. (2003a)</td>
</tr>
<tr>
<td>$\omega_c$</td>
<td>The vegetation single scattering albedo</td>
<td>0.05</td>
<td>Pathmathevan et al. (2003a)</td>
</tr>
<tr>
<td>$W_c$</td>
<td>Vegetation water content (kgm $^{-2}$)</td>
<td>0.0–0.25</td>
<td>GAME–Tibet and optimization</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Root-mean-square rms height of surface roughness (cm)</td>
<td>2.0–4.0</td>
<td>GAME–Tibet and optimization</td>
</tr>
<tr>
<td>$\rho_b$</td>
<td>Soil bulk density (kg m $^{-3}$)</td>
<td>1.4</td>
<td>Pathmathevan et al. (2003a)</td>
</tr>
<tr>
<td>$\rho_s$</td>
<td>Soil specific density (kg m $^{-3}$);</td>
<td>2.65</td>
<td>Pathmathevan et al. (2003a)</td>
</tr>
<tr>
<td>$S$</td>
<td>Sand percentage</td>
<td>60</td>
<td>GAME–Tibet</td>
</tr>
<tr>
<td>$C$</td>
<td>Sand percentage</td>
<td>20</td>
<td>GAME–Tibet</td>
</tr>
</tbody>
</table>

**Fig. 3.** The location and overpass time of TMI measurement around the site of MS3478, MS3608, and MS3637 from July 19 (Julian Day 190) to August 7 (Julian Day 219), 1998.
and aerodynamic parameters. Most of the static parameters are directly derived from Sellers et al. (1996b), while other parameters are obtained by field measurements in GAME-Tibet experiment and optimization. A thorough description of the parameters has been presented in other papers (Li & Koike, 2003; Pathmathevan et al., 2003a). As for the observation operator, we require 3 variables and 12 parameters to calculate soil surface brightness temperature. The input variables are directly from SiB2 model, including soil surface temperature, soil water content, and canopy temperature. Some parameters are directly extracted from reference (Pathmathevan et al., 2003a) and the other parameters are obtained by field measurements in GAME-Tibet experiment and optimization. The input parameters required by observation operator are listed in Table 1.

### 3.4. TMI data

The Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) is a passive microwave sensor designed to provide quantitative rainfall information over a wide swath under the TRMM satellite. The TMI measures the brightness

<table>
<thead>
<tr>
<th>Site</th>
<th>Simulation</th>
<th>Observation</th>
<th>Difference between simulation and observation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
<td>Mean</td>
</tr>
<tr>
<td>MS3478</td>
<td>Surface</td>
<td>0.3691</td>
<td>1.79E-04</td>
</tr>
<tr>
<td></td>
<td>Root zone</td>
<td>0.3732</td>
<td>7.90E-05</td>
</tr>
<tr>
<td></td>
<td>Deep layer</td>
<td>0.3085</td>
<td>2.48E-06</td>
</tr>
<tr>
<td>MS3608</td>
<td>Surface</td>
<td>0.2132</td>
<td>1.93E-04</td>
</tr>
<tr>
<td></td>
<td>Root zone</td>
<td>0.2202</td>
<td>9.70E-05</td>
</tr>
<tr>
<td></td>
<td>Deep layer</td>
<td>0.2879</td>
<td>2.82E-05</td>
</tr>
<tr>
<td>MS3637</td>
<td>Surface</td>
<td>0.2132</td>
<td>4.98E-04</td>
</tr>
<tr>
<td></td>
<td>Root zone</td>
<td>0.2318</td>
<td>3.23E-04</td>
</tr>
<tr>
<td></td>
<td>Deep layer</td>
<td>0.3070</td>
<td>2.35E-06</td>
</tr>
</tbody>
</table>

![Fig. 4. Results of assimilating in situ soil moisture observations (4 cm) at MS3478 site, Tibetan Plateau, from July 19 (Julian Day 200) to August 7 (Julian Day 219), 1998.](image-url)
temperature at five separate frequencies: 10.7, 19.4, 37, 85.5 GHz (dual polarization), and 21.3 (vertical polarization only). The low frequency (10.7 GHz) is known to be sensitive to soil moisture and the effects of atmosphere can be negligible in the frequency, so it is selected for our study.

We obtained TMI Calibrated Brightness temperature (1B11) Product (http://daac.gsfc.nasa.gov) during the period of July 9 to August 7, 1998. Considering the uniformity of the land surface condition in the Tibetan Plateau, the effect of the difference in footprint might be negligible. In this study, all the measurements within a circle with radius 0.1° (about 10 km) are sampled to increase the sampling number, so there are twice or three times TMI measurements per day. Furthermore, the overpass time of TMI varies, so it is very useful to capture the diurnal variation of surface soil moisture. Fig. 3 shows the location and overpass time of TMI measurement around the site of MS3478, MS3608, and MS3637 from July 9 to August 7, 1998.

4. Result

4.1. The determination of error covariance

In data assimilation scheme, the prior knowledge of background, model, and observation errors is crucial for correctly assimilating observations into a model. However, observation errors mainly include instrument error, and representative error. Model errors are usually caused by the uncertainty of model structure, soil parameters (such as hydraulic conductivity and porosity) and atmospheric forcing data (such as precipitation, air humidity, and radiation). So it is very difficult to quantify the model and observation errors.

In our study, the assimilated observations are in situ surface soil moisture observations and TMI horizontal and vertical polarization brightness temperature at 10.7 GHz, respectively. The Model operator and Observation Operator have been calibrated using in situ observations at each site before conducting our assimilation experiment, so we assumed that the in situ observations are perfect. Additionally, we added stochastic noise with zero mean and variance $2.5E-5$ to in situ observations to generate a set of new observations which were used to perform the experiment of assimilating in situ surface observation. As for the experiment of assimilating the brightness temperature of TMI, observation errors may be caused by several factors such as the scale difference between observation and simulation, sensor errors, atmospheric condition, and observation operator uncertainty. Based on calibration results, the standard deviation of TMI horizontal and vertical polarization brightness temperature at 10.7 GHz were set to 3 K and 2 K, respectively.

We only collected 30-day atmospheric forcing data and precipitation data available at each site from July 9 (Julian Day

![Fig. 5. Results of assimilating in situ soil moisture observations (4 cm) at MS3608 site, Tibetan Plateau, from July 19 (Julian Day 200) to August 7 (Julian Day 219), 1998.](image)
to August 7 (Julian Day 229), 1998, so the first 10-day (July 9 to 18) data were used to calculate model and background errors and the last 20-day (July 19 to August 7) data were used to conduct assimilation experiments. We ran SiB2 model by the first 10-day data and obtained the corresponding soil moisture simulation results at surface layer, root zone, and deep layer, respectively. Furthermore, the errors statistic of simulation, observation, and the differences between simulation and observation at these sites were calculated and listed in Table 2. Ignoring soil moisture correlation among different soil layers, the mean and variance of soil moisture observation at each site was adopted as the initial soil moisture and background errors covariance for our assimilation experiments, respectively. Because the EnKF algorithm cannot directly deal with model bias error, so we neglected the influence of model bias in our system and used the variance of the difference between the simulated and observed soil moisture at each soil layer to construct model errors covariance at each site.

### 4.2. Assimilation of in situ surface soil moisture observation

In this experiment, we assimilate in situ surface soil moisture observation twice per day (0UTC and 12UTC) into our data assimilation system to update soil moisture profile beginning on

<table>
<thead>
<tr>
<th>Site</th>
<th>Simulation</th>
<th>Assimilation of in situ surface soil moisture</th>
<th>Assimilation TMI brightness temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>Average error</td>
<td>RMSE</td>
</tr>
<tr>
<td>MS3478</td>
<td>Surface</td>
<td>0.0226</td>
<td>0.0153</td>
</tr>
<tr>
<td></td>
<td>Root zone</td>
<td>0.0316</td>
<td>0.02957</td>
</tr>
<tr>
<td></td>
<td>Deep layer</td>
<td>0.0146</td>
<td>−0.0141</td>
</tr>
<tr>
<td>MS3608</td>
<td>Surface</td>
<td>0.0311</td>
<td>0.0286</td>
</tr>
<tr>
<td></td>
<td>Root zone</td>
<td>0.0763</td>
<td>0.0751</td>
</tr>
<tr>
<td></td>
<td>Deep layer</td>
<td>0.0151</td>
<td>−0.0139</td>
</tr>
<tr>
<td>MS3637</td>
<td>Surface</td>
<td>0.0418</td>
<td>0.0402</td>
</tr>
<tr>
<td></td>
<td>Root zone</td>
<td>0.0319</td>
<td>−0.0164</td>
</tr>
<tr>
<td></td>
<td>Deep layer</td>
<td>0.0078</td>
<td>0.0065</td>
</tr>
</tbody>
</table>

Fig. 6. Results of assimilating in situ soil moisture observations (4 cm) at MS3637 site, Tibetan Plateau, from July 19 (Julian Day 200) to August 7 (Julian Day 219), 1998.
July 19 (Julian Day 200), 1998. The ensemble size is set to 100 according to prior study (not shown in this paper). The time series of assimilation and simulation results of soil moisture in surface layer, root zone and deep layer from July 19 (Julian Day 200) to August 7 (Julian Day 219), 1998 are given in Figs. 4–6. The RMSE (root mean square error) and average errors of the simulation and assimilation results at surface layer, root zone, and deep layer for these sites are summarized in Table 3.

At the surface layer, the estimation of soil moisture has significant improvement in these sites. In our experiment, the observation errors are smaller than model errors, so the soil moisture is pulled closely to in situ observation at the depth of 4 cm when observation data available. We can see that the simulation results of surface layer at these sites are overestimated, but the assimilation results follow in situ observations more closely than the simulation results. Furthermore, the results of assimilation can capture exactly the diurnal variation of surface soil moisture at MS3608 and MS3637 sites. At root zone, compared to in situ observation at the depth of 20 cm, assimilation results are better than simulation results at MS3478 and MS3608 site, but largely underestimate relative to the results of the simulation at MS3637 site. As shown in Figs. 4–6, the evolution trend of simulated soil moisture is the same at surface layer and root zone, so it is positive correlation between the simulation results of soil moisture at surface layer and root zone. If the simulation results of surface soil moisture overestimate, the assimilation results of surface soil moisture will be close to in situ observation by decreasing surface soil moisture when observation is available. Furthermore, this information is propagated to the deeper soil layer such as root zone and makes the assimilation results lower than the simulation results at root zone. At the sites of MS3478 and MS3608, the simulation results of soil moisture at root zone are overestimated, so the assimilation results are better than the simulation results. However, at the site MS3637, the simulation results of soil moisture at root zone are underestimated, so the assimilation results are worse than simulation results. At deep layer, on the one hand soil moisture does not vary significantly; on the other hand, owing to the given small model errors in deep layer, the propagated information from surface layer is weak. So the results of simulation and assimilation are in good agreement with in situ observation at the depth of 100 cm.

4.3. Assimilation of TMI brightness temperature

In this experiment, we try to assimilate TMI low-frequency brightness temperature into our data assimilation system (Fig. 1) to improve soil moisture profile. The Figs. 7–9 show the results of assimilating brightness temperature of horizontal and vertical polarization at 10.7 GHz from July 19 (Julian Day 200) to August 7 (Julian Day 219), 1998. The Table 3 also lists the

![Fig. 7. Results of assimilating TMI brightness temperature (10.7 GHz, H&V) at MS3478 site, Tibetan Plateau from July 19 (Julian Day 200) to August 7 (Julian Day 219), 1998.](image-url)
errors statistic results of assimilating TMI brightness temperature.

The results show that this scheme can accurately retrieve surface soil moisture by assimilating remote-sensed brightness temperature. At surface layer, compared to simulation results, the assimilation results have significant improvement. Additionally, we find that when the soil moisture is less that 20%, the assimilation results are a little far from the observations at MS3608 site. This may be attributed to TMI shallower sensing depth compared to in situ soil moisture observation (4 cm). So the assimilation results have fast dry-down during dry season (from Julian Day 207 to 211). At root zone, the assimilation results of soil moisture are complicated. The assimilation results have significant improvement comparing to the simulation results at the sites of MS3478 and MS3608. However, the assimilation results are worse than the simulation results at the sites of MS3637. At deep layer, the assimilation results slightly differ from the simulation results because soil moisture in deep layer is stable and hardly influenced by moisture in surface layer.

5. Summary and conclusions

In this study we develop a one-dimensional soil moisture assimilation system based on ensemble Kalman filter algorithm and SiB2. A physically-based model, AIEM, is adopted as an observation operator, which can accurately calculate the emissivity of land surface. This system can assimilate in situ soil moisture observations and remote sensing brightness temperature. According to TMI brightness temperature observation and soil moisture in situ observation from July 9 to August 7, 1998, we have done two experiments to evaluate our data assimilation system: (a) to assimilate in situ surface soil moisture observations (4 cm) every 12 hour; (b) to assimilate TMI brightness temperature of horizontal and vertical polarization at 10.7 GHz. From the results of these experiments, the following conclusions can be made:

(1) The ensemble Kalman Filter is practical and effective for assimilating in situ observation and passive microwave remote sensing observation into land surface models. It can effectively solve the nonlinearity that existed in the model operator and observation operator.

(2) Assimilation of both in situ surface soil moisture observation and low-frequency brightness temperature will significantly improve the estimation of soil moisture in surface layer. At root zone, the assimilation results of soil moisture are complicated. If soil moisture in surface layer and root zone are both overestimated or underestimated in comparison with observation, the assimilation results of soil moisture at root zone will be improved, such as the results at MS3608 and MS3678 sites. However, if soil

Fig. 8. Results of assimilating TMI brightness temperature (10.7 GHz, H&V) at MS3608 site, Tibetan Plateau from July 19 (Julian Day 200) to August 7 (Julian Day 219), 1998.
moisture in surface is overestimated and soil moisture at root zone is underestimated (or just the opposite), the assimilation results will be worse than the simulation results, such as the results at MS3478 site. For soil moisture in deep layer, the assimilation results slightly differ from the simulation results because soil moisture in deep layer is stable and hardly influenced by moisture in surface layer.

(3) Though assimilation results are close to observation when there are bias errors in model operator, the larger bias would make system unstable, break up water balance, and propagate improper information to deeper soil layer as shown in Fig. 6. So it is very important to couple bias correction method with EnKF algorithm Moradkhani et al., 2005, which will be helpful to improve the estimation of soil moisture profile.

(4) Soil physical parameters such as the saturated hydraulic conductivity and the porosity in the three soil layers are considered to be homogeneous in SiB2, this assumption will bring significant error to the simulation of soil moisture. Since the soil physical parameters have different values in the layers, the current model needs to be adjusted to cope with heterogeneous soil characteristics.

(5) Both the quality of TMI observation and the accuracy of radiation transfer model influence the assimilation results. It will cause the improvement slightly, even worse. So we need further study on how to identify the quality of observation data and decrease the uncertainty of radiation transfer model.

(6) Low-frequency brightness temperature observations from passive microwave sensors onboard satellite have higher temporal resolution and are suitable for near-surface layer soil moisture retrieval, so remote sensing brightness temperature observations will become the major source to be assimilated into land data assimilation system. However, if there are not enough observations assimilated, results of assimilation would be worse than those of simulation. So we can make full use of multisensors such as SSM/I, AMSR-E and TMI to increase observations. Assimilating remote sensing observations from multifrequency, multipolarization and multisensor will be essential steps for developing operational land data assimilation system.

Acknowledgements

This work is supported by the China State Key basic Research Project (grant number: 2001CB309404), the NSFC (National Science Foundation of China) project (grant number: 90202014 and 40401044), CAS (Chinese Academy of Science) International Partnership Project “The Basic Research for Water Issues of Inland River Basin in Arid Region” (CXTD-Z2005-2),
References


