Advances in Water Resources 000 (2015) 1–14

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Comparison of ensemble-based state and parameter estimation methods for soil moisture data assimilation

Weijing Chen [a,b], Chunlin Huang [b,c,*], Huanfeng Shen [a], Xin Li [b,d]

[a] School of Resource and Environmental Science, Wuhan University, Wuhan, Hubei 430079, China
[b] Key Laboratory of Remote Sensing of Gansu Province, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, Gansu 730000, China
[c] Heihe Remote Sensing Experimental Research Station, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, Gansu 730000, China
[d] CAS Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Data assimilation
Soil moisture
Brightness temperature
State-parameter estimation
Common Land Model

ABSTRACT

Model parameters are a source of uncertainty that can easily cause systematic deviation and significantly affect the accuracy of soil moisture generation in assimilation systems. This study addresses the issue of retrieving model parameters related to soil moisture via the simultaneous estimation of states and parameters based on the Common Land Model (CoLM). The state-parameter estimation algorithms AEnKF (Augmented Ensemble Kalman Filter), DEnKF (Dual Ensemble Kalman Filter) and SODA (Simultaneous optimization and data assimilation) are entirely implemented within an EnKF framework to investigate how the three algorithms can correct model parameters and improve the accuracy of soil moisture estimation. The analysis is illustrated by assimilating the surface soil moisture levels from varying observation intervals using data from Mongolian plateau sites. Furthermore, a radiation transfer model is introduced as an observation operator to analyze the influence of brightness temperature assimilation on states and parameters that are estimated at different microwave signal frequencies. Three cases were analyzed for both soil moisture and brightness temperature assimilation, focusing on the progressive incorporation of parameter uncertainty, forcing data uncertainty and model uncertainty. It has been demonstrated that EnKF is outperformed by all other methods, as it consistently maintains a bias. State-parameter estimation algorithms can provide a more accurate estimation of soil moisture than EnKF. AEnKF is the most robust method, with the lowest RMSE values for retrieving states and parameters dealing only with parameter uncertainty, but it possesses disadvantages related to increasing sources of uncertainty and decreasing numbers of observations. SODA performs well under the complex situations in which DEnKF shows slight disadvantages in terms of statistical indicators; however, the former consumes far more memory and time than the latter.

1. Introduction

Soil moisture is a key variable in understanding land surface hydrological processes that partition precipitation into runoff and infiltration and that control water storage and drainage [20]. As a vital element in the water and energy cycle, soil moisture forms the foundation of meteorological research, water resource regulation and agricultural management [9,13,37,54]. Modeling provides temporally and spatially continuous simulations and predictions of soil moisture but lacks precision. Meanwhile, many types of observations have uncertain accuracy and poor resolution, which is due to limited finances and the capabilities of the utilized instruments themselves. Many studies have noted that data assimilation has the potential to produce continuous and accurate soil moisture data sets that are reconciled in temporal and spatial resolution [6,18,22,27,28,41,52].

The assimilation of data originating from atmospheric and oceanographic sciences [15,21] takes full advantage of imperfect models and finite data in an optimal way by merging the information embodied in remote-sensing or ground-based networks into a dynamic model to improve forecast trajectory. Many experiments have been conducted to improve soil moisture estimation using *in situ* observations at the beginning of the development of land or hydrology data assimilation [5,19,26,55]. However, remote-sensing techniques dominate over *in situ* measurements in terms of the scope of an observed area. Low-frequency microwave brightness temperature is highly related to near-surface soil moisture and is only weakly affected by the atmosphere and clouds. Recently, many soil

* Corresponding author at: Key Laboratory of Remote Sensing of Gansu Province, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, Gansu 730000, China. Tel.: +869314967975.
E-mail address: huangcl@lzb.ac.cn (C. Huang).

http://dx.doi.org/10.1016/j.advwatres.2015.08.003
0309-1708/

moisture products have been applied to enhance model simulations on a regional scale because of the successful launch of a series of satellites with microwave sensors, such as AMSR-E (Advanced Microwave Scanning Radiometer for Earth Observing System), ASCAT (Advanced Scatterometer), and SMOS (Soil Moisture and Ocean Salinity) [3,4,12,39]. However, the large uncertainties that exist in retrieval algorithms may contaminate the quality of soil moisture products, which are expected to be used during assimilation. An alternative method is to directly assimilate brightness temperature into land surface or hydrological models to yield optimal soil moisture estimation [8,22,24,29,40].

As a fusion method, data assimilation improves model simulation by effectively handling background error, which is not the sole factor that influences the capability of data assimilation. Most applications of state assimilation focus on managing the errors that are embodied in the background, on observations and on forcing data, as opposed to employing model structure or parameters [1,25,48,53]. However, the outcome of data assimilation relies on an unbiased prediction of a model state, which is largely dependent on the effectiveness and representativeness of the model. Discrepancies between model parameterization and actual land surface processes account for model errors, but many studies are conducted under the assumption of a state-of-the-art model. Considering that parameter uncertainty affects state estimation to a large extent, the proper specification of model parameters as functions of variables when characterizing a state has become a crucial aspect of recent studies.

It is generally recognized that parameter calibration can diminish long-term bias, while state updating can weaken stochastic error. Thus, a calibration period is usually necessary to optimize the parameters of a hydrological model. Recently, many scientists have focused their attention on minimizing parameter errors in land data assimilation by performing simultaneous estimations of states and parameters [31,32,34,38,46,47,56]. This joint method expands the data assimilation framework from only updating model states to updating both model states and parameters.

Within the framework of EnKF-based assimilation, three types of algorithms are typically used for simultaneous state and parameter estimation. EnKF, which was originally proposed by Evensen [14], is a commonly used sequential algorithm for data assimilation and has shown strength in dealing with non-linear models because of its reliance on the propagation of a random ensemble of retrieved variables. EnKF is also an advantageous approach for highly dimensional applications, mainly because it captures the relevant parts of an error structure by means of a comparatively small ensemble of model trajectories, including (1) The state augmentation approach [2,17,33]. Monsivais-Huertero et al. [33] employed both synthetic and field observations to understand the effects of simultaneous state-parameter estimation using an augmented state vector, spatial and temporal update frequency and forcing data uncertainty in root-zone soil moisture. (2) The dual filter approach ([30,35,44]; Lü et al., 2008). Moradkhani et al. [35] presented a dual state-parameter estimation approach for the sequential estimation of parameters and states in a conceptual rainfall-runoff model (HyMOD) using observed streamflow. The algorithm is recursive, updating parameters and states in turn, and is mutually affective. (3) The parameter optimization and state assimilation approach [45,50]. Vrugt et al. [50] proposed the combined usage of parameter optimization and sequential data assimilation to facilitate valid treatment of input, output, parameter, and model structural errors in a Sacramento model, which was designated as the simultaneous optimization and data assimilation method (SODA).

Given the abundance of studies on joint state and parameter estimation and the paucity of investigations about algorithm applicability, the main objective of this study was to evaluate the performance abilities of all three of the above-discussed algorithms in a series of comparative experiments. We developed a data assimilation framework based on the common land model (CoLM), with soil moisture as the state variable of concern. First, we utilized *in situ* soil moisture to diagnose the performance of state assimilation at different observation intervals; at the same time, we examined the applicability of retrieving information regarding three soil property parameters (volume percentage of sand, volume percentage of clay and porosity). Second, we coupled the land surface model to a radiative transfer model (RTM), which acted as an observation operator, and added the standard deviation of the surface height into the parameter space. Brightness temperature was assimilated at different frequency combinations to judge the validity of each method. All of the soil moisture and brightness temperature experiments were implemented for three different cases: parameter uncertainty, atmospheric forcing data uncertainty and model uncertainty.

This paper is structurally organized as follows: models and methods are introduced in Section 2, in which the study area and experimental design are also described. The results of and discussion about the experiments are explained in Section 3. Section 4 presents other related discussions and the final conclusions.

2. Data assimilation scheme

2.1. Land surface model

The CoLM [10] is an improved version of the Community Land Model, with one vegetation layer, 10 unevenly spaced vertical soil layers, and up to 5 snow layers (depending on the total snow depth). We employed the CoLM as a dynamic model (model operator) to maintain its prognostic variables, which represent soil moisture in this work. The soil water equation is:

$$\frac{\Delta z_j}{\Delta t}\Delta\theta_j = [q_{j-1} - q_j] - f_{root,j*}E_{tr} \tag{1}$$

where $\Delta\theta_j$ is the change in water content as a result of the last time step in layer j, and Δz_j is the thickness of layer j. $f_{root,j*}$ and E_{tr} represent effective root fraction and transpiration, respectively. q_j is the water flow at the depth of the $z_{h,j}$ interface between layer j and layer j+1, as calculated by Darcy's law:

$$q = -K\left(\frac{\partial\psi}{\partial z} - 1\right) \tag{2}$$

K and ψ are the hydraulic conductivity and matric potential of soil, respectively, which vary with soil water content, θ, and soil texture based on the scheme proposed by Clapp and Hornberger [7].

(a) The hydraulic conductivity of soil, K, is:

$$K = K_{sat}s^{2B+3} \tag{3}$$

where the wetness (liquid water degree of saturation) is defined as:

$$s = \left[\frac{\theta_1}{1-\theta_d-\theta_i}\right] \tag{4}$$

where$1-\theta_d$represents porosity, and the exponent B is defined as $B = 2.91 + 0.159(\%clay)$. For numerical reasons, when the effective porosity, $(1-\theta_d-\theta_i)$, is less than 0.05 in any of two neighboring layers, or when the liquid content is less than 0.001, then $K = 0$.

(b) The matric potential of soil is ψ, and the matric potential of unfrozen soil is:

$$\psi = \psi_{sat}s^{-B} \tag{5}$$

CoLM establishes the relationship between soil texture and soil thermal and hydraulic parameters as follows.

The specific heat capacity of a soil solid is:

$$\rho_d c_c = (2.128\%\text{sand} + 2.385\%\text{clay}) \times 10^6/(\%\text{sand} + \%\text{clay}), (\text{Jm}^{-3}\,\text{K}^{-1}) \quad (6)$$

The thermal conductivity of a soil solid is:

$$\lambda_d = (8.80\%\text{sand} + 2.92\%\text{clay})/(\%\text{sand} + \%\text{clay}), (\text{Wm}^{-1}\,\text{K}^{-1}) \quad (7)$$

The saturated matrix potential is:

$$\psi_{sat} = -10 \times 10^{1.88-0.013(\%\text{sand})}, (\text{mm}) \quad (8)$$

The saturated hydraulic conductivity is:

$$K_{sat} = 0.0070556 \times 10^{-0.884+0.0153(\%\text{sand})}, (\text{mm}\,\text{s}^{-1}) \quad (9)$$

2.2. Microwave radiative transfer model

To consider the influence of a canopy on the emitted brightness temperature from soil, the Q-h model, which is an empirical model, was revised and selected as an observation operator to link the soil moisture output from the CoLM with the brightness temperature:

$$T_{B,H(V)} = T_g(1 - \Gamma_{H(V)})\exp(-\tau_c) + T_c(1-\omega)[1 - \exp(-\tau_c)][1 + \Gamma_{H(V)} \times \exp(-\tau_c)] \quad (10)$$

in which T_B is brightness temperature, T_g is ground temperature, and the subscript $H(V)$ denotes vertical (horizontal) polarization. τ_c represents vegetation optical thickness relative to vegetation water content, as proposed by Jackson and Schmugge [23]:

$$\tau_c = b(100\lambda)^{\chi} w_c/\cos\gamma \quad (11)$$

in which ω is the single-scattering albedo of the vegetation calculated by the following empirical formula obtained by Fujii [16]

$$\omega = 0.00083/\lambda \quad (12)$$

in which γ is the incident angle, and $\lambda[m]$ is the wavelength. b and χ are empirical coefficients, which are set to 3.98 and −1.41, respectively, and are identical to the results of the sensitivity case published by Yang et al. [56]. w_c is vegetation water content estimated according to Paloscia and Pampaloni [36].

$$w_c = \exp(LAI/3.3) - 1 \quad (13)$$

where LAI (in meters squared per meters squared) is the leaf area index.

Γ is soil reflectivity written as:

$$\Gamma_{H(V)} = [(1-Q)\cdot R_{H(V)} + Q \cdot R_{V(H)}]\exp(-h) \quad (14)$$

where Q and h are empirically determined surface roughness parameters, defined as $h = (2ks\cos\gamma)^2$, $Q = 0.35[1 - \exp(-0.6s^2\lambda)]$. k is the wave number defined as $2\pi/\lambda$, and s is the standard deviation of the surface height. R is the rough reflectivity calculated by:

$$R_H = \left|\frac{\cos\gamma - \sqrt{\varepsilon_r - \sin^2\gamma}}{\cos\gamma + \sqrt{\varepsilon_r - \sin^2\gamma}}\right|^2 \quad (15)$$

$$R_V = \left|\frac{\varepsilon_r\cos\gamma - \sqrt{\varepsilon_r - \sin^2\gamma}}{\varepsilon_r\cos\gamma + \sqrt{\varepsilon_r - \sin^2\gamma}}\right|^2 \quad (16)$$

ε_r is an angle soil dielectric constant that depends on the soil moisture θ and is calculated according to Dobson et al. [11]:

$$\varepsilon_r = [1 + (1-\rho)(\varepsilon_s^{\alpha} - 1) + \theta^{\beta}\varepsilon_{fw}^{\alpha} - \theta]^{1/\alpha} \quad (17)$$

$\varepsilon_s = 4.7 + 0.0j$ denotes the dielectric constant for a mineral soil in which j represents the imaginary part of plural, ρ is the soil porosity, ε_{fw} is the dielectric constant of free water, $\alpha = 0.65$, and β is a soil texture-dependent coefficient [49]. S and C indicate the percentages of sand and clay, respectively.

$$\beta = (127.48 - 0.519\times S - 0.152\times C)/100 + (1.33797 - 0.603\times S - 0.166\times C)/100\,j \quad (18)$$

2.3. Ensemble-based state-parameter estimation methods

2.3.1. Ensemble kalman filter

In this study, EnKF is the foundation of all three of the included simultaneous state and parameter estimation algorithms and is divided into two steps: forecast and analysis. An overview of the EnKF procedure follows.

The initial state can be defined as X_0^a. Then, the ith member, $X_{i,0}^a$, of the initial state ensemble is obtained by adding the random noise to X_0^a.

$$X_{i,0}^a = X_0^a + u_i \quad u_i \sim N(0, P_0) \quad (19)$$

where u_i is the background error vector, which conforms to a Gaussian distribution with a zero mean and the covariance matrix P_0.

Forecast: In the forecast step, each member of the state realizations is propagated according to:

$$X_{i,k+1}^f = M(X_{i,k}^a, \alpha_{k+1}, \beta_{k+1}) + w_i \quad w_i \sim N(0, Q) \quad (20)$$

Here, $M(\bullet)$ is a model operator and represents the CoLM in this case. The superscripts a and f refer to the states of analysis and forecast, respectively. α and β are atmospheric forcing data and model parameters, which are used to run the CoLM. The model error is indicated by w_i, with a zero mean and the covariance matrix Q, and represents all of the uncertainties related to the forcing data and the model structure.

Analysis: When an observation is available, an observation vector is assimilated into the model.

$$\hat{Y}_{i,k+1} = H(X_{i,k+1}^f) \quad (21)$$

Here, $H(\bullet)$ is the observation operator, which establishes a relationship between model states and observations and is linear, constituting 0 and 1 in soil moisture assimilation experiments, but is substituted by a non-linear Radiative Transfer Model when directly assimilating brightness temperature observations. $\hat{Y}_{i,k+1}$ is the so-called projection of the model state within the observational space.

A linear correction equation is used according to a standard Kalman filter to update the forecasted state ensemble members:

$$X_{i,k+1}^a = X_{i,k+1}^f + K_{k+1}(Y_{i,K+1} - \hat{Y}_{i,k+1}) \quad (22)$$

$Y_{i,k+1}$ is generated by adding the stochastic perturbation into the actual observation Y_{k+1} at the time k+1, with a zero mean and the covariance matrix R, according to:

$$Y_{i,k+1} = Y_{k+1} + v_i \quad v_i \sim N(0, R) \quad (23)$$

K_{k+1} is the Kalman gain matrix at the time k+1, which is calculated as:

$$K_{k+1} = P_{k+1}^f H^T (H P_{k+1}^f H^T + R)^{-1} \quad (24)$$

P_{k+1}^f is the forecast error covariance matrix at the time k+1. $P_{k+1}^f H^T$ is the cross covariance between the model state forecasts X_{k+1}^f and their projections $H(X_{k+1}^f)$ within the observation space, while $HP_{k+1}^f H^T$ is the covariance of $H(X_{k+1}^f)$.

Finally, the analysis state variable estimated at the time k+1 is given by the averaged value of the ensemble members. The analyzed ensemble is then integrated forward until the next observation appears and the process is repeated.

2.3.2. Augmented ensemble Kalman filter

AEnKF treats parameter as a type of special state variable by augmenting the state vector to transform the parameter estimation problem into a state estimation problem. States and parameters are concatenated into a single higher-dimensional joint state vector from

$X = [x_1, x_2 \ldots x_m]$ to $X = [x_1, x_2 \ldots x_m, \beta_1, \beta_2 \ldots \beta_n]$, where m and n are the dimensions of x and β, which represent the state and the parameter, respectively. Thus, parameters are updated with state updates via EnKF, and the new analysis values of parameters act as forecast values until the next update process. A detailed description is in Monsivais-Huertero et al. [33].

2.3.3. Dual ensemble Kalman filter

The DEnKF process is designed to recursively estimate both states and parameters using two parallel filters. Furthermore, the update processes between states and parameters are mutually interactive each time new observations appear. Because this method has been completely described by Moradkhani et al. [35], only several key steps are given in this paper.

Parameters are sampled at each time step by Kernel Smoothing to avoid over-dispersion of the parameter ensemble and a loss of information between time points, leading to an inherent gradual change of the parameters.

$$\beta_{t+1}^{i-} \sim N(a\beta_t^{i+} + (1-a)\beta_t^+, h^2 V_t^+) \quad V_t^+ = \mathrm{var}(\beta_t^+) \tag{25}$$

where $a = \frac{3\delta-1}{2\delta}$ and $h = \sqrt{1-a^2}$; δ is a factor between 0 and 1.

The sampled parameter ensemble from above is used to force the model to generate forecast values of states, which are applied to an observation operator to produce the forecast observation values required in the Kalman gain equation for parameters.

$$K_{t+1}^{\beta} = \sum\nolimits_{t+1}^{\beta y} \left[\sum\nolimits_{t+1}^{yy} + \sum\nolimits_{t+1}^{y} \right]^{-1} \tag{26}$$

Then, a brand new parameter ensemble that is acquired from the first filter is imported to backroll the model to obtain new model state forecasts. Correspondingly, the state ensemble is similarly updated using the Kalman gain to correct the state trajectories.

$$K_{t+1}^{x} = \sum\nolimits_{t+1}^{xy} \left[\sum\nolimits_{t+1}^{yy} + \sum\nolimits_{t+1}^{y} \right]^{-1} \tag{27}$$

Σ is the operation used to calculate the covariance or cross covariance.

2.3.4. Simultaneous optimization and data assimilation

SODA is a combination of a parameter optimization algorithm and a data assimilation algorithm. An inner EnKF loop is set up for state estimation, while an outer stochastic global optimization loop produces an assumed parameter set under EnKF conditions by using the batch calibration strategy SCEM-UA. A detailed description of this method has been presented by Vrugt et al. [50] and mainly includes the following steps.

(1) Sample according to the prior probability distribution of a parameter.
(2) Run the model with parameter sets from step (1) in succession, while concurrently implementing EnKF for state updating.
(3) Sort all of the parameter sets according to a predefined cost function that reflects the discrepancy between the simulation and observation of a given parameter set. Subsequently, partition them into a specific number of complexes and shuffle the complexes with new candidates generated by SEM (Sequence Evolution Metropolis).
(4) Check the convergence criteria. If satisfied, terminate the iteration; if unsatisfied, return to step (2).

Theoretically, SODA uses calibration to correct long-term systematic biases due to parameter uncertainties and ensemble data assimilation to correct short-term or instantaneous system biases associated with model structure, input data, and other sources of errors.

2.4. Description of study area and data

CEOP (Coordinated Energy and water cycle Observations Project) provides consistent research-quality data sets with error descriptions of the Earth's energy budget and water cycle and their variability and trends on interannual to decadal time scales, with the aim of understanding and predicting continental to local-scale hydroclimates for hydrologic applications. In this study, we chose to evaluate the CEOP reference sites located in Mandal Govi, Mongolia, which are covered by a flat area of semiarid grassland. There are five automatic weather stations (AWS) and twelve automatic stations for soil hydrology (ASSH) in this area. The CEOP website (http://data.eol.ucar.edu/master_list/?project=CEOP/EOP-3/4) provides additional details about the corresponding instrumentation and data collection methodology.

Meteorological data (wind speed, air temperature, specific humidity, precipitation, and atmospheric pressure) that were measured directly at ground sites (BTS, DRS, DGS, and MGS) from October 1, 2002 (Julian day 274), to September 30, 2003 (Julian day 273), were used to drive the CoLM. Missed *in situ* observations and downward shortwave and longwave radiation were obtained from hourly intermediate gridded model output data produced by the JMA (Japanese Meteorological Agency) global data assimilation system. JMA-MOLTS is a shortened product from the six-hour model forecast in the data assimilation cycle that has responded to the CEOP model output request since Oct 1st, 2002. Soil temperature and soil moisture measurements are available for BTS at depths of 3, 10, 20, and 40 cm and for the other three stations at depths of 3, 10, 40, and 100 cm.

To take atmospheric forcing data uncertainty into consideration, another series of meteorological data that were extracted from GLDAS (Global Land Data Assimilation System) were employed, with supplementary precipitation from the TRMM (Tropical Rainfall Measurement Mission) because the precipitation data from GLDAS are missing. All of the meteorological data sets are consolidated into hourly data; thus, the CoLM runs and outputs hourly.

2.5. Experimental design

2.5.1. Assimilation experiments

Given the limited comprehensive annual atmospheric data due to data gaps caused by the failure of instruments, one year of repeated forcing data are required to spin-up the CoLM in a 20-year period. This study did not consider the frozen soil process; thus, the assimilation experiments were conducted over 120 days, commencing on May 31, 2003 (Julian day 151). The solum of the model was restratified to conform to the observation depths at the four ground sites before the assimilation experiments and all ten layers of soil moisture compose the state vector to be estimated. For the purpose of evaluating the feasibility of AEnKF, DEnKF and SODA, soil moisture and brightness temperature assimilation experiments were designed, and three cases were performed according to the different sources of uncertainty in both experiments.

Case 1. We hypothesized that the discrepancy that existed between the simulation and the actual situation originated only from parameter uncertainty. Synthetic experiments provided a controlled environment to mimic reality but focused on the parameter uncertainty. Therefore, the initial conditions for the reference run and the experimental run were derived from the spin-up period that was implemented with 'true' parameters (determined through an experiment performed by Yang et al. [56]) and parameter sets that were uniformly sampled in the specific ranges (defined according to the model default and the physical range). The 'true' observations were generated by the reference run and were imported into RTM to export 'true' brightness temperatures. All of the experiments were forced by the

same forcing data set that was observed *in situ* to disregard the uncertainties in the forcing data.

Case 2. We hypothesized that a discrepancy existing between the simulation and the actual situation originated from both parameter and atmospheric forcing data uncertainties. The "true" states and brightness temperature were produced in the same manner as those in Case 1. The only difference between Case 1 and Case 2 is that the meteorological data that were used in the experimental runs in Case 2 were produced from GLDAS and TRMM to include atmospheric forcing data uncertainty. Apart from that, the postulates were the same as those for Case 1.

Case 3. We hypothesized that parameter, atmospheric forcing data and model uncertainties together contributed to the discrepancy existing between the simulation and the actual situation. Due to these uncertainties, the measured soil moisture will be inconsistent with the simulated soil moisture. Therefore, the measured soil moisture at each site was taken as the "truth" in this case, and the other conditions are the same as those in Case 1. Furthermore, synthetic brightness temperature was generated by replacing soil moisture and soil temperature (the input variables of RTM) with the corresponding *in situ* values.

In the soil moisture assimilation experiment, the first layer soil volumetric water content from the reference run (Case 1 and Case 2) or the *in situ* observation (Case 3) served as the "truth" observation, with the observation ensemble being acquired by disturbing the "truth" with a 10% multiplicative error, and the assimilation frequencies being fixed to 0.5 day, 1 day, 3 days and 8 days to investigate the effect of the quantity of observations. In this part of the experiment, soil texture (volume percentage of sand, volume percentage of clay) and porosity from the CoLM are estimated using observed soil moisture. To guarantee the physical meaning of the parameters in the update process, the sand and clay content were re-checked to ensure that their sum was less than 100%. If the sum exceeded 100% after updating, the sand and clay content were adjusted by subtracting the quantity$((\text{sand}+\text{clay})-100)/2.0$. In addition, "Sim" represents the ensemble average of the model simulations without assimilation using the initial state and parameter condition stemmed from the spin-up for the experimental run (assimilation experiments).

In the brightness temperature assimilation experiment, the synthetic brightness temperature observation was assumed to have white Gaussian noise with a standard deviation of 2 K. Because of the high sensitivity of horizontal polarization to vegetation and the low sensitivity of higher frequencies to soil moisture, 6.9 and 10.7 GHz frequencies with vertical polarization (hereafter referred as 6.9 V and 10.7 V) were merged via assimilation to retrieve states and parameters. In this paper, the experimental results were compared among different frequency combinations (6.9 V, 10.7 V, 6.9 V + 10.7 V). To imitate authentic conditions, synthetic brightness temperatures were treated as AMSR-E data and were assimilated at 1:00 a.m. (descending orbit) every day. The parameters to be estimated in this part of the experiment include volume percentage of sand, volume percentage of clay and porosity, as well as the standard deviation of the surface height from the RTM.

The initial state ensemble is selected for up to 50 realizations for all cases in the spin-up period before the assimilation experiment. The outputs from the CoLM in the experimental run were postulated to be mostly influenced by the parameters and forcing data in our study. To account for errors associated with forcing data, normally distributed additive perturbations or log-normally distributed multiplicative perturbations were applied, depending on the variable. Under that assumption, a positive perturbation of the downward shortwave radiation tends to be associated with negative perturbations of the longwave radiation and the precipitation, and vice versa. Table 1 displays the standard deviations and cross-correlations for the perturbations of precipitation, shortwave radiation, longwave radiation and air temperature. Table 2 lists the specified 'true' parameters and sampling range used in the three cases. As a result, each of the mean values of the perturbed factors was equal to zero for the additive case and to one for the multiplicative case.

Table 1
Summary of perturbation parameters for atmospheric forcing data and the cross correlation coefficients used to generate random perturbations for the different variables.

Variables	Noise type	Standard deviation	Cross correlation
Precipitation	Multiplicative	0.5	[1.0 −0.8 0.5 0.0,
Shortwave radiation	Multiplicative	0.3	−0.8 1.0 −0.5 0.4,
Longwave radiation	Additive	30 W/m^2	0.5 −0.5 1.0 0.4,
Air temperature	Additive	2 K	0.0 0.4 0.4 1.0]

2.5.2. Evaluation metrics

To assess the performance levels of the different algorithms (the results of soil moisture and other essential states in water and energy balance), we defined several evaluation factors, including the root mean square error (RMSE), the mean bias error (MBE) and the correlation coefficient (R), which are described as follows:

$$RMSE=\sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t-X_{true,t})^2} \tag{28}$$

$$MBE=\frac{1}{T}\sum_{t=1}^{T}(X_t-X_{true,t}) \tag{28}$$

$$R=\frac{\sum_{t=1}^{T}(X_t-\overline{X})(X_{true,t}-\overline{X_{true}})}{\sqrt{(X_t-\overline{X})^2(X_{true,t}-\overline{X_{true}})^2}} \tag{29}$$

in which T is the total number of steps, and X_t and $X_{true,t}$ represent the simulation or assimilation values of states in different cases and the true values at step t, respectively. The horizontal line above the expression indicates mean value.

To deploy the widespread application of simultaneous state and parameter estimation into fields involving tremendous data, such as in regional data assimilation, experiments spanning long spans of time and coupling with other models, the computational time (CT) was also used to compare computational efficiencies. The algorithms that were run with a 2.80-GHz Intel 860 CPU and CT are basically counted in seconds, except that the time-consuming SODA is calculated in minutes.

3. Results and discussion

To evaluate the validity of each of the three above-discussed methods of resolving the parameter problem, we conducted three cases of both soil moisture and brightness temperature assimilation, and the results of these experiments were subsequently analyzed and compared. The four sites showed similar results in most cases; therefore, the results for BTS are considered to be representative in the following discussion.

3.1. Assimilation of surface soil moisture

Table 3 displays the statistical indicators RMSE and MBE of the soil moisture predictions in the first layer derived from Case 1. It is apparent that all three methods perform well in state estimation, while the RMSEs of AEnKF, DEnKF and SODA change from 0.0047, 0.0048 and 0.0036 to 0.0109, 0.0120 and 0.0065, respectively, with the observation interval varying from 0.5 to 8 days. It is well known that a cutback in observation quantity can degenerate assimilation

Please cite this article as: W. Chen et al., Comparison of ensemble-based state and parameter estimation methods for soil moisture data assimilation, Advances in Water Resources (2015), http://dx.doi.org/10.1016/j.advwatres.2015.08.003

Table 2
List of the specified 'true' parameters and sampling range used in the three cases.

Sites	Parameters				Longitude	Latitude
	Sand	Clay	Porosity	RMS(cm)		
BTS	47	28	0.456	0.06	107°08′32.2″	46°46′35.4″
DGS	47	28	0.377	0.06	106°22′06.8″	46°07′38.3″
DRS	47	28	0.4105	0.06	106°42′53.0″	46°12′31.2″
MGS	47	28	0.361	0.06	106°15′52.2″	45°44′34.9″
Sampling range	[0,60]	[20,80]	[0.1,0.5]	[0.01,0.2]	/	/

Table 3
The RMSEs and MBEs of soil moisture varying with observation intervals between the results of simulation/assimilation and the 'truth' during May 31, 2003 to September 27, 2003 in Case 1. Note: soil moisture serves as observation.

BTS	RMSE				MBE			
	0.5D	1D	3D	8D	0.5D	1D	3D	8D
Sim	0.0161				0.0108			
AEnKF	0.0047	0.0043	0.0076	0.0109	−1.3e−04	5.5e−04	6.0e−04	−0.0025
DEnKF	0.0048	0.0053	0.0067	0.0120	1.6e−05	3.8e−04	0.0011	−0.0018
SODA	0.0036	0.0038	0.0055	0.0065	0.0019	0.0013	0.0026	0.0031

results. All three methods showed similar performance for the densest observations, but SODA outperformed the other two in cases with less observations, mainly because the optimization process that is embedded in SODA uses information regarding entire time series. The extremely small values of MBEs from AEnKF and DEnKF do not contradict the above conclusions, but the continuous adjustment of parameters with AEnKF and DEnKF causes an offset of model parameters. In addition, SODA provides a unique set of optimum parameters during the assimilation period, causing a deviation in the same direction, and only RMSE acts as an objective function. The superiority of AEnKF and DEnKF has been discussed since the two methods first appeared. In principle, AEnKF should provide better estimates of states and parameters than DEnKF because it explicitly accounts for their cross-covariance. However, the estimation process can lead to unstable results because of complex interactions between states and parameters in nonlinear dynamic systems. In Case 1, AEnKF produces more accurate soil moisture estimations than DEnKF, which can be ascribed to the simple hypothesis of sole uncertainty from model parameters. The SODA-optimized values for sand content, clay content and porosity generated by one observation per day are 46.98, 27.54 and 0.45, respectively. Correspondingly, the parameters of AEnKF and DEnKF gradually stabilized to 37.77, 30.30, 0.40 and 39.24, 31.76, 0.38. SODA produces the best parameter estimation because of its more extensive usage of information, but AEnKF and DEnKF progressively converge more closely to the 'true' values.

For Case 2, forcing data uncertainty is introduced into the experiments, and the characteristics of the results agree with the conclusions drawn from Case 1. Compared with the 'true' soil moisture from layer one, AEnKF (DEnKF, SODA) reduces the RMSE from 0.0478 (Sim) to 0.0364 (0.0349, 0.0351) when the observation interval is one day. AEnKF suffers from uncertainties arising from multiple aspects, but the three methods still produce superior soil moisture estimates relative to those produced by the simulation. Parameter estimation leads to excellent results from SODA for sand content, clay content and porosity, which were 47.72, 27.77, and 0.46, respectively, while the corresponding values from AEnKF (DEnKF) are approximately 38.24, 32.37 and 0.39 (50.29, 29.80 and 0.46).

The soil moisture assimilation results from four layers (3, 10, 20, and 40 cm) compared with the CoLM simulations, and *in situ* measurements from Case 3 are shown in Fig. 2. AEnKF, DEnKF and SODA improve open loop estimations of surface-layer soil moisture to a large extent, the curves of which draw near to the true values. In the deeper layers (Fig. 2(b)–(d)), all of the assimilation algorithms performed fairly well compared with the simulations, except that AEnKF produced very little improvement. This is illustrated by the parameter estimations. Table 4 shows quantitative analyses of the results shown in Fig. 2(a). All three of the methods considerably reduce the RMSEs by more than 50% (52% for AEnKF, 60% for DEnKF and 63% for SODA). The MBEs show great improvement in assimilation, which confirms the RMSE evaluation. The correlation coefficient (R) obtained with SODA (AEnKF and DEnKF) increased to 0.7968 (0.7449 and 0.7678) from 0.6291. Taking computational time into account, which is also displayed in Table 4, AEnKF and DEnKF are highly efficient, but DEnKF has a higher CT value because of its two separate update procedures for states and parameters. The optimization procedure with repeated resampling of SODA cuts down its computational efficiency. The positive influence of observation quantity on assimilation was verified in Case 1, and the consistent results in Cases 2 and 3 using different observation intervals are omitted.

Figs. 3 and 4 describe the parameter estimations and evolution processes for the parameter ensembles corresponding to the above results for the states. It can be observed from Fig. 3 that DEnKF and SODA produce similar parameter estimations, while AEnKF shows a bit of discrepancy. As previously observed, overestimated porosity leads to overestimated soil moisture in the deeper layers because soil porosity determines saturated water content. The larger the porosity, the more water the soil is able to store. Considering model structure error, the optimum parameters obtained by the three methods apparently differ from *in situ* measurements. The reliability of each of the three parameter sets is discussed in the final part of this section. The 50th and 100th percentile confidence intervals are indicated in Fig. 4 by varying grayscales. The convergence for AEnKF is worse than for DEnKF, especially when observational information is deficient (not shown). DEnKF can avoid the static evolution of parameter

Table 4
Statistic indicators of soil moisture between the results of simulation/assimilation and the measurements in situ during May 31, 2003 to September 27, 2003 in Case 3. Note: soil moisture serves as observation one per day.

	RMSE	MBE	R	CT
Sim	0.0578	0.0508	0.6291	14.0 s
AEnKF	0.0276	0.0108	0.7449	15.4 s
DEnKF	0.0230	0.0049	0.7678	31.5 s
SODA	0.0211	0.0053	0.7968	126 m

Please cite this article as: W. Chen et al., Comparison of ensemble-based state and parameter estimation methods for soil moisture data assimilation, Advances in Water Resources (2015), http://dx.doi.org/10.1016/j.advwatres.2015.08.003

Table 5
The RMSEs and MBEs of soil moisture varying with observation frequencies between the results of simulation/assimilation and the 'truth' during May 31, 2003 to September 27, 2003 in Case 1. Note: brightness temperatures serve as observation.

BTS	RMSE			MBE		
	6.9 V	10.7 V	6.9V + 10.7 V	6.9 V	10.7 V	6.9V + 10.7 V
Sim	0.0161			0.0108		
AEnKF	0.0064	0.0079	0.0063	−0.0029	0.0043	−0.0016
DEnKF	0.0083	0.0187	0.0082	0.0061	0.0171	0.0038
SODA	0.0072	0.0070	0.0068	0.0033	0.0043	0.0022

ensembles in a few observations because of the inherent convergence of the Kernel Smoothing Sampling that is implemented within it.

The Taylor diagrams pictured in Fig. 5 show the results for all four study sites in Case 3 when assimilating soil moisture once per day. Taylor diagrams [43] provide a way of graphically summarizing how closely a pattern (or a set of patterns) matches a set of observations. The position of each letter on the plot summarizes the performance of the corresponding experimental result. The similarity between two groups of soil moisture is quantified in terms of the correlation (R), the centered root-mean-square difference (RMSD) and the amplitude of their variations, represented by their standard deviations (SD). 'OBS' indicates the *in situ* measurement values, and the letters 'A', 'B', 'C' and 'D' refer to the results from Sim, AEnKF, DEnKF and SODA, respectively. The green contours indicate the RMSD values, which are proportional to the distance to a point on the *x*-axis that is identified as 'OBS'. For BTS, it can be observed that SODA has the best agreement with the observations, as assessed by the most approximate SD value (0.0322) (SD of the observations indicated by dashed arc), with the largest *R* value and smallest RMSD value (0.0204). AEnKF and DEnKF show similar R values; however, DEnKF produces better agreement with the observations (with a slightly lower RMSD of 0.0225 and an SD of 0.0340). For the other sites, each of the three methods has approximately the same correlation with the observations (0.9), which is distinctly better than that produced by Sim. Although there is not much difference between the RMSD values for AEnKF, DEnKF and SODA in DGS and DRS, DEnKF has less variability (0.0318 in DGS, 0.0308 in DRS) compared to the observed values (0.0359 in DGS, 0.0358 in DRS) than the others. AEnKF performed poorly for MGS, with an RMSD of 0.0280 and an *R* of 0.8676, both of which are inferior to DEnKF and SODA. The latter two methods both have observation correlations of more than 0.88 and RMSD values less than 0.02.

3.2. Assimilation of AMSR-E brightness temperature

Table 5 displays the RMSE and MBE values in the first layer of Case 1, in which the observations are set as 6.9 V, 10.7 V, and both 6.9 V and 10.7 V. SODA settles the RMSE in the proximity of 0.007 and shows a slight advantage when two frequencies are used simultaneously (0.0068). The RMSEs of AEnKF (DEnKF) range from 0.0064 (0.0083) to 0.0079 (0.0187), reaching minimums of 0.0063 and 0.0082 when assimilating both 6.9 V and 10.7 V. This clearly indicates that AEnKF is always better than the other two methods, which have finite uncertainties. Assimilating brightness temperatures at 6.9 V achieves better performance than doing so at 10.7 V without exception because low frequency is more sensitive to changes in soil water content and standard deviation of surface height. The improvement of statistical indicators from Sim demonstrates that more accurate soil moisture estimation is achieved when parameter estimation is taken into consideration. MBE confirms the analysis of RMSE, and the SODA parameters (47.05 for sand, 27.83 for clay, 0.46 for porosity and 0.066 for RMS) were closest to the 'true values' in the 6.9 V assimilation experiment. The same parameters for AEnKF (DEnKF) converged at 39.57 (38.50), 27.91 (30.24), 0.43 (0.43) and 0.080 (0.085).

Table 6
Statistic indicators of soil moisture between the results of simulation/assimilation and the measurements in situ during May 31, 2003 to September 27, 2003 in Case 3. Note: brightness temperatures of 6.9 GHz serve as observation.

	RMSE	MBE	*R*	CT
Sim	0.0578	0.0508	0.6291	14.6 s
AEnKF	0.0319	−0.0071	0.3990	16.0 s
DEnKF	0.0312	−0.0120	0.4811	33.2 s
SODA	0.0281	0.0030	0.5436	158 m

In Case 2, in which brightness temperature was assimilated at 6.9 V, the soil moisture RMSE values of AEnKF and DEnKF drop from 0.0535 of Sim to 0.0407 and 0.0375. SODA reaches the lowest RMSE of 0.0332 and produces a precise estimation of soil parameters (47.20 for sand, 33.96 for clay, 0.46 for porosity and 0.059 for RMS). Moreover, the porosity obtained from AEnKF (DEnKF) still produces a satisfying value of 0.47 (0.49), but sand, clay and RMS steadily reach unrealistic values of 37.65, 54.87 and 0.030 (58.05, 38.58 and 0.100), respectively.

In Fig. 6 (a)–(d), the *in situ* measurements and soil moisture curves from Case 3 at four different depths (3, 10, 20, and 40 cm) are plotted, where the brightness temperature was assimilated at 6.9 V. AEnKF, DEnKF and SODA each enhance soil moisture estimation not only at the surface layer but also at deeper layers on the basis of state-parameter estimation. The reliable estimations of the three algorithms for the deeper layers mostly depend on the settlement of biased parameters. Table 6 provides further insights into the statistical results in Fig. 6. AEnKF, DEnKF and SODA successively reduce the RMSE values from 0.0578 to 0.0319, 0.0312 and 0.0281, and both RMSEs and MBEs demonstrate remarkable improvement in soil moisture estimation. Nevertheless, the correlation coefficients for all three algorithms are far from satisfactory, showing no enhancement over Sim. As can be observed in Fig. 6, the lag of the truth curve results in poor *R* values. Compared with the precipitation shown in Fig. 1, we are of the opinion that the new sets of parameters estimated by the three algorithms quickly respond to the forcing data, while in reality the delivery of soil water happens at specific times after rainfall.

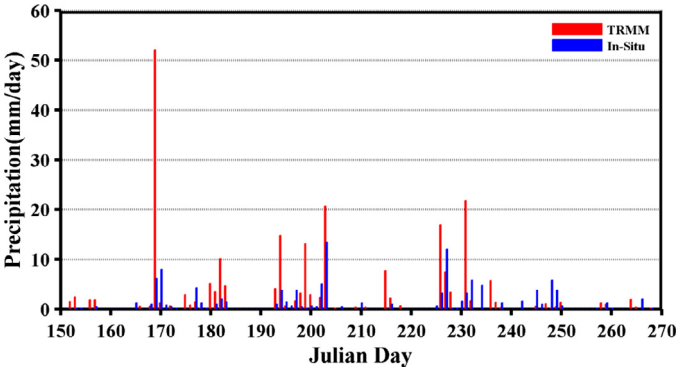

Fig. 1. Daily precipitations of AWS (*In-Situ* observation) and TRMM in BTS site.

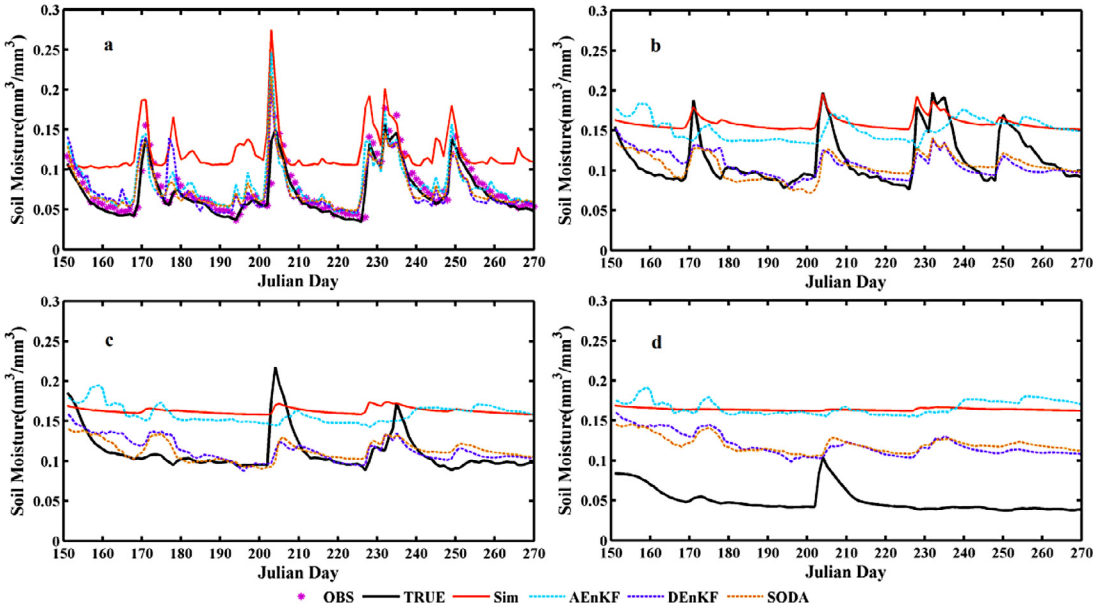

Fig. 2. Comparisons of day-average soil moisture of simulation (Sim), both states and parameters update (AEnKF, DEnKF and SODA) with the measurements in situ corresponding to Table 4. (a) 5 cm, (b) 10 cm, (c) 20 cm, (d) 40 cm.

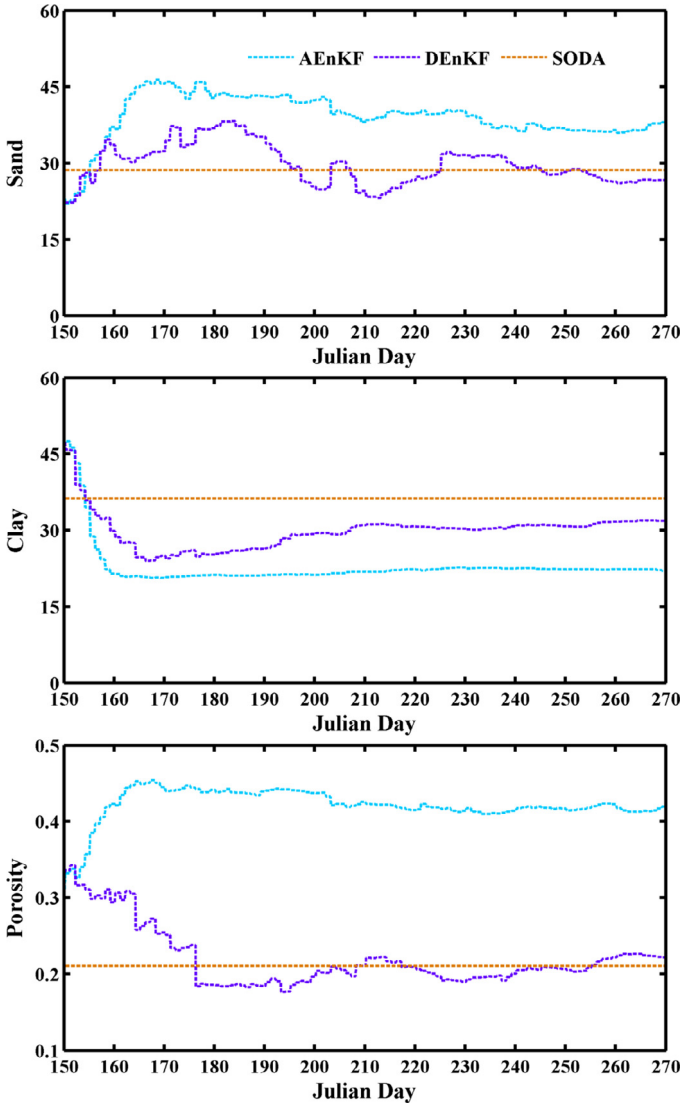

Fig. 3. The results of parameters with assimilating soil moisture one per day of BTS in Case 3.

Similar to the soil moisture assimilation experiment, the computation time of AEnKF, DEnKF and SODA in this case increases progressively, as shown in Table 6.

The parameters displayed in Fig. 7 correspond to the aforementioned experimental results shown in Fig. 6. It is clear that the sand content, clay content and porosity of AEnKF and DEnKF present significant coincidence and that SODA strongly improves the estimation values, except the extreme RMS. Parameterization contributing to model error determines the nonconformity between the calculated parameters and the true parameters. The reliability of the three sets of parameter values will also be discussed in the final part of this section. Fig. 8 illustrates the 50th and 100th percentiles of the parameter ensembles and shows that the evolution speed of AEnKF is similar to that of DEnKF. It is apparent that parameter variability appears to converge when assimilating brightness temperatures.

The Taylor diagrams (Fig. 9) show the RMSD, SD and *R* values for soil moisture at the four sites assessed in Case 3 when the brightness temperature was assimilated at 6.9 V. 'OBS' indicates the true *in situ* values, and the letters 'A', 'B', 'C' and 'D' refer to Sim, AEnKF, DEnKF and SODA, respectively. The correlation coefficient for SODA improved from 0.7773 to 0.8250 for DRS, while the *R* values of the other algorithms or at the other sites demonstrate a failure in enhancing the matching of the assimilated and measured values. Based on the values included in Table 6, this consequence does not deteriorate the other statistical indicators, as in BTS. Actually, discussion regarding the other sites has been omitted because of the striking similarity with BTS in terms of RMSE. Considering the overall performance levels of the three algorithms, SODA stands out for its distinctive parameter optimization, and DEnKF is superior to AEnKF when assessing complicated conditions.

Six sets of parameter estimation values (three algorithms multiplied by two types of observation assimilations) from Case 3 were imported into the CoLM to replacing the original values that were used to simulate soil moisture. These soil moisture results were compared with the *in situ* measurements that are displayed in Fig. 10. A rapid decrease in soil moisture occurs in the first layer at the beginning of assimilation period, as shown in Fig. 10(a) and (b), revealing a rapid modification from the overestimation of open loop values. In Fig. 10(a), the curves produced by DEnKF and SODA are well-matched, and their tight connection with the black curve (representing the *in situ* measurements) demonstrates that the parameters are trustworthy. The overestimation of porosity by AEnKF

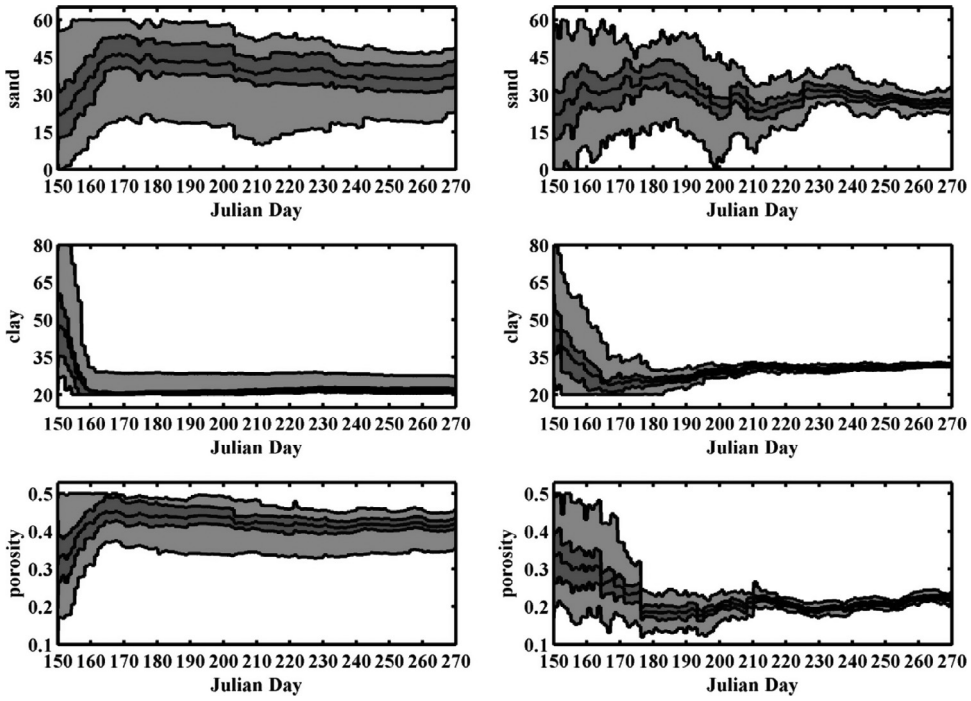

Fig. 4. The parameters ensemble evolutions of AEnKF (left) and DEnKF (right) corresponding to Fig. 3.

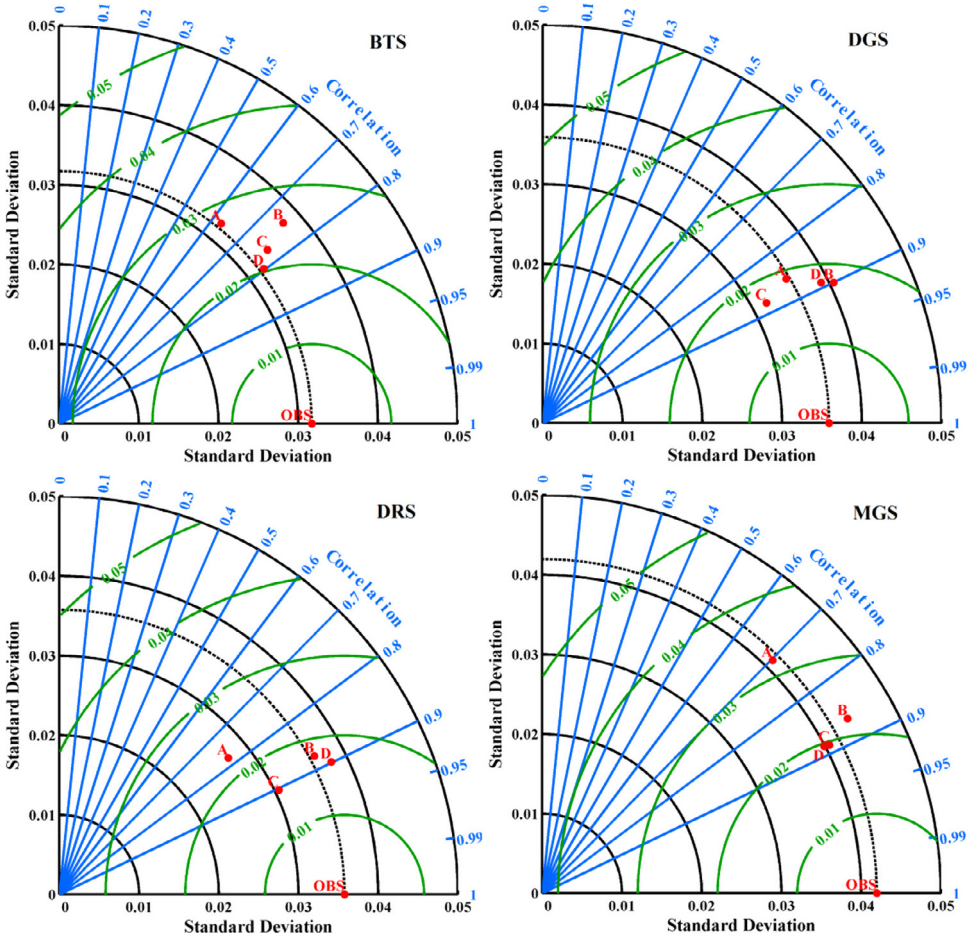

Fig 5. Taylor diagrams of soil moisture of BTS, DGS, DRS and MGS in Case 3 with assimilating soil moisture once a day ('A' to 'D' refer to the results from Sim, AEnKF, DEnKF and SODA, respectively).

Please cite this article as: W. Chen et al., Comparison of ensemble-based state and parameter estimation methods for soil moisture data assimilation, Advances in Water Resources (2015), http://dx.doi.org/10.1016/j.advwatres.2015.08.003

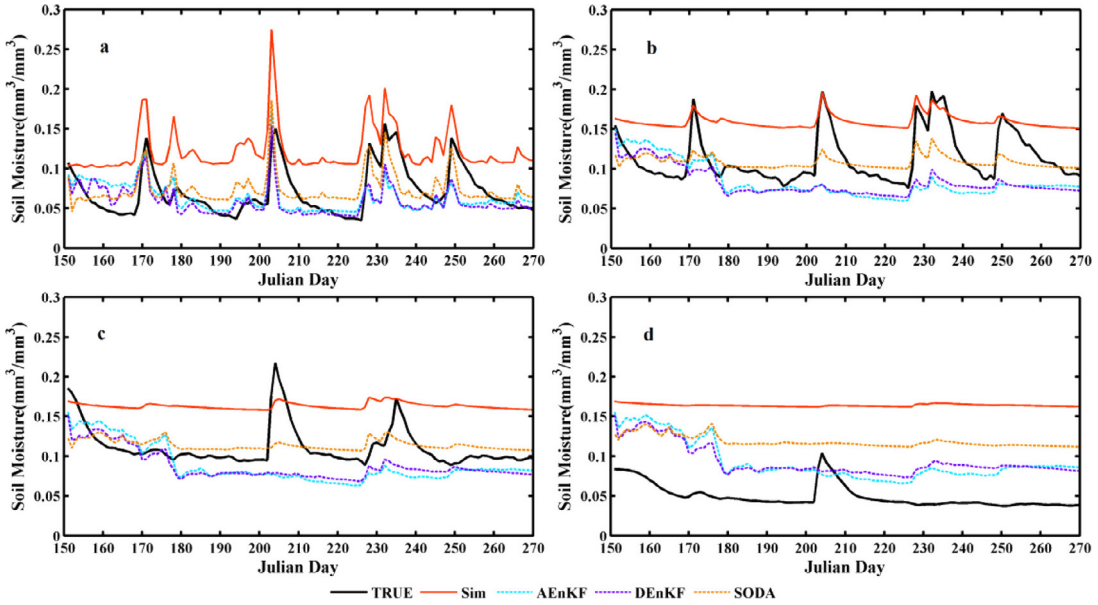

Fig. 6. Comparisons of daily mean soil moisture of simulation (Sim), both states and parameters update (AEnKF, DEnKF and SODA) with the measurements in situ corresponding to Table 6. (a) 5 cm, (b) 10 cm, (c) 20 cm, (d) 40 cm.

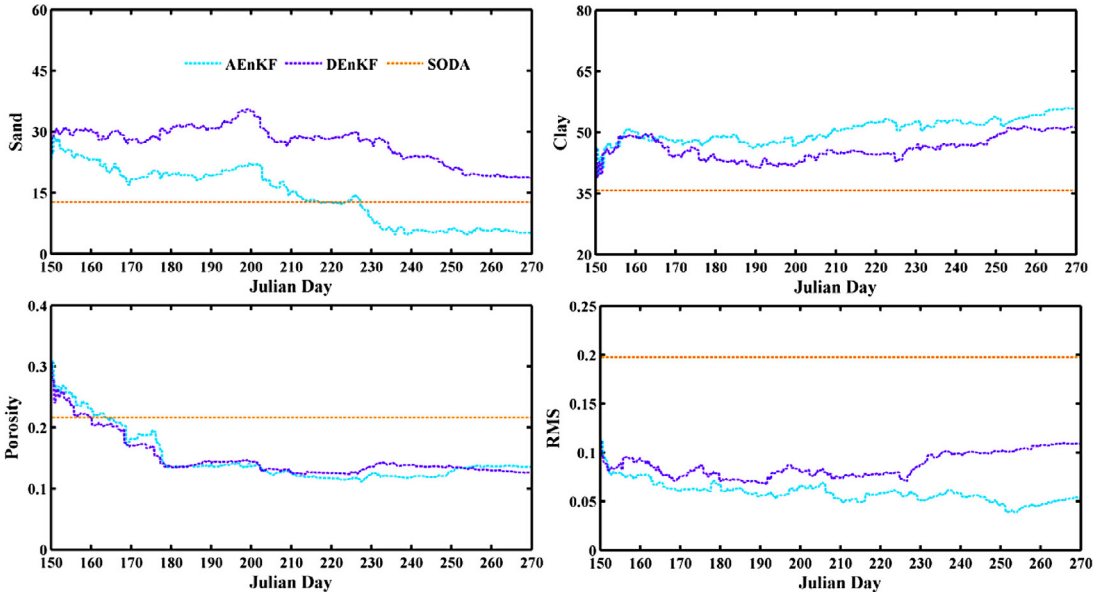

Fig. 7. The results of parameters assimilating 6.9 GHz brightness temperature of BTS in Case 3.

is attributed to an overestimation of soil moisture. In Fig. 10(b), the curve generated by SODA is close to the truth and to the other two curves and is also similar to the curve in Fig. 10(a) due to the similar parameter estimations that were used. Careful inspection of the behavior patterns of the states and parameters confirms the opinion that it is rare for either to simultaneously reach interconsistency with the truth due to the model structure error.

Another possible reason for such results is that *in situ* measurements of soil water dynamics in arid areas contain insufficient information to warrant a successful estimation of parameters [42]. As shown in Figs. 3 and 7, the maximum surface soil water content is less than 0.2, leading to an absence of soil water saturation at the indicated porosity estimation. The factors that contribute toward confining the porosity to approximately 0.2 may also in many occasions affect other parameters. A wide range of soil moisture states is required to reliably constrain soil hydraulic function. Moreover, the use of a single metric also conspires against achieving the desired identifiability, as noted by Vrugt et al. [51].

3.3. Issues related to state-parameter estimation

A remarkable systematic deviation in soil moisture exists in the original simulation used in this study, as shown in Figs. 2 and 6. It is clear that EnKF alone is outperformed by all other methods (not shown), as it consistently maintains a bias. Only state update (hereafter referred as EnKF) is inappropriate under the framework of EnKF, which has a basic assumption that both prediction and observation are unbiased. Although the state results (soil moisture) of EnKF appear to be reasonable and more valid than open loop values (not shown), they violate the basic assumption of Kalman filter approaches, which is also reflected in the inferiority of the results produced by AEnKF, DEnKF and SODA in most cases. Therefore, a joint

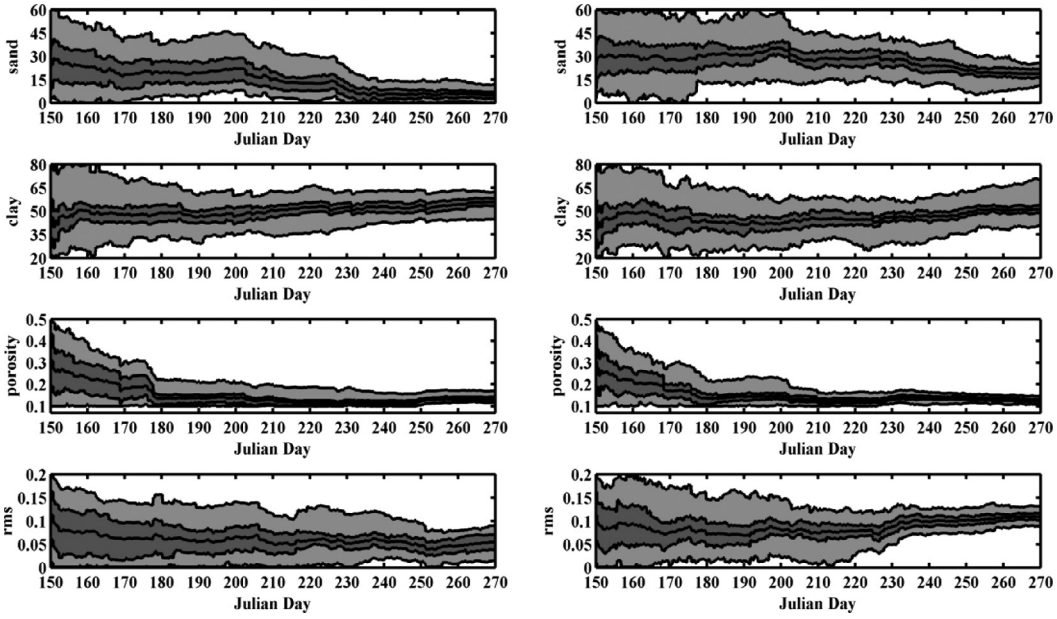

Fig. 8. The parameters ensemble evolution of AEnKF (left) and DEnKF (right) corresponding to Fig. 7.

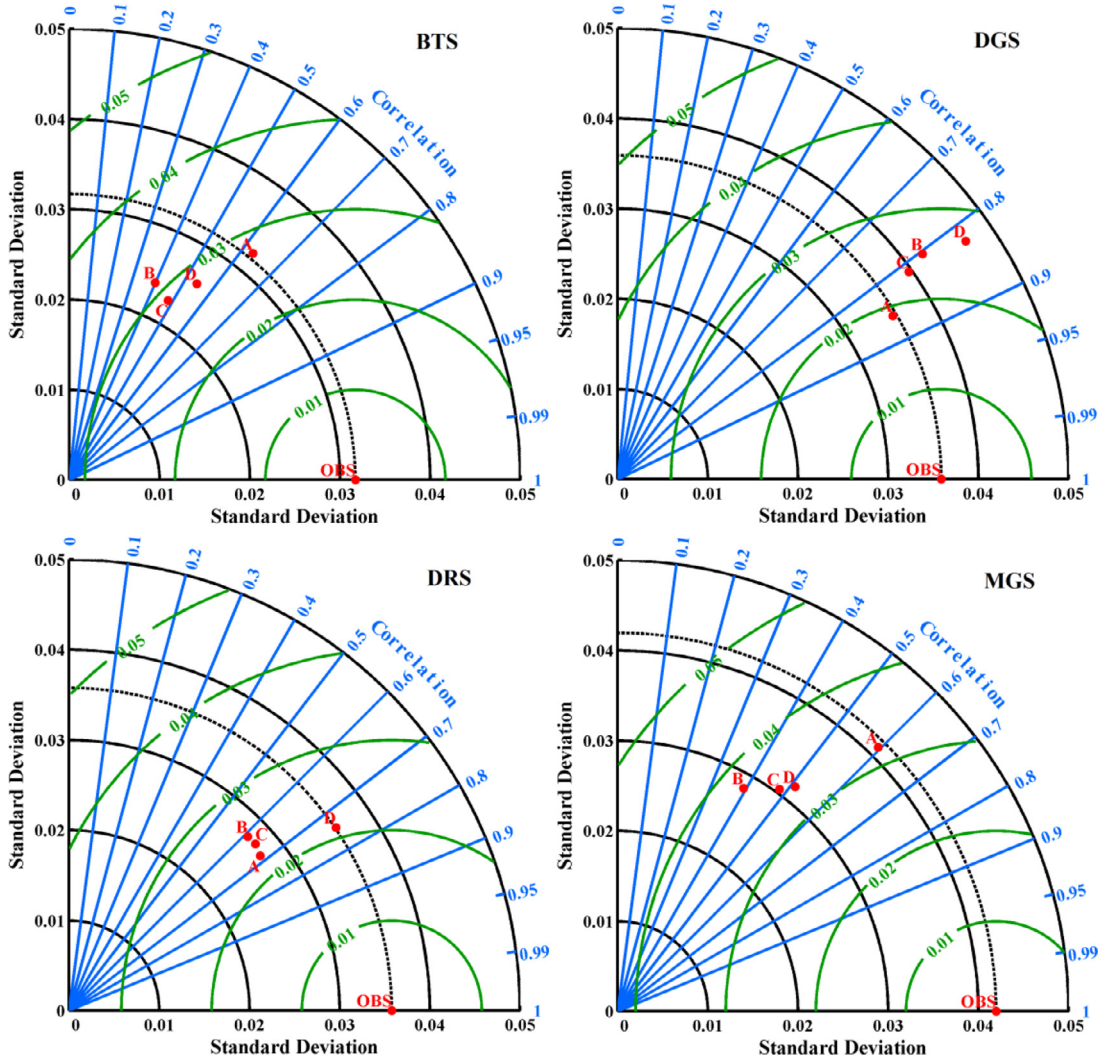

Fig 9. Taylor diagrams of soil moisture of BTS, DGS, DRS and MGS in Case 3 with assimilating 6.9 GHz brightness temperature ('A' to 'D' refer to the results from Sim, AEnKF, DEnKF and SODA, respectively).

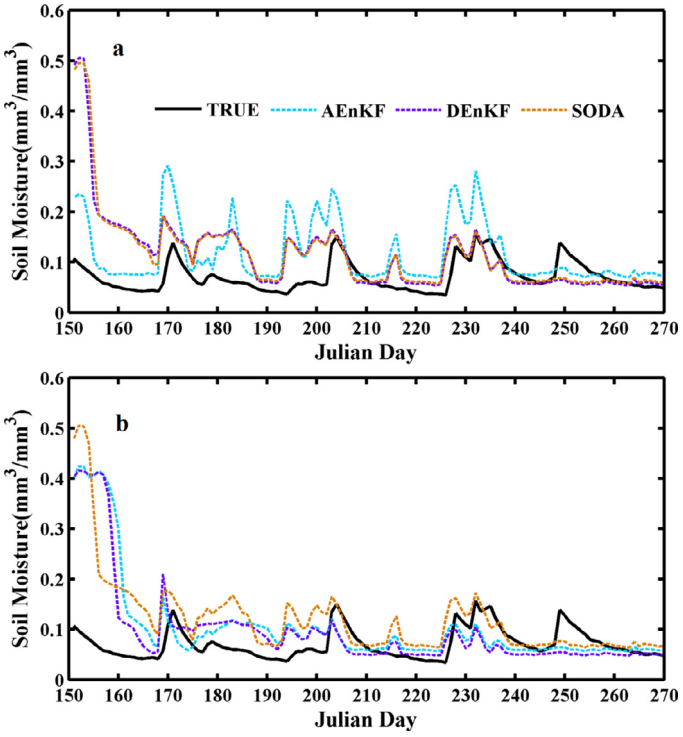

Fig 10. Comparisons of the simulations calculated by the parameters set derived from AEnKF, DEnKF and SODA with assimilating (a) soil moisture and (b) brightness temperature in Case 3 to the measurements in situ.

states-parameter estimation is essential to correct the long-term bias produced by parameter estimation and to correct the random error produced by state assimilation when a model simulation is subjected to a large bias. Unlike hydrologic models that use historical discharge data to calibrate parameters before assimilation, a calibration process is rarely carried out with land-data assimilation. The intention of parameter estimation in this study was to correct parameters in a land-surface model. This serves the same purpose as calibration in hydrologic models, with the only difference being that this process is carried out during assimilation instead of preceding it. In addition, this is the exact signification of using joint states-parameters estimation methods in land-data assimilation, as it does not require substantial historical data.

A noticeable issue of Dual-EnKF is that the same observation is used twice during one assimilation process (first updating parameters, then updating model states), and the updated parameters indeed impact the model states to be updated. This may result in an over-correction issue. However, in our experiments, although the impact of overusing observational information may occur at the beginning of an experimental window, when the parameters progressively converge to stable values, the observations have little effect on the parameters and therefore this type of impact gradually vanishes. Moreover, DEnKF still shows good performances in both soil moisture and brightness temperature assimilation experiments, which confirms the feasibility of DEnKF. The main problem that exists in DEnKF is that model states may not represent an unbiased estimation because of the excessive usage of observation in this method.

Computational efficiency is also a remarkable issue, especially for an operational data assimilation system. SODA generally achieved the best performance in estimating model parameters and states in all of the experiments, which is because in principle SODA undergoes calibration and thus removes bias due to parameter errors by using the information contained within an entire time series, while AEnKF and DEnKF only employ observational information that was collected at the time of an available observation. It seems that AEnKF and DEnKF are not comparable to SODA. However, AEnKF and DEnKF also achieved acceptable levels of accuracy in these experiments and have their own advantages, including flexibility, implementation simplicity, and reduced inner storage and operational time requirements. These benefits make AEnKF and DEnKF more suitable for operational data assimilation systems on regional and global scales.

4. Conclusions

In this paper, three simultaneous state-parameter estimation algorithms were applied to both soil moisture and brightness temperature assimilation to compare their performance levels. Each has its own pros and cons. First, AEnKF is highly efficient at producing accurate results of both states and parameters under situations in which an explicit relationship exists between them. However, when a high nonlinearity exists between states and observations (brightness temperature assimilation) or multiple uncertainties make the relationship between states and parameters ambiguous (Cases 2 and 3), the superiority of AEnKF vanishes. In addition, AEnKF shows no improvement in cases with sparse observations. Second, SODA requires a lot of time, although it produces the most accurate states and parameters in complicated cases (especially in Case 3), and shows a slight change when the quantity and sensitivity of observations declines (the 8-day soil moisture and the vertical brightness temperature of 10.7 GHz assimilation), both on account of its optimization process. Third, DEnKF possesses similar characteristics to the other two. In short, DEnKF possesses adaptability to conditions in which AEnKF loses effectiveness, but it never transcends SODA in terms of accuracy. However, considering the computational time requirement of SODA, DEnKF still has relative merit. Thus, under the condition that a numerical model or simple physical model is used for data assimilation or when an obvious relationship exists between states and parameters, AEnKF is the optimal choice. However, on a large-scale or under a long-term assimilation timeframe, DEnKF should be chosen because of its high efficiency and excellent error handling in intricate situations. SODA should be selected when pursuing accuracy in the case of scarce observations, regardless of its computational time requirements.

We chose soil texture as a parameter in this study and adopted internal parameterization to calculate soil thermal and hydraulic parameters. However, whether such parameterization should be inherent in a model remains questionable; thus, direct estimations of soil thermal and hydraulic parameters or alterations in parameterization will be addressed in future work. As discussed above, the limited variability in the observations that were assimilated can definitely affect proper parameter identification. Therefore, using more than a single state variable may warrant a reliable estimation of parameters.

In terms of brightness temperature assimilation, there is still much work to be performed in the future. Soil temperature as an imported variable of a radiation transfer model is also a significant factor that determines brightness temperature. Moreover, the details of the Radiative Transfer Model used here should be adjusted for different regions; for example, microwaves have deeper penetration depths in arid areas. To eventually assimilate brightness temperatures observed from satellites, such issues should be explored in our ongoing research.

Acknowledgments

This work is supported by the National Science Foundation of China under Grants 91325106 and 41271358, the Major State Basic Research Development Program under Grant 2011CB707103, the Hundred Talent Program of the Chinese Academy of Sciences under grant 29Y127D01, and the Cross-disciplinary Collaborative Teams Program for Science, Technology and Innovation of the Chinese Academy of Sciences.

References

[1] Bai Y, Li X. Evolutionary algorithm-based error parameterization methods for data assimilation. Mon Weather Rev 2011;139(8):2668–85. http://dx.doi.org/10.1175/2011MWR3641.1.
[2] Bateni SM, Entekhabi D. Surface heat flux estimation with the ensemble Kalman smoother: joint estimation of state and parameters. Water Resour Res 2012;48(8):W08521.. http://dx.doi.org/10.1029/2011WR011542.
[3] Brocca L, Melone F, Moramarco T, Wagner W, Naeimi V, Bartalis Z, et al. Improving runoff prediction through the assimilation of the ASCAT soil moisture product. Hydrol Earth Syst Sci 2010;7(4):4113–44. http://dx.doi.org/0.5194/hessd-7-4113-2010.
[4] Brocca L, Moramarco T, Melone F, Wagner W, Hasenauer S, Hahn S. Assimilation of surface- and root-zone ASCAT soil moisture products into rainfall-runoff modeling. IEEE Trans Geosci Remote Sens 2012;50(7):2542–55. http://dx.doi.org/10.1109/TGRS.2011.2177468.
[5] Chen F, Crow WT, Starks PJ, Moriasi DN. Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. Adv Water Resour 2011;34(4):526–36. http://dx.doi.org/10.1016/j.advwatres.2011.01.011.
[6] Chirico GB, Medina H, Romano N. Kalman filters for assimilating near-surface observations into the Richards equation-Part 1: retrieving state profiles with linear and nonlinear numerical schemes. Hydrol Earth Syst Sci 2014;18(7):2503–20. http://dx.doi.org/10.5194/hess-18-2503-2014.
[7] Clapp RB, Hornberger GM. Empirical equations for some soil hydraulic properties. Water Resour Res 1978;14(4):601–4. http://dx.doi.org/10.1029/WR014i004p00601.
[8] Crow WT. Correcting land surface model predictions for the impact of temporally sparse rainfall rate measurements using an ensemble Kalman filter and surface brightness temperature observations. J Hydrometeorol 2003;4(5):960–73. http://dx.doi.org/10.1175/1525-7541(2003)004<0960:CLSMPF>2.0.CO;2.
[9] Collow TW, Robock A, Wu W. Influences of soil moisture and vegetation on convective precipitation forecasts over the United States great plains. J Geophys Res 2014;119(15):9338–58. http://dx.doi.org/10.1002/2014JD021454.
[10] Dai Y, Zeng X, Dickinson RE, Baker I, Bonan GB, Bosilovich MG, et al. The common land model. Bull Amer Meteor Soc 2003;84:1013–23. http://dx.doi.org/10.1175/BAMS-84-8-1013.
[11] Dobson MC, Ulaby FT, Hallikainen MT, El-Rayes MA. Microwave dielectric behavior of wet soil-Part II: dielectric mixing models. IEEE Trans Geosci Remote Sens 1985(1):35–46. http://dx.doi.org/10.1109/TGRS.1985.289498.
[12] Drusch M. 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. J Geophys Res 2007;112(D3). http://dx.doi.org/10.1029/2006JD007478.
[13] Entekhabi D, Rodriguez-Iturbe I, Castelli F. Mutual interaction of soil moisture state and atmospheric processes. J Hydrol 1996;184(1):3–17. http://dx.doi.org/10.1016/0022-1694(95)02965-6.
[14] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res 1994;99(C5):10143–62. http://dx.doi.org/10.1029/94JC00572.
[15] Evensen G, Van Leeuwen PJ. Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. Mon Weather Rev 1996;124(1):85–96. http://dx.doi.org/10.1175/1520-0493(1996)124<0085:AOGADF>2.0.CO;2.
[16] Fujii H. Development of a microwave radiative transfer model for vegetated land surface based on comprehensive in-situ observations. PhD diss, Tokyo Univ.; 2005.
[17] Han X, Franssen HJH, Montzka C, Vereecken H. Soil moisture and soil properties estimation in the Community Land Model with synthetic brightness temperature observations. Water Resour Res 2014;50(7):6081–105. http://dx.doi.org/10.1002/2013WR014586.
[18] Han XJ, Li X, Franssen HH, Vereecken H, Montzka C. Spatial horizontal correlation characteristics in the land data assimilation of soil moisture. Hydrol Earth Syst Sci 2012;16(5):1349–63. http://dx.doi.org/10.5194/hess-16-1349-2012.
[19] Heathman GC, Starks PJ, Ahuja LR, Jackson TJ. Assimilation of surface soil moisture to estimate profile soil water content. J Hydrol 2003;279(1):1–17. http://dx.doi.org/10.1016/S0022-1694(03)00088-X.
[20] Heathman GC, Cosh MH, Han E, Jackson TJ, McKee L, McAfee S. Field scale spatiotemporal analysis of surface soil moisture for evaluating point-scale in situ networks. Geoderma 2012;170:195–205. http://dx.doi.org/10.1016/j.geoderma.2011.11.004.
[21] Houtekamer PL, Mitchell HL. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon Weather Rev 2001;129(1):123–37. http://dx.doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.
[22] Huang CL, Li X, Lu L, Gu J. Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter. Remote Sens Environ 2008;112(3):888–900. http://dx.doi.org/10.1016/j.rse.2007.06.026.
[23] Jackson TJ, Schmugge TJ. Vegetation effects on the microwave emission of soils. Remote Sens Environ 1991;36(3):203–12. http://dx.doi.org/10.1016/0034-4257(91)90057-D.
[24] Jia B, Xie Z, Tian X, Shi C. A soil moisture assimilation scheme based on the ensemble Kalman filter using microwave brightness temperature. Sci China Ser D-Earth Sci 2009;52(11):1835–48. http://dx.doi.org/10.1007/s11430-009-0122-z.
[25] Lei F, Huang C, Shen H, Li X. Improving the estimation of hydrological states in the SWAT model via the ensemble Kalman smoother: synthetic experiments for the Heihe River Basin in northwest China. Adv Water Resour 2014;67:32–45. http://dx.doi.org/10.1016/j.advwatres.2014.02.008.
[26] Li J, Islam S. Estimation of root zone soil moisture and surface fluxes partitioning using near surface soil moisture measurements. J Hydrol 2002;259(1):1–14. http://dx.doi.org/10.1016/S0022-1694(01)00589-3.
[27] Li X, Huang CL, Che T, Jin R, Wang SG, Wang JM, et al. Development of a Chinese land data assimilation system: its progress and prospects. Prog Nat Sci 2007;17(8):881–92. http://dx.doi.org/10.1080/10002007088537487.
[28] Liu Y, Weerts A, Clark M, Hendricks Franssen HJ, Kumar S, Moradkhani H, et al. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrol Earth Syst Sci 2012;16(10):3863–87. http://dx.doi.org/10.5194/hess-16-3863-2012.
[29] Lu H, Koike K, Yang K, Hu ZY, Xu XD, Rasmy M, et al. Improving land surface soil moisture and energy flux simulations over the Tibetan plateau by the assimilation of the microwave remote sensing data and the GCM output into a land surface model. Int J Appl Earth Obs 2012;17:43–54. http://dx.doi.org/10.1016/j.jag.2011.09.006.
[30] Lü H, Yu Z, Zhu Y, Drake S, Hao Z, Sudicky EA. Dual state-parameter estimation of root zone soil moisture by optimal parameter estimation and extended Kalman filter data assimilation. Adv Water Resour 2011;34(3):395–406. http://dx.doi.org/10.1016/j.advwatres.2010.12.005.
[31] Medina H, Romano N, Chirico GB. Kalman filters for assimilating near-surface observations in the Richards equation-Part 2: a dual filter approach for simultaneous retrieval of states and parameters. Hydrol Earth Syst Sci Discuss 2012;9(12):13329–72. http://dx.doi.org/10.5194/hess-18-2521-2014.
[32] Medina H, Romano N, Chirico GB. Kalman filters for assimilating near-surface observations in the Richards equation-Part 3: retrieving states and parameters from laboratory evaporation experiments. Hydrol Earth Syst Sci Discuss 2012;9(12):13373–414. http://dx.doi.org/10.5194/hess-18-2543-2014.
[33] Monsivais-Huertero A, Graham WD, Judge J, Agrawal D. Effect of simultaneous state–parameter estimation and forcing uncertainties on root-zone soil moisture for dynamic vegetation using EnKF. Adv Water Resour 2010;33(4):468–84. http://dx.doi.org/10.1016/j.advwatres.2010.01.011.
[34] Moradkhani H, Hsu KL, Gupta H, Sorooshian S. Uncertainty assessment of hydrologic model states and parameters: sequential data assimilation using the particle filter. Water Resour Res 2005;41(5):W05012. http://dx.doi.org/10.1029/2004WR003604.
[35] Moradkhani H, Sorooshian S, Gupta HV, Houser PR. Dual state–parameter estimation of hydrological models using ensemble Kalman filter. Adv Water Resour 2005;28(2):135–47. http://dx.doi.org/10.1016/j.advwatres.2004.09.002.
[36] Paloscia S, Pampaloni P. Microwave polarization index for monitoring vegetation growth. IEEE Trans Geosci Remot Sen 1988;26(5):617–21. http://dx.doi.org/10.1109/36.7687.
[37] Pauwels V, Hoeben R, Verhoest NE, De Troch FP. The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation. J Hydrol 2001;251(1):88–102. http://dx.doi.org/10.1016/S0022-1694(01)00440-1.
[38] Qin J, Liang S, Yang K, Kaihotsu I, Liu R, Koike T. Simultaneous estimation of both soil moisture and model parameters using particle filtering method through the assimilation of microwave signal. J Geophys Res 2009;114(D15):D15103. http://dx.doi.org/10.1029/2008JD011358.
[39] Reichle RH, Koster RD, Liu P, Mahanama SP, Njoku EG, Owe M. Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). J Geophys Res 2007;112(D9). http://dx.doi.org/10.1029/2006JD008033.
[40] Reichle RH, McLaughlin DB, Entekhabi D. Variational data assimilation of microwave radiobrightness observations for land surface hydrology applications. IEEE Trans Geosci Remot Sen 2001;39(8):1708–18. http://dx.doi.org/10.1109/10.1109/36.942549.
[41] Reichle RH, McLaughlin DB, Entekhabi D. Hydrologic data assimilation with the ensemble Kalman filter. Mon Weather Rev 2002;130(1):103–14 http://dx.doi.org/10.1175/1520-0493(2002)130≤0103:HDAWTE≥2.0.CO;2.
[42] Scharnagl B, Vrugt JA, Vereecken H, Herbst M. Inverse modelling of in situ soil water dynamics: Investigating the effect of different prior distributions of the soil hydraulic parameters. Hydrol Earth Syst Sci 2011;15(10):3043–59. http://dx.doi.org/10.1016/10.5194/hess-15-3043-2011.
[43] Taylor KE. Summarizing multiple aspects of model performance in a single diagram. J Geophys Res 2001;106(D7). http://dx.doi.org/7183-7192.10.1029/2000JD900719.
[44] Tian X, Xie Z, Dai A. A land surface soil moisture data assimilation system based on the dual-UKF method and the Community Land Model. J Geophys Res 2008;113:D14127. http://dx.doi.org/10.1029/2007JD009650.
[45] Tian X, Xie Z, Dai A, Shi C, Jia B, Chen F, et al. A dual-pass variational data assimilation framework for estimating soil moisture profiles from AMSR-E microwave brightness temperature. J Geophys Res 2009;114(D16). http://dx.doi.org/10.1029/2008JD011600.
[46] Tong M, Xue M. Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part I: sensitivity Analysis and Parameter Identifiability. Mon Weather Rev 2008;136(5):1630–48. http://dx.doi.org/10.1175/2007MWR2070.1.
[47] Tong M, Xue M. Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part II: Parameter Estimation Experiments. Mon Weather Rev 2008;136(5):1649–68. http://dx.doi.org/10.1175/2007MWR2071.1.
[48] Ueno G, Higuchi T, Kagimoto T, Hirose N. Maximum likelihood estimation of error covariances in ensemble-based filters and its application to a coupled atmosphere-ocean model. Q J R Meteorol Soc 2010;136(650):1316–43. http://dx.doi.org/10.1002/qj.654.

Please cite this article as: W. Chen et al., Comparison of ensemble-based state and parameter estimation methods for soil moisture data assimilation, Advances in Water Resources (2015), http://dx.doi.org/10.1016/j.advwatres.2015.08.003

[49] Ulaby FT, Moore RK, Fung AK. Microwave remote sensing active and passive-volume iii: from theory to applications. Norwood, Mass.: Artech House; 1986.
[50] Vrugt JA, Diks C, Gupta HV, Bouten W, Verstraten JM. Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. Water Resour Res 2005;41(1):W01017. http://dx.doi.org/10.1029/2004WR003059.
[51] Vrugt JA, ter Braak CJF, Diks CGH, Schoups G. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: theory, concepts and applications. Adv Water Resour 2013;51:457–78. http://dx.doi.org/10.1016/j.advwatres.2012.04.002.
[52] Walker JP, Houser PR. A methodology for initializing soil moisture in a global climate model: assimilation of near-surface soil moisture observations. J Geophys Res 2001;106(D11):11761–74. http://dx.doi.org/10.1029/2001JD900149 .
[53] Wu G, Zheng X, Wang L, Zhang S, Liang X, Li Y. A new structure for error covariance matrices and their adaptive estimation in EnKF assimilation. Quart J Roy Meteorol Soc 2013;139(672):795–804. http://dx.doi.org/10.1002/qj.2000.
[54] Wu W, Dickinson RE. Time scales of layered soil moisture memory in the context of land-atmosphere interaction. J Clim 2004;17(14):2752–64. http://dx.doi.org/10.1175/1520-0442(2004)017<2752:TSOLSM>2.0.CO;2.
[55] Xie X, Zhang D. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. Adv Water Resour 2010;33(6):678–90. http://dx.doi.org/10.1016/j.advwatres.2010.03.012.
[56] Yang K, Koike T, Kaihotsu I, Qin J. Validation of a dual-pass microwave land data assimilation system for estimating surface soil moisture in semiarid regions. J Hydrometeorol 2009;10(3):780–93. http://dx.doi.org/10.1175/2008JHM1065.1.

Please cite this article as: W. Chen et al., Comparison of ensemble-based state and parameter estimation methods for soil moisture data assimilation, Advances in Water Resources (2015), http://dx.doi.org/10.1016/j.advwatres.2015.08.003