Data assimilation of soil water flow via ensemble Kalman filter: infusing soil moisture data at different scales

Citation for published version:
Zhu , P, Shi, L, Zhu, Y, Zhang, Q, Huang, K & Williams, M 2017, 'Data assimilation of soil water flow via ensemble Kalman filter: infusing soil moisture data at different scales', *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2017.10.078

Digital Object Identifier (DOI):
10.1016/j.jhydrol.2017.10.078

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Hydrology

Data assimilation of soil water flow via ensemble Kalman filter: infusing soil moisture data at different scales

Penghui Zhu[a], Liangsheng Shi[a*], Yan Zhu[a], Qiuru Zhang[a], Kai Huang[b], Mathew Williams[c]

[a] State Key Laboratory of Water Resources and Hydropower Engineering Sciences, Wuhan University, Wuhan, Hubei 430072, China

[b] Guangxi Institute of Hydraulic Research, Nanning, Guangxi 530023, China

[c] School of GeoSciences and National Centre for Earth Observation, University of Edinburgh, Edinburgh EH9 3FF, United Kingdom

*Corresponding author, E-mail: liangshs@whu.edu.cn

Abstract

This paper assesses the value of multi-scale near-surface (0~5cm) soil moisture observations to improve state-only or state-parameter estimation based on the ensemble Kalman filter (EnKF). To the best of our knowledge, studies on assimilating multi-scale soil moisture data into a distributed hydrological model with a series of detailed vertical soil moisture profiles are rare. Our analysis factors include spatial measurement scales, soil spatial heterogeneity, multi-scale data with contrasting information and systematic measurement errors. Results show that coarse-scale soil moisture data are also very useful for identifying finer-scale parameters and states given biased initial parameter fields, but it becomes increasingly difficult to recover the finer-scale spatial heterogeneity of soil property as the observation grids become coarser. In state-only estimation, near-surface soil moisture data result in improvement for shallow soil moisture profiles and degradation for deeper soil moisture profiles, with stronger influences from finer-scale data. With the decrease of background spatial heterogeneity of soil property, the value of coarse-scale data increases notably. Soil moisture data at two scales with contrasting information are found to be both useful. By updating spatially correlated soil hydraulic parameters, deviated observations still contain considerably useful information for finer-scale state-parameter estimation. More importantly, by presenting a difference information assimilation method we successfully extract useful information from soil moisture data containing systematic measurement errors. The current study can be extended to consider more complex atmosphere input and topography, etc.

Key words: data assimilation, multi-scale soil moisture data, distributed hydrologic model

1. **Introduction**

Data assimilation (DA), as a tool to improve model parameters and state predictions using observational data, has been frequently applied to hydrological practices (Chen and Zhang, 2006; Clark et al., 2008; Houser et al., 1998; Liu et al., 2012; Moradkhani et al., 2005; Samuel et al., 2014; Shi et al., 2012; Weerts and El Serafy, 2006; Xu and Gómez-Hernández, 2016). Soil moisture is a key variable in the land surface system and also an important source of observable data in DA. There still exist several aspects that might increase the difficulty of using soil moisture data for the improvement of hydrological simulations. One is that different measurement techniques yield soil moisture data of different scales, resolutions and accuracies (Susha et al., 2014; Vereecken et al., 2008); the second is that soil moisture itself exhibits high spatial and temporal variability at a variety of scales (Crow and Wood, 1999; Famiglietti et al., 1999; Gaur and Mohanty, 2013; Hu et al., 1998; Korres et al., 2015); another is that the scale mismatch between monitoring and modeling often occurs (Blöschl and Sivapalan, 1995; Western and Blöschl, 1999). Based on the aforementioned reasons, it remains a challenge to assimilate soil moisture data from multiple scales into modeling results to optimize estimation efficiently and effectively.

In order to capture the spatiotemporal characteristics (correlation length, variability, mean value, etc.) of soil moisture, a variety of measurement techniques have been developed. Vereecken et al. (2008) classified soil moisture measurements into two main categories: contact-based and contact-free methods. The former requires direct contact with the soil (e.g. time domain reflectometry), and typically provides point-scale measurements with high temporal and spatial resolutions as well as field-scale spatiotemporal soil moisture dynamics. The latter mainly includes remote sensing methods and hydro-geophysical methods (e.g. ground penetrating radar), and is more suitable for large and medium-scale monitoring (Romano, 2014; Vereecken et al., 2008). A critical evaluation of almost all the classical and modern soil moisture measurement means was presented by Susha et al. (2014), which reconfirmed that "both classical and modern techniques exhibit uncertainty related to the accuracy, precision, coverage and volume of measurements". Many other remarkable reviews are also available for interested readers, among which Fang and Lakshmi (2014), Robinson et al. (2008) and Romano (2014) are highly recommended. The emergence of various soil moisture measurement techniques provides a good opportunity for hydrological data assimilation, but how to evaluate the value of soil moisture data from multiple scales is challenging work.

As the term "scale" appears on different occasions, it is imperative to present the general meaning of it. In hydrology, the term "scale" may be defined from three perspectives, i.e. process scale (or characteristic scale of a process), observation scale and modeling scale (Blöschl and Sivapalan, 1995). The process scale is the scale that natural phenomena exhibit, and for a stochastic process, it refers to the scale of natural variability, which can be quantified by the correlation length of a natural process or variable (Western and Blöschl, 1999). The correlation length can be represented by the "range", which is a key parameter of the variogram. The range is the maximum distance of correlation. First proposed by Blöschl and Sivapalan (1995) and later adopted and improved by Korres et al. (2015), Romano (2014) and Vereecken et al. (2008), the concept of observation and modeling scale consists of a triplet of "support", "spacing", and "extent", and applies to both spatial and temporal dimensions. Support refers to the integration volume or area (or time) of a single sample or model element, spacing to the distance (or time interval) between samples, and extent to the overall measurement or simulation domain. In this study, the scale of soil moisture observations is specified with the “support” component, which is related to the spatial resolution in the sensors’ terminology.

The horizontal supports of frequently used techniques are on the order of centimeters for the handheld probes (e.g. ECH2O and FDR), meters for the geophysical methods (e.g. GPR), decameters or hectometers for the air-borne sensors (e.g. SAR, synthetic aperture radar and PBMR, L band push broom microwave radiometer), and hectometers or kilometers for the space-borne sensors (e.g. SMOS) (Fang and Lakshmi, 2014; Korres et al., 2015; Koyama et al., 2009; Koyama et al., 2010; Vereecken et al., 2008). Multi-scale soil moisture data may contain useful information of the surface-subsurface hydrological system at different spatio-temporal levels, and methods that can assimilate multi-scale data as well as accessing the value of them are needed. Durand and Margulis (2007) assimilate synthetic 25 km passive microwave (PM) observations and synthetic 1 km near infrared (NIR) narrowband albedo observations into a land surface model with a resolution of 1 km based on the EnKF approach. Lievens et al. (2015) provide an algorithm that deals with the assimilation of 25 km SMOS soil moisture data into the Variable Infiltration Capacity (VIC) model with a resolution of 12.5 km. Montzka et al. (2012) give an overview of multivariate and multi-scale data assimilation in terrestrial systems and state that both the PF (Particle Filter) and the EnKF are useful algorithms that can infuse multi-scale data. They note that multi-scale data assimilation can be performed in two ways: to use the observation operator, or to rescale the observations to the model scale prior to assimilation.

Although methods already exist for the assimilation of multi-scale data, their applications in terrestrial systems are limited (Montzka et al., 2012). One reason is that there might exist a mismatch between the scale at which data are measured and the scale at which simulations are conducted.

Synthetic or real-world studies concerning the “scale-mismatch” problem in multi-scale data assimilation are highly required. Another reason is that soil moisture data measured at different scales may conflict with each other, and insights on how to deal with conflicting data are lacking (Montzka et al., 2012).A third reason is that soil moisture at different scales exhibit spatial and temporal variability which is affected by several factors, such as soil, land use, meteorology, topography, and measurement scale (De Lannoy et al., 2006; Korres et al., 2013; Korres et al., 2015; Western et al., 1998). For the top-layer soil moisture data set of the OPE^3 field in De Lannoy et al. (2006), the horizontal range of soil moisture increases in wetter periods, during which a vertical flux of precipitation exists. Korres et al. (2013) find the combined influences of soil property, precipitation, land use pattern, evapotranspiration and analysis scale on surface soil moisture patterns in the modeling study on an agricultural field. For the soil moisture data sets of Rs15mCatchCrop and Rs150mCatchCrop in Korres et al. (2015), the mean range value changes from 432m to 711m, indicating that the correlation length increases with the measurement support. The above factors that affect soil moisture variability will also affect data assimilation efficiency, and it is too complicated to comprehensively consider their influences.

In addition, biased data (data with systematic measurement errors) at a certain scale may impede the successful utilization of data at other scales and lead to deterioration of data assimilation. Existing studies with respect to bias estimation and correction in DA can be seen in Dee (2005), Pauwels et al. (2013) and Ridler et al. (2014), etc. These studies, although based on different assumptions, present very insightful and effective approaches that can be applied in DA. As there is certain limitations for different methods, how to eliminate the data biases in DA is still worth study.

This paper is an attempt to conduct state-only or dual state-parameter estimation in subsurface hydrology using multi-scale soil moisture observations. Under the ensemble Kalman filter (EnKF) framework, synthetic soil moisture observations from three support scales 600 m, 3000 m and 9000 m are assimilated into a fully coupled distributed unsaturated-saturated water flow model (Zhu et al., 2012) with a resolution of 600 m. We will investigate the influences of measurement scale (horizontal support), soil spatial heterogeneity (in terms of parameter correlation length), conflicting soil moisture data from two scales (caused by different precipitation/irrigation time series) and systematic measurement errors on retrieving soil moisture profiles and estimating saturated soil hydraulic conductivities.

2. Methodology

2.1. Fully coupled unsaturated-saturated water flow model

A fully coupled unsaturated-saturated water flow model developed by Zhu et al. (2012) is selected to simulate the soil water and groundwater flow. The validity and efficiency of the model have been demonstrated by comparing its simulation results with those of Hydrus1D, the Variably-Saturated Two-Dimensional Water Flow and Transport Model (SWMS2D), the 3D model HydroGeoSphere, and FEFLOW. Moreover, by applying to a practical irrigation district, the Yonglian Irrigation District, Inner Mongolia, China, the model reveals its applicability in simulating large-scale unsaturated-saturated water flow.

According to the experimental findings which demonstrate that the vertical fluxes are often dominant over the lateral fluxes in the unsaturated zone at the hillslope scale (Sherlock et al., 2002), it is usually considered reasonable in large-scale simulations to care only about the vertical flow and neglect the horizontal flux in the vadose zone (Chen et al., 1994). Therefore, the heavy computational burden of numerically modeling large-scale water flow can be reduced by simplifying the three-dimensional (3D) Richards' equation in the unsaturated zone to the 1D equation. In the model, the whole unsaturated-saturated domain is horizontally divided into several sub-areas according to the spatially distributed inputs such as soil type, vegetation, meteorological condition and topography. For each sub-area, a 1D vertical soil column is used to represent the averaged unsaturated flow in that area. It is also assumed that only vertical fluxes exist between the unsaturated zone and the saturated zone. Then, the 1D Richards' equation of each column is coupled with the 3D groundwater flow equation through the vertical flux from the unsaturated zone to the groundwater table.

The Richards' equation is used to describe the simplified vertical flow through the unsaturated zone (Vogel et al, 1996),

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left(K_h\left(\frac{\partial h}{\partial z} - 1\right)\right) - S \tag{1}$$

where θ is the volumetric water content; h is the pressure head; t is time; K_h is the unsaturated soil hydraulic conductivity, which varies with the pressure head; z is the vertical coordinate and S is the source/sink terms.

For the saturated zone, the 3D groundwater flow equation is applied,

$$\mu_1 \frac{\partial H}{\partial t} = \frac{\partial}{\partial x_j}\left(K_s \frac{\partial H}{\partial x_i}\right) - S \tag{2}$$

where μ_1 is the elastic storage coefficient; H is the total water head; t is time; x_i and x_j are the spatial coordinates (x_i, $x_j = x, y, z$); K_s is the saturated hydraulic conductivity. This 3D groundwater flow equation is simplified using the concept of Vertical/Horizontal Splitting (Lardner and Cekirge, 1988), and then solved using water balance analysis method.

The vertical flux between the unsaturated zone and the groundwater table is expressed by the head gradient between the adjacent nodes in the unsaturated and saturated zones. The head matrix of the unsaturated and saturated zones are put together to form the unified global matrix, some of whose elements should be revised according to the water-balance-based coupling between the two zones. After solving the global head matrix, soil moistures in the unsaturated zone are acquired using the famous van Genuchten model. Detailed descriptions of the model construction can be seen in Zhu et al. (2012).

We select this model because on one hand, it is a distributed subsurface flow model which is suitable for investigating the impacts of horizontal observation scales in data assimilation practices, and on the other hand, it can simultaneously give detailed vertical soil moisture profiles for different sub-areas.

2.2. Ensemble Kalman Filter (EnKF)

Data assimilation (DA) is the process that combines modelling results and observations to generate the optimal states. The traditional standard Kalman filter is a widely applied sequential data assimilation approach suitable for small and linear systems with Gaussian error statistics. When implementing DA for large and nonlinear problems, some variants of the standard Kalman filter are believed to be more capable. The ensemble Kalman filter (EnKF), first proposed by Evensen (1994), is a Monte Carlo variant of the standard Kalman filter, and has proved highly applicable in complicated nonlinear hydrological problems (Komma et al., 2008; Pathiraja et al., 2016; Reichle et al., 2002; Shi et al., 2015; Song et al., 2014; Xie and Zhang, 2010; Xu and Gómez-Hernández, 2016). Different from the standard Kalman filter's explicit computation of the prior covariance matrix, the EnKF uses an ensemble of model realizations to approximate the covariance of the state vector.

In this study, the profile pressure head and soil moisture of the 1D soil columns are to be calibrated. The augmented state vector S_k that will be updated at time step k is,

$$S_k = \left(m_k{}^T, h_k{}^T, \theta_k{}^T\right)^T \tag{3}$$

where m_k is the parameter vector, h_k and θ_k are the variable (pressure head and soil moisture) vectors. The dimension of the state vector is $N_s = N_m + N_h + N_\theta$, where N_m is the number of unknown parameters of all the soil columns; N_h or N_θ is the total number of one dimensional nodes of the soil columns. In this study, we choose the simultaneous updating of h_k and θ_k. The updated soil moisture θ_k is used for result analysis, and the updated pressure head h_k are inserted back in the flow model because the pressure head is selected as the main variable to be directly solved in the model.

Whenever the observations are available, the state vector of each ensemble member i should be updated via,

$$S_{k,i}^a = S_{k,i}^b + K_k\left(d_{obsk,i} - H_k S_{k,i}^b\right) \tag{4}$$

where $S_{k,i}^b$ and $S_{k,i}^a$ denote the state vectors before and after assimilation, respectively; H_k is the observation operator mapping the model states to the observation space $H_k S_{k,i}^b$, which in other words, is the observation prediction. Let d_{obsk} denote the observation with a dimension of N_d at time step k, then for each realization *i* the observation vector is,

$$d_{obsk,i} = d_{obsk} + \varepsilon_{k,i} \tag{5}$$

where $\varepsilon_{k,i}$ is the independent white noise of the observation, which varies among realizations (Burgers et al., 1998). Serving as a weighting factor between model predictions and observations, the Kalman gain K_k is calculated by,

$$K_k = C_k^b H_k{}^T\left(H_k C_k^b H_k{}^T + R_k\right)^{-1} \tag{6}$$

where R_k is the error covariance matrix of the observations at time step k; $\boldsymbol{C}_k^b$ is the prior error covariance matrix of the state vector and can be approximated by,

$$C_k^b \approx \frac{1}{N_e-1}\sum_{i=1}^{N_e}\left[\left(S_{k,i}^b - \overline{S_k^b}\right)\left(S_{k,i}^b - \overline{S_k^b}\right)^T\right] \tag{7}$$

$$\overline{S_k^b} \approx \frac{1}{N_e}\sum_{i=1}^{N_e} S_{k,i}^b \tag{8}$$

where N_e is the ensemble size; $\overline{S_k^b}$ is the ensemble mean of the state vector before assimilation.

2.3. Method of assimilating multi-scale soil moisture observations

Recalling section 2.2, it can be found that the observation operator H_k and the covariance matrix C_k always appear together as the product $C_k^b {H_k}^T$ or $H_k C_k^b {H_k}^T$ in the updating step. If the observational variables are just part of the state variables to be updated, H_k will be a $N_d \times N_s$ matrix with an element of 1 where there is an observation prediction and 0 where there isn't. Under this condition, $C_k^b {H_k}^T$ and $H_k C_k^b {H_k}^T$ can actually be obtained by directly selecting several lines from C_k instead of calculating the whole of it, therefore the computational burden can be greatly reduced (Chen and Zhang, 2006). However, in our study the multi-scale soil moistures are not the direct state variables to be solved in the governing equations of the model, thus the whole state error covariance matrix C_k is supposed to be calculated and additional handling of H_k, $H_k C_k^b {H_k}^T$, as well as $C_k^b {H_k}^T$ is needed. We avoid this by augmenting the state vector S_k with the multi-scale observation d_k, which can be constructed from the direct model state variables using a "sub-model". A sub-model herein refers to the method and process used before data assimilation to transform the direct model variables to the predicted measurements when the direct model variables are not observable. The augmented state vector will then become,

$$S_k = \left({m_k}^T, {h_k}^T, {\theta_k}^T, d_k^T\right)^T \tag{9}$$

where d_k is the constructed model prediction of the multi-scale observation d_{obsk}. Thus, the elements of H are still 1s and 0s, and the convenience as stated in Chen and Zhang (2006) is retained. Note that by using different sub-models, different indirect model predictions can be constructed according to their relationships with direct model predictions. Another advantage of the augmented form of S_k is that data from two or more scales and of different types can be assimilated simultaneously.

In our synthetic study, coarse-scale soil moisture data is constructed by aggregating several finer-scale soil moisture data. The Area-Weighted-Average method is adopted to generate the aggregated coarse-scale soil moisture with the following expression,

$$ASM = \frac{\sum_i^n A_i \theta_i}{\sum_i^n A_i} \tag{10}$$

where n is the number of model grids within a same parent coarse observation grid; A_i is the area of a finer grid, that is, the area of a horizontal sub-area of the modeling domain; θ_i is soil moisture of a finer grid; ASM is the aggregated coarse-scale soil moisture. Note that the construction of area-averaged soil moisture by Eq. (10) is only for generating synthetic observations (in the reference modeling) and observation predictions (in the uncertain modeling) to drive the data assimilation of soil water flow in our synthetic study, in other cases the weights of finer grids in the aggregation of model results to coarse-scale grids do not necessarily depend on the area of finer girds.

2.4. Method of treating biased data—difference information assimilation method

In order to deal with the possible systematic measurement errors, we present a very simple and easy to use method based on EnKF, which is termed as “difference information assimilation”. The term difference information means the difference between observations, whether temporally or spatially. In our present study, only the spatial difference is involved. Assume that at a certain time point, N observational grids are measured by the same sensor and that these N measurements correspond to the following truth vector:

$$d_t = (d_t^1,\ d_t^2,\ \ldots,\ d_t^N)^T \tag{11}$$

where the superscript 1, 2, and N denote different physical measurement locations. If the systematic measurement error δ is failed to be eliminated, the original observation vector can be expressed as:

$$d_{obs} = (d_{obs}^1,\ d_{obs}^2,\ \ldots,\ d_{obs}^N)^T = (d_t^1 + \delta + \varepsilon^1,\ d_t^2 + \delta + \varepsilon^2,\ \ldots,\ d_t^N + \delta + \varepsilon^N)^T \tag{12}$$

where ε^i (i =1, 2, …, N) is random error. If this d_{obs} is directly assimilated, severe damage may be caused. Therefor the observation vector is transformed to such a form:

$$\widetilde{d_{obs}} = (d_{obs}^1 - d_{obs}^2,\ d_{obs}^2 - d_{obs}^3,\ \ldots,\ d_{obs}^{N-1} - d_{obs}^N,\ d_{obs}^N - d_{obs}^1)^T \tag{13}$$

$$\widetilde{d_{obs}} = (d_t^1 - d_t^2 + \varepsilon^1 - \varepsilon^2, d_t^2 - d_t^3 + \varepsilon^2 - \varepsilon^3, \ldots, d_t^N - t_o^1 + \varepsilon^N - \varepsilon^1)^T \tag{14}$$

$$\widetilde{d_{obs}} = (\widetilde{d_{obs}^1},\ \widetilde{d_{obs}^2},\ \ldots,\ \widetilde{d_{obs}^N})^T \tag{15}$$

where $\widetilde{d_{obs}^i}$ is the *ith* reconstructed observational data, representing the information difference between d_{obs}^i and d_{obs}^{i+1}. Note that if the random error ε^i of the original observations obeys a normal distribution $N(0, \sigma^2)$, and different observations are independent, then the random error of the constructed observational data $\widetilde{d_{obs}^i}$ will also obey a normal distribution, but the variance will be $2\sigma^2$.

The difference information is assimilated into the physical model using the form of augmented state vector described in Section 2.3, and can be jointly assimilated with other observational data with different scales or types. More exactly, the observation differences will be included in d_k of formula (9), and the model states will be updated not based on the original observations but on the observation differences.

3. Numerical experiments

Synthetic experiments are designed to explore the value of multi-scale near-surface (0~5 cm) soil moisture observations in state-parameter estimation. A reference modeling or "true run" is performed firstly, in which parameters and state variables are seen as "true" values. The EnKF runs are then conducted for the same time period using wrong soil hydraulic parameters. Initial and boundary conditions of the EnKF runs and the "true run" are set to be identical, as this study only focuses on parameter errors. The EnKF runs assimilate soil moisture observations draw from the reference modeling to compensate for errors arising from wrong parameters. In addition, the open-loop run without assimilating any observational data and with just the same configurations as the EnKF runs is performed in ensemble mode for comparison.

3.1 Flow domain description and boundary conditions

A 9000 ×9000 ×3 m cuboid domain with a number of 225 600×600 m sub-areas is created. Soil materials in the vertical direction are set to be uniform for simplification, since only the horizontal scale is the target of our study. The 1D vertical soil columns are divided into 31 elements with 5 cm thickness for the top two elements near soil surface and 10 cm thickness for the rest. The soil type of all sub-areas in the reference modeling is selected as sandy loam from Carsel and Parrish (1988). Soil parameters are as follows: $\alpha = 7.5\ \mathrm{m}^{-1}, \mathrm{n} = 1.89, \theta_r = 0.065, \theta_s = 0.41$, except the saturated soil hydraulic conductivity K_s, which varies among sub-areas and will be specified in Section 3.3. Initial total heads of all the simulation domain are -210 cm. By setting a constant water table of 210 cm below soil surface, the number of unsaturated nodes will not change during the whole simulation period and no water flux exists between horizontal sub-areas. Fig. 1 shows the daily-averaged time series of precipitation and potential evaporation, in which (a) and (c) are used as the upper boundary conditions for all the cases unless otherwise specified. The unsaturated-saturated flow in this study is simulated with a time step of 0.01 days.

[Fig. 1]

3.2 Observations

In this synthetic study, near-surface soil moisture observations from three horizontal scales 600 m, 3000 m and 9000 m, denoted by θ_{600}, θ_{3000} and θ_{9000} are used. The 600 m soil moisture data can correspond to some sensing instrument with a footprint of intermediate scale, for example, the cosmic-ray soil moisture probe (Zreda et al., 2008; Zreda et al., 2012). The 9000 m soil moisture data can correspond to the SMAP mission (Das et al., 2011; Entekhabi et al., 2010). The 3000 m measurement scale may represent the future 3000 m soil moisture product from SMAP or other missions although not mature at present. The 600 m-scale near-surface soil moisture observations are drawn from the linear mean of the top two nodes of the vertical soil columns in the reference modeling, representing an observation depth of 5 cm. The 3000 m and 9000 m-scale near-surface soil moisture observations are generated by Equation (10). Soil moisture observations from the three scales are all assumed to be unbiased and only suffer a random measurement error of 0.04 m^3/m^3 unless otherwise stated. The generation of these soil moisture data is under simplified conditions, since in reality the sensing depths of instruments will change with soil moisture content and coarse-scale observations are not necessarily the area-average of finer-scale observations. This simplification will not affect the main purpose of our study.

3.3 Experimental setup and data assimilation scenarios

The parameter K_s (m/day) is taken as the unknown factor. There are in total 225 parameters to be estimated for the whole study area. It is assumed that the logarithmic hydraulic conductivity field $Y(x) = ln\,Ks(x)$ obeys a normal distribution and is second-order stationary with a two-dimensional covariance function defined by a separable exponential form:

$$C_Y(\boldsymbol{h}) = \sigma_Y^2 \exp\left(-\frac{|h_x|}{\lambda_\mathrm{x}} - \frac{|h_y|}{\lambda_\mathrm{y}}\right) = \sigma_Y^2 \exp\left(-\frac{|x_1-x_2|}{\lambda_\mathrm{x}} - \frac{|y_1-y_2|}{\lambda_\mathrm{y}}\right) \quad (16)$$

where (x_1, y_1) and (x_2, y_2) are the 2D coordinates, σ_Y^2 is the variance, λ_x and λ_y are the correlation lengths in x and y directions. The prior mean and variance of the logarithmic hydraulic conductivity field are selected to be 0.5 and 1. The correlation lengths λ_x and λ_y are specified in Table 1, considering different soil spatial heterogeneities. Initial realizations of the logarithmic hydraulic conductivity field are generated using the above statistics. The reference field is given by randomly selecting a realization from realizations generated using a mean value of -0.5 and the same variance and correlation lengths as the initial field of the EnKF system. The model structural errors are ignored in this study since the same model is applied in the reference modeling and the EnKF runs. An ensemble size of 200 is selected. The total simulation time is 80 days, and the assimilation frequency is once a day.

Concerning the measurement scale (horizontal support), soil spatial heterogeneity, conflicting soil moisture data from two scales and systematic measurement errors，four scenarios are considered.

Scenario 1

Under a given background condition (correlation length of the ln Ks filed is 9000 m), soil moisture observations from different scales are available, the data value of these soil moisture observations need to be accessed. Two sub-scenarios are analyzed, the first is updating state variables only, while the unknown parameters are not cared, the second is simultaneously updating unknown parameters and state variables.

Scenario 2

For a given soil moisture product, data assimilation efficiency under different background conditions needs to be accessed. In this study, the background soil heterogeneity, in terms of the spatial correlation length of the ln Ks filed is considered.

Scenario 3

Under a given background condition, multi-scale soil moisture observations with contrasting information are available, the assimilation results need to be compared. Finer-scale data assimilation can be driven by different coarse-scale observations, which may provide contrasting soil moisture information with completely different temporal trends. Intuitively, detailed spatial soil moisture features can be better captured by finer-scale soil moisture data. However, due to the commonly existing spatial heterogeneity of soil properties, precipitation or evapotranspiration, etc., soil moisture on a particular area may not be represented by the observation of a given scale. It is not clear that which data scale is optimal if the scale of study areas does not match with the observation scales.

Scenario 4

Under a given background condition, soil moisture observations from two scales are available, but one data has systematic errors, and the other is unbiased with only random errors.

In corresponding to the four scenarios, a series of experiments are conducted, the detailed specifications of which are listed in Table 1, and described in Section 4. It should be mentioned that the sources of uncertainty in a hydrological modeling are not limited to the soil hydraulic conductivity only, and other factors such as the meteorological input, land use type, topography and the van Genuchten parameters, etc. can also result in great uncertainty of the model states. As the main purpose of this study is to explore the idea and method of assimilating multi-scale soil moisture observations, we reasonably select the soil hydraulic conductivity as the only unknown factor to simplify the research. Sensitivity of the value of multi-scale soil moisture observations to different factors mentioned above will be the topic of a future study, using the idea and method presented in the current study.

[Table 1]

3.4 Performance assessment

To evaluate the data assimilation effectiveness, root mean square error (RMSE) relative to the "true run" are computed based on the ensemble mean values of the unknown parameters and state variables:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[E(x_i) - x_i^{true})]^2} \tag{17}$$

where N is the number of nodes or the number of unknown parameters, $E(x_i)$ is the ensemble mean value of the *ith* state variable or parameter, x_i^{true} is the synthetic "true value" in the reference modeling. Because the saturated soil moisture content θ_s is treated as a known and correct parameter, the focus should be on the unsaturated vertical nodes. Unless otherwise stated, N is $22 \times 225 = 4950$ when calculating the RMSE of profile soil moistures for the whole simulation domain, where 22 is the number of unsaturated vertical nodes in each sub-area and 225 is the number of sub-areas. N is 225 when calculating the RMSE of unknown parameters. If only shallow layer, say 0~50 cm soil moistures are cared, then N will be $7 \times 225 = 1575$. If the soil moisture RMSE of certain sub-areas is cared, then N will equal the total number of unsaturated or cared vertical nodes of these sub-areas.

4. Results and discussion

4.1 Dual state-parameter estimation using soil moisture observations from different scales

In Case 1~3, near-surface (0~5 cm) soil moisture observations from three footprints (600 m, 3000 m and 9000 m) are assimilated into the same uncertain modeling, respectively. Both the saturated soil hydraulic conductivities and state variables are updated. The reference ln Ks field, the ensemble mean ln Ks field of the initial realizations and the estimated ensemble mean ln Ks fields (at the end of the simulation period) are shown in Fig. 2. The 225 estimated ensemble mean values of ln Ks versus their reference values at day 1, 10, 50 and 80 are plotted in Fig. 3.

From Fig. 2 and 3 it can be seen that the initial ensemble mean ln Ks field does not show any spatial tendency, compared with the reference field. At the end of the simulation period, the ln Ks field confined by the 600 m-scale soil moisture data is almost the same as the reference field. Major features of the reference ln Ks field can also be captured by the 3000 m-scale data. When the observation scale rises to 9000 m, the capability to recover the 600 m-scale ln Ks field decreases dramatically (Fig. 2). Generally, the estimated ln Ks values using finer-scale soil moisture data approach the reference values more rapidly and accurately (Fig. 3). In Fig. 3, the ln Ks estimates from coarser-scale soil moisture data are more concentrated with respect to their reference counterparts, indicating that the estimated ln Ks spatial variance is underestimated by assimilating coarse-scale data, which can also be seen in Fig. 2. Note that the reference ln Ks filed in Case 1~3 has a spatial mean and a spatial standard deviation of -0.225 and 0.737. While the 9000 m-scale soil moisture data can drive the spatial mean of the ln Ks field close to that of the reference ln Ks field (from the initial value 0.504 to the final -0.147), it cannot recover the spatial variance (from the initial spatial standard deviation 0.004 to the final 0.078). In conclusion, it is difficult to use coarse-scale soil moisture data to capture the finer-scale spatial heterogeneity of soil property.

[Fig. 2]

[Fig. 3]

The temporal evolution of RMSEs of ln Ks and profile soil moisture for Case 1~3 and the open-loop run are plotted in Fig. 4. Note that at beginning profile soil moistures of all the cases are set to be the same as that of the reference modeling and during the early time period precipitation haven’t yet infiltrated into deeper soil. It’s easy to see that soil moisture observations from all the three scales have positive effects on reducing profile soil moisture RMSE, but with the increase of observation scale, the efficiency decreases obviously. Improvements for profile soil moisture are in accordance with improvements for parameters. Detailed soil moisture profiles of a representative sub-area (Sub-area 183) at the end of the simulation time are plotted in Fig. 5 (a), including the reference modeling, the open-loop run and the EnKF runs.

[Fig. 4]

[Fig. 5]

4.2 Soil moisture profile retrieval without updating unknown parameters using soil moisture observations from different scales

As stated by Moradkhani et al. (2005) and Xie and Zhang (2010), in many data assimilation practices only dynamic state variables are updated while parameters are not. In Case 4~6 of this study, the saturated soil hydraulic conductivity is not updated, and other settings are the same as those of Case 1~3. The 0~200 cm and 0~50 cm profile soil moisture RMSEs for Case 4~6 and the open-loop run are shown in Fig. 6 (a) and (b), respectively. As the initial profile soil moistures are set to be correct, with the infiltration of precipitation under wrong soil hydraulic conductivity fields, soil moisture RMSEs all gradually increase with time at the early stage. From Fig. 6 (a), at the early period (about 0~10 days) when precipitation has not yet infiltrate into deeper soil, near-surface soil moisture data from all the three scales are found to improve profile soil moisture estimation compared with the open-loop run, and finer-scale data is more efficient. However, the RMSE using θ_{600} grows larger than that using θ_{3000} after about 10 days, and then larger than that using θ_{9000} at day 16, and later it grows distinctly beyond the soil moisture RMSE of the open-loop run. The RMSE using θ_{3000} also grows larger than that of the open-loop run. In contrast, for Case 6 using θ_{9000}, there is always a slight drop of the RMSE from that of the open-loop run during the whole simulation period. The above results tell that wrong hydraulic conductivity can lead to spurious soil moisture correlations between surface and deep nodes of the soil profile, and therefore assimilating near-surface soil moisture data can actually worsen soil moisture estimation.

[Fig. 6]

From Fig. 6 (b), it can be seen that at most assimilation steps, near-surface soil moisture data can improve the 0~50 cm profile soil moisture, and generally improvement from finer-scale observations is larger, except during day 12~40, when soil moisture RMSE using θ_{600} exhibits a greater fluctuation. In the long run, for shallow-layer soil moisture estimation finer-scale data is more efficient. Detailed soil moisture profiles (at the end of the simulation period) of a representative soil column without updating the soil hydraulic conductivity fields is plotted in Fig. 5 (b), for comparison with Fig. 5 (a). Related studies can be found as for using surface soil moisture data to modify deeper soil moisture profiles (Chen et al., 2011; Lievens et al., 2015; Walker et al., 2001), among which Chen et al. (2011) reveal similar results with this Section.

Combining the results of Section 4.1 and 4.2, it can be concluded that finer-scale soil moisture data have greater influence on data assimilation, under the premise that the observation grid is not smaller than the modeling grid. It should be noted that the “greater influence” can be positive (Fig. 4), but it can also be negative (Fig. 6 (a)).

4.3 Data assimilation under different degrees of soil spatial heterogeneity in terms of soil hydraulic conductivity

In Case 7, Case 3 and Case 8, soil moisture data from the 9000 m-scale covering a number of 225 model grids is used. The difference of the three cases lies in the background parameter correlation length (see Table 1). We artificially select these three parameter correlation lengths to make the comparison more distinct. The RMSE evolutions for ln Ks fields and profile soil moisture are exhibited in Fig. 7. It can be seen that soil moisture data have no effect, slight positive effect, and obvious positive effect on parameter estimation under a parameter correlation length of 1800 m, 9000 m and 60000 m, respectively. The RMSE evolution of profile soil moistures is in accordance with that of ln Ks fields. The above results indicate that it’s hard to use the 9000-scale data to improve the 600 m-scale state and parameter estimation with a strong spatial heterogeneity of soil property. But when the spatial heterogeneity of soil property becomes weaker, the 9000-scale data can provide rather valuable information for even the much finer 600 m-scale model grids.

The information gain from the 3000 m-scale soil moisture observation in respect of the spatial heterogeneity of the parameter field is also tested (Cases 2, 9 and 10). Similar phenomenon is observed (results not shown), except that the 3000 m-scale data is also useful when the spatial correlation length of the ln Ks field is 1800 m. In conclusion, the value of coarse-scale soil moisture observations for finer-scale state-parameter estimation greatly depends on the degree of background soil spatial heterogeneity.

[Fig. 7]

4.4 Data assimilation using multi-scale soil moisture observations with contrasting temporal trends

In Cases 11~13, we mimic Scenario 3 in which the upper boundary of the simulation filed is controlled by two different precipitation/irrigation strategies, which is demonstrated in Fig. 8. Most of the sub-areas in Fig. 8 (shallow grey areas) still receive the precipitation series in Fig. 1 (a), while 20 sub-areas (dark grey areas) in the top left corner of the study domain receive a different precipitation series in Fig. 1 (b). A 3000 m-scale soil moisture observation covering the top left 25 model grids and a 9000 m-scale observation covering all the domain are given. The correlation length of the ln Ks field is 9000 m. Fig. 9 gives the temporal evolution of the 3000 m-scale and the 9000 m-scale soil moisture observations, as well as the near-surface soil moisture changes of Sub-areas 61~65 in the reference modeling. It is obvious that the trend of 9000 m-scale soil moisture observation is much more similar to those of Sub-areas 61~65, while the 3000 m-scale observation exhibits a totally different temporal trend. For Sub-areas 61~65, it is natural to question which one of the 9000 m-scale (Case 12) and the 3000 m-scale (Case 11) observations can provide better estimation results, and whether simultaneously assimilation of these two data set can yield further improvement (Case 13).

[Fig. 8]

[Fig. 9]

[Fig. 10]

The RMSE of ln Ks versus time as well as that of profile soil moisture for Sub-areas 61~65 are demonstrated in Fig. 10. Results show that during the early period (about 0~9 days) the RMSE of ln Ks conditioned on the 9000 m-scale soil moisture data drops faster than that conditioned on the 3000 m-scale data, probably due to the similar temporal trend of the 9000 m-scale data with those of Sub-areas 61~65. But in the long run, the temporally deviated 3000 m-scale data gives better estimation, which might be attributed to the smaller scale-mismatch compared with the 9000 m-scale data, and the horizontal correlation of the ln Ks field. Considering both a short and a relatively long assimilation period, the simultaneously assimilation of 3000 m- and the 9000 m- scale observations is advantageous, because the corresponding RMSE curve always keeps close to the better one of the other two curves by the separate assimilation. The result of profile soil moisture follow that of parameter estimation. In conclusion, the influences of both the scale-mismatch and the contrast of observable information should be considered when assimilating multi-scale soil moisture data.

In practice, the usefulness of soil moisture data from a certain scale depends on several factors, including the spatial heterogeneity of soil properties, the spatial variation of precipitation or evapotranspiration, the degree of scale-mismatch between observations and simulations, etc. To judge the data value of multi-scale soil moisture data with contrasting information, it is not enough to consider only one factor. Our results demonstrate that by updating spatially correlated soil hydraulic parameters, deviated observations still contain considerably useful information to identify finer-scale states and parameters. The limitation of this section is that the influencing factors mentioned previously are not thoroughly considered. Taking a more systematic analysis of the data value of multi-scale data with contrasting information in DA can be the subject of a separate study.

4.5 Data assimilation using soil moisture data with systematic measurement errors

In real-world problems, soil moisture observations are subjected to both random errors and systematic errors. Systematic errors of observations should be removed before data are used. However, sometimes elimination of systematic observation errors cannot be guaranteed because of the complex error components.

In a virtual experiment we assume that the 600 m-scale soil moisture observation suffer a systematic bias of 0.03 m^3/m^3 from the true value. The random error is still 0.04 m^3/m^3. This data is assimilated with model results to test the impact of systematic observation errors on dual state-parameter estimation through Case 14. In Case 15, the result of EnKF by simultaneously utilizing the unbiased 3000 m-scale soil moisture data and the 600 m-scale data with a bias of 0.03 m^3/m^3, is tested. Note that by applying the augmented form of the state vector stated in Section 2.3 (formula (9)), data from different sources and of different types can be assimilated simultaneously. In Case 16, the 600 m-scale biased soil moisture data are assimilated using the difference information method described is Section 2.4. Other settings of Case 14~16 are identical with those of Case 1~3.

[Fig. 11]

The RMSEs of ln Ks fields and profile soil moisture for Case 14~16 and Case 2 are plotted in Fig. 11. It can be seen that the direct assimilation of biased 600 m-scale soil moisture observation severely damages the estimate of ln Ks fields and profile soil moisture. Even when the unbiased 3000 m-scale data is integrated together, the assimilation result does not get better obviously, indicating the decisive effect of the biased 600 m-scale observation over the unbiased 3000 m-scale observation. The above results illustrate that directly assimilating soil moisture data with systematic measurement errors can not only lead to deterioration of data assimilation but also impede the successful utilization of data at other scales. By applying the difference information assimilation method (Case 16), the 600 m-scale biased-data results in great improvement of parameter and soil moisture estimation. The limitation of the method used here is that the systematic observational errors are assumed to be constant at different spatial locations. Another limitation is that for unbiased observational data with only random errors, part of the information content can be reduced by assimilating the observation difference instead of the original data. The difference information assimilation method can be classified as the bias-blind systems stated in Dee (2005), with the observational data reprocessed before assimilation. Bias-aware assimilation methods, on the other hand, is advantageous in that they can explicitly give online bias estimation (Pauwels et al., 2013; Ridler et al., 2014)and can also take into consideration the forecast biases, but they are also based on specific assumptions, for example, assumptions about the source and nature of the biases in the system (Dee, 2005). The forecast bias in this study caused by wrong initial model parameters are implicitly reduced by jointly update the unknown parameters with state variables. The observation bias is implicitly eliminated by assimilating the difference information instead of the original information. To explicitly estimate the forecast and observation biases falls outside the scope of this study.

5. Conclusions

In this paper we present a multi-scale data assimilation scheme based on the EnKF method and a distributed subsurface water flow model, focusing on unsaturated zone state-only or state-parameter estimation with near-surface (0~5 cm) soil moisture observations. The value of near-surface soil moisture data from three measurement scales, namely 600 m, 3000 m and 9000 m, on reducing the 600 m-scale model errors are accessed (Scenario 1). Using the 9000 m and the 3000 m-scale soil moisture observations, the influence of soil spatial heterogeneity in terms of saturated soil hydraulic conductivity on data assimilation efficiency is considered (Scenario 2). The results of assimilating 3000 m-scale and 9000 m-scale soil moisture data which exhibit obviously different temporal trends, are compared (Scenario 3). In addition, the severe damage of directly assimilating soil moisture data with systematic measurement errors is demonstrated and a difference information method based on the multi-scale EnKF scheme (Scenario 4) is proposed.

Results and conclusions are summarized as follows:

Coarse-scale soil moisture data also contain very useful information for finer-scale state and parameter estimation with biased initial ln Ks fields, but with the increasing of measurement scales, the data assimilation efficiency decreases a lot (RMSE of soil moisture increases from 0.002 using 600 m data to 0.012 using 9000 m data). From Case 1~3 (Section 4.1), it can be seen that the a soil moisture observation scale of 3000 m still brings great improvements to the 600 m-scale state-parameter estimation (RMSEs of ln Ks and soil moisture reduced to 0.373 and 0.007 from 1.035 and 0.014 of the open-loop run). The 9000 m-scale soil moisture data can drive the spatial mean of the ln Ks field to the reference field, but it cannot recover the spatial variability. Soil heterogeneity have great effects on the efficiency of data assimilation. When the correlation length of the ln Ks field increases from 1800 m to 9000 m and to 60000 m, notable improvement can be seen using the 9000 m-scale soil moisture data to estimate the 600 m-scale states and parameters.

In dual state-parameter estimation, the profile soil moisture estimation is in accordance with the estimation of the ln Ks field. Without updating the ln Ks field, assimilation of near-surface soil moisture data can lead to improvement for shallow soil moisture profiles and damage for deeper (>50cm in this study) soil moisture profiles, and the smaller the measurement scale is, the larger the influence will be, given that the measurement scale is not smaller than the model scale.

When data from different scales are available but with contrasting temporal trends, their influences on data assimilation are subtle, and factors should be considered simultaneously. In Section 4.4, compared with the 9000 m-scale soil moisture data, the 3000m-scale data exhibits a more different temporal trend with the soil moisture temporal evolution of study areas, but the letter still brings much greater improvements (RMSEs of ln Ks and soil moisture further reduced to 0.446 and 0.010 from 1.048 and 0.166) except during the early period (0~9days). Joint assimilation of multi-scale soil moisture data with contrasting information is found to be advantageous but need to be further investigated.

Given that the measurement scale is not smaller than the model scale, finer-scale data is more efficiency on driving data assimilation, but should be used with caution. The direct assimilation of the 600 m-scale soil moisture data with systematic measurement errors results in the deterioration of data assimilation and also causes the failure of assimilating unbiased 3000 m-scale soil moisture data. By applying a spatial difference information assimilation method, we successfully eliminate the disadvantageous effect of the biased 600 m-scale observational data and prove that the multi-scale EnKF data assimilation scheme is able to take full advantage of data, even with systematic measurement errors.

Based on the results of this study, the general conclusion is that the EnKF approach is proved to provide a promising framework to use multi-scale soil moisture data. The current study only covers a few aspects in DA with multi-scale data, and should extended to consider unbiased initial parameter ensemble, or/and other factors such as meteorological input, land use type, topography, etc.

Acknowledgement

This study was supported by the Excellent Young Scientist Foundation of the National Natural Science Foundation of China Grant 51522904 and the National Natural Science Foundation of China Grant 51479144. The fifth author was supported in part through Guangxi Agricultural Technology Transformation Fund Grant 14125004-4. MW was supported by the UK National Centre for Earth Observation.

References:

Blöschl, G. and Sivapalan, M., 1995. Scale issues in hydrological modelling: A review. Hydrological Processes, 9(3-4): 251-290.

Burgers, G., Jan Van Leeuwen, P. and Evensen, G., 1998. Analysis Scheme in the Ensemble Kalman Filter. Monthly Weather Review, 126(6): 1719-1724.

Chen, F., Crow, W.T., Starks, P.J. and Moriasi, D.N., 2011. Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. Advances in Water Resources, 34(4): 526-536.

Chen, Y. and Zhang, D., 2006. Data assimilation for transient flow in geologic formations via ensemble Kalman filter. Advances in Water Resources, 29(8): 1107-1122.

Chen, Z., Govindaraju, R.S. and Kavvas, M.L., 1994. Spatial averaging of unsaturated flow equations under infiltration conditions over areally heterogeneous fields. 1. Development of models. Water Resources Research, 30(2): 523-533.

Clark, M.P. et al., 2008. Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. Advances in Water Resources, 31(10): 1309-1324.

Crow, W.T. and Wood, E.F., 1999. Multi-scale dynamics of soil moisture variability observed during SGP'97. Geophysical Research Letters, 26(23): 3485–3488.

Das, N.N., Entekhabi, D. and Njoku, E.G., 2011. An algorithm for merging SMAP radiometer and radar data for high-resolution soil-moisture retrieval. IEEE Transactions on Geoscience and Remote Sensing, 49(5): 1504-1512.

De Lannoy, G.J.M., Verhoest, N.E.C., Houser, P.R., Gish, T.J. and Van Meirvenne, M., 2006.

Spatial and temporal characteristics of soil moisture in an intensively monitored agricultural field (OPE3). Journal of Hydrology, 331(3-4): 719-730.

Dee, D. P., 2005. Bias and data assimilation. Q.J.R. Meteorol. Soc., 131: 3323–3343. doi:10.1256/qj.05.137

Durand, M. and Margulis, S.A., 2007. Correcting first‐order errors in snow water equivalent estimates using a multifrequency, multiscale radiometric data assimilation scheme. Journal of Geophysical Research Atmospheres, 112(D13): 3710-3711.

Entekhabi, D. et al., 2010. The soil moisture active passive (SMAP) mission. Proceedings of the IEEE, 98(5): 704-716.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research Oceans, 99(C5): 10143-10162.

Famiglietti, J.S. et al., 1999. Ground-based investigation of soil moisture variability within remote sensing footprints during the Southern Great Plains 1997 (SGP97) Hydrology Experiment. Water Resources Research, 35(6): 1839-1851.

Fang, B. and Lakshmi, V., 2014. Soil moisture at watershed scale: Remote sensing techniques. Journal of Hydrology, 516(6): 258-272.

Gaur, N. and Mohanty, B.P., 2013. Evolution of physical controls for soil moisture in humid and subhumid watersheds. Water Resources Research, 49(3): 1244-1258.

Houser, P.R. et al., 1998. Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. Water Resources Research, 34(12): 3405-3420.

Hu, Z., Chen, Y. and Islam, S., 1998. Multiscaling properties of soil moisture images and decomposition of large- and small-scale features using wavelet transforms. International Journal of Remote Sensing, 19(13):2451-2467.

Komma, J., Blöschl, G. and Reszler, C., 2008. Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting. Journal of Hydrology, 357(3-4): 228-242.

Korres, W., Reichenau, T.G. and Schneider, K., 2013. Patterns and scaling properties of surface soil moisture in an agricultural landscape: An ecohydrological modeling study. Journal of Hydrology, 498: 89-102.

Korres, W. et al., 2015. Spatio-temporal soil moisture patterns – A meta-analysis using plot to catchment scale data. Journal of Hydrology, 520: 326-341.

Koyama, C.N., Korres, W., Fiener, P. and Schneider, K., 2009. High-resolution soil moisture estimation from ALOS PALSAR Fine Mode (Dual Polarization) data in agricultural areas. EGU General Assembly Conference Abstracts, 11: 4980.

Koyama, C.N., Korres, W., Fiener, P. and Schneider, K., 2010. Variability of surface soil moisture observed from multitemporal C-band synthetic aperture radar and field data. Vadose Zone Journal, 9(9): 1014--1024.

Lardner, R.W. and Cekirge, H.M., 1988. A new algorithm for three-dimensional tidal and storm surge computations. Applied mathematical modelling, 12(5): 471-481.

Lievens, H. et al., 2015. SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. Remote Sensing of Environment, 168: 146-162.

Liu, Y. et al., 2012. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrology and Earth System Sciences, 16(10): 3863-3887.

Montzka, C., Pauwels, V.R., Franssen, H.J., Han, X. and Vereecken, H., 2012. Multivariate and multiscale data assimilation in terrestrial systems: a review. Sensors, 12(12): 16291-16333.

Moradkhani, H., Sorooshian, S., Gupta, H.V. and Houser, P.R., 2005. Dual state–parameter estimation of hydrological models using ensemble Kalman filter. Advances in Water Resources, 28(2): 135-147.

Moradkhani, H., Hsu, K., Gupta, H. and Sorooshian, S., 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. Water Resources Research, 41(5):237-246.

Pathiraja, S., Marshall, L., Sharma, A. and Moradkhani, H., 2016. Hydrologic modeling in dynamic catchments: A data assimilation approach. Water Resources Research, 52(5): 3350-3372.

Pauwels, V. R. N., De Lannoy, G. J. M., Hendricks Franssen, H.-J., and Vereecken, H., 2013. Simultaneous estimation of model state variables and observation and forecast biases using a two-stage hybrid Kalman filter. Hydrol. Earth Syst. Sci., 17: 3499-3521.

Reichle, R.H., McLaughlin, D.B. and Entekhabi, D., 2002. Hydrologic Data Assimilation with the Ensemble Kalman Filter. Monthly Weather Review, 130(1): 103-114.

Ridler, M.-E., H. Madsen, S. Stisen, S. Bircher, and R. Fensholt., 2014. Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark. Water Resour. Res., 50: 8962–8981.

Robinson, D.A. et al., 2008. Soil Moisture Measurement for Ecological and Hydrological Watershed-Scale Observatories: A Review. Vadose Zone Journal, 7(1): 358.

Romano, N., 2014. Soil moisture at local scale: Measurements and simulations. Journal of Hydrology, 516: 6-20.

Samuel, J., Coulibaly, P., Dumedah, G. and Moradkhani, H., 2014. Assessing model state and forecasts variation in hydrologic data assimilation. Journal of Hydrology, 513: 127-141.

Sherlock, M.D., McDonnell, J.J., Curry, D.S. and Zumbuhl, A.T., 2002. Physical controls on septic leachate movement in the vadose zone at the hillslope scale, Putnam County, New York, USA. Hydrological Processes, 16(13): 2559-2575.

Shi, L., Zeng, L., Zhang, D., and Yang, J., 2012. Multiscale-finite-element-based ensemble Kalman filter for large-scale groundwater flow. Journal of hydrology, 468: 22-34.

Shi, L., Song, X., Tong, J., Zhu, Y. and Zhang, Q., 2015. Impacts of different types of measurements on estimating unsaturated flow parameters. Journal of Hydrology, 524: 549-561.

Song, X., Shi, L., Ye, M., Yang, J. and Navon, I.M., 2014. Numerical Comparison of Iterative Ensemble Kalman Filters for Unsaturated Flow Inverse Modeling. Vadose Zone Journal, 13(2): 1-12.

Susha Lekshmi, S.U., Singh, D. N., and Maryam Shojaei Baghini, 2014. A critical review of soil moisture measurement. Measurement, 54: 92-105.

Vereecken, H. et al., 2008. On the value of soil moisture measurements in vadose zone hydrology: A review. Water Resources Research, 44(4): 253-270.

Vogel, T., Huang, K., Zhang, R., van Genuchten, M.Th., 1996. The HYDRUS Code for Simulating One-dimensional Water Flow, Solute Transport, and Heat Movement in Variably-saturated Media, Version 5.0, Research Report No. 140. U.S. Salinity Laboratory Agricultural Research Service U.S. Department of Agriculture Riverside, California.

Walker, J.P., Willgoose, G.R. and Kalma, J.D., 2001. One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: a comparison of retrieval algorithms. Advances in Water Resources, 24(6): 631-650.

Weerts, A.H. and El Serafy, G.Y.H., 2006. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resources Research, 42(9): 123-154.

Western, A.W., Blöschl, G. and Grayson, R.B., 1998. Geostatistical characterisation of soil moisture patterns in the Tarrawarra catchment. Journal of Hydrology, 205(1–2): 20-37.

Western, A.W. and Blöschl, G., 1999. On the spatial scaling of soil moisture. Journal of Hydrology, 217(3-4): 203-224.

Xie, X. and Zhang, D., 2010. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. Advances in Water Resources, 33(6): 678-690.

Xu, T. and Gómez-Hernández, J.J., 2016. Joint identification of contaminant source location, initial release time and initial solute concentration in an aquifer via ensemble Kalman filtering. Water Resources Research, 52(8): 6587-6595.

Zhu, Y., Shi, L., Lin, L., Yang, J. and Ye, M., 2012. A fully coupled numerical modeling for regional unsaturated–saturated water flow. Journal of Hydrology, 475: 188-203.

Zreda, M., Desilets, D., Ferré, T.P.A. and Scott, R.L., 2008. Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. Geophysical research letters, 35(21), L21402. http://dx.doi.org/10.1029/2008GL035655.

Zreda, M. et al., 2012. COSMOS: the COsmic-ray Soil Moisture Observing System. Hydrology and Earth System Sciences, 16(11): 4079 - 4099.

Figure and Table Captions

Fig. 1. Time series of daily precipitation and potential evaporation, (a) and (c) are used in all the cases; (b) is only used in Case 11~13 for Scenario 3.

Fig. 2. Illustration of the ln Ks fields in Scenario 1 (Case 1~3): (a) reference filed; (b) ensemble mean of the initial ensemble members; (c) ~ (e) estimated ensemble mean ln Ks fields at the end of the simulation period by 600 m-, 3000 m- and 9000 m-scale soil moisture data, respectively.

Fig. 3. Estimated ensemble mean values of ln Ks by soil moisture data from 600 m-, 3000 m- and 9000 m- scales in Case 1~3 of Scenario 1. (a) ~ (d) Represent the results of day 1, 10, 50 and 80, respectively. Results corresponding to different observation scales are denoted by different data tags.

Fig. 4. Temporal evolution of RMSEs for the ln Ks fields and profile soil moistures in Scenarios 1 (Case 1~3 and the open-loop run). Different lines represent results by soil moisture data from 600 m, 3000 m, and 9000 m scales, respectively.

Fig. 5. Soil moisture profiles of the representative sub-area for the reference modeling, the open-loop run and the EnKF runs in Scenario 1 (Case 1~6 and the open-loop run): (a) parameters are updated; (b) parameters are not updated.

Fig. 6. Temporal evolution of RMSEs for profile soil moistures in Case 4~6: (a) the whole unsaturated zone; (b) 0~50 cm soil depth.

Fig. 7. Temporal evolution of RMSEs for the ln Ks fields and profile soil moisture in Scenario 2 (Case 3, 7 and 8), given different parameter correlation lengths λ.

Fig. 8. Illustration of the upper boundary conditions used in Case 11~13 of Scenario 3. Most of the sub-areas (shallow grey areas) still receive the precipitation series in Fig. 1 (a), while 20 sub-areas (dark grey areas) in the top left corner of the study domain receive a different precipitation series in Fig. 1 (b). The locations of Sub-area 61~65 are labeled with the red Arabic numerals.

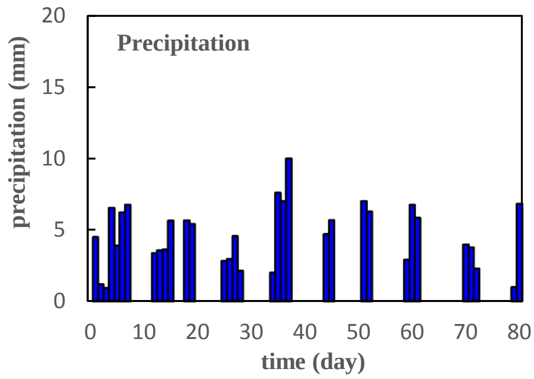
Fig. 9. Temporal trends of the 3000 m-scale (Case 11) and the 9000 m-scale (Case 12) soil moisture data in comparison with the average soil moisture changes of Sub-areas 61~65 in Scenario 3.

Fig. 10. Temporal evolution of RMSEs for the ln Ks fields and profile soil moistures of Sub-areas 61~65 in Scenarios 3 (Case 11~13 and the open-loop run). Different lines represent results by soil moisture data from a 3000 m- grid, a 9000 m- grid and the combined 3000 m- and 9000 m- grids, respectively.

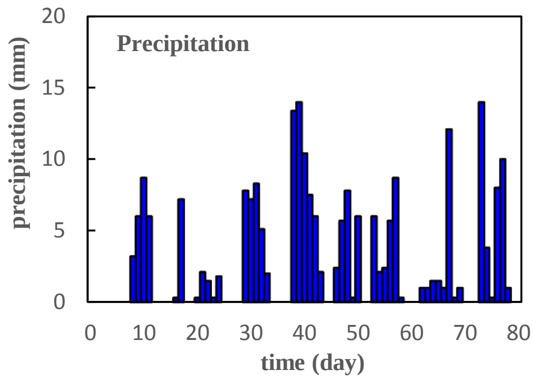
Fig .11. Temporal evolutions of RMSEs for the ln Ks fields and profile soil moistures of Case 2, 14, 15 and 16, as well as the open loop run in Scenario 4.

Table 1 Specifications of all the cases

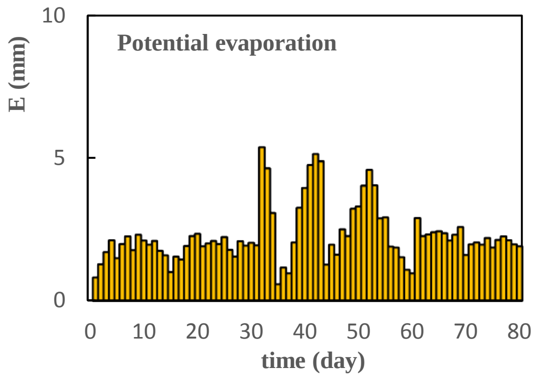
Fig. 1.



(a)



(b)



(c)

Fig. 1. Time series of daily precipitation and potential evaporation, (a) and (c) are used in all the cases; (b) is only used in Case 11~13 for Scenario 3.

Fig. 2.

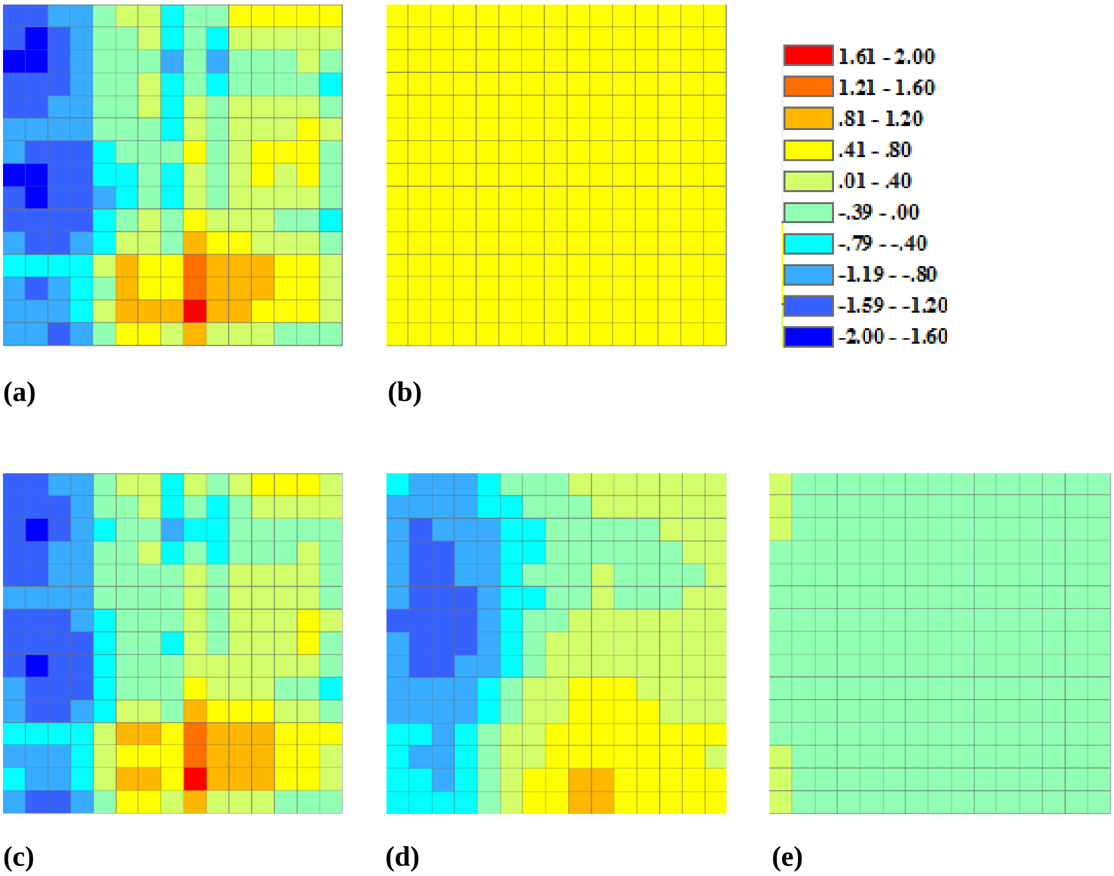
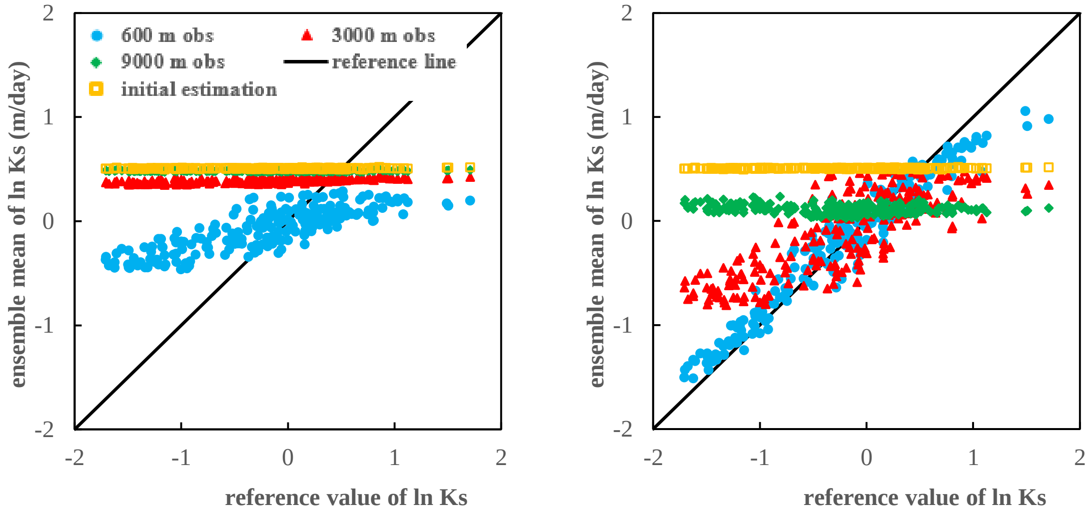
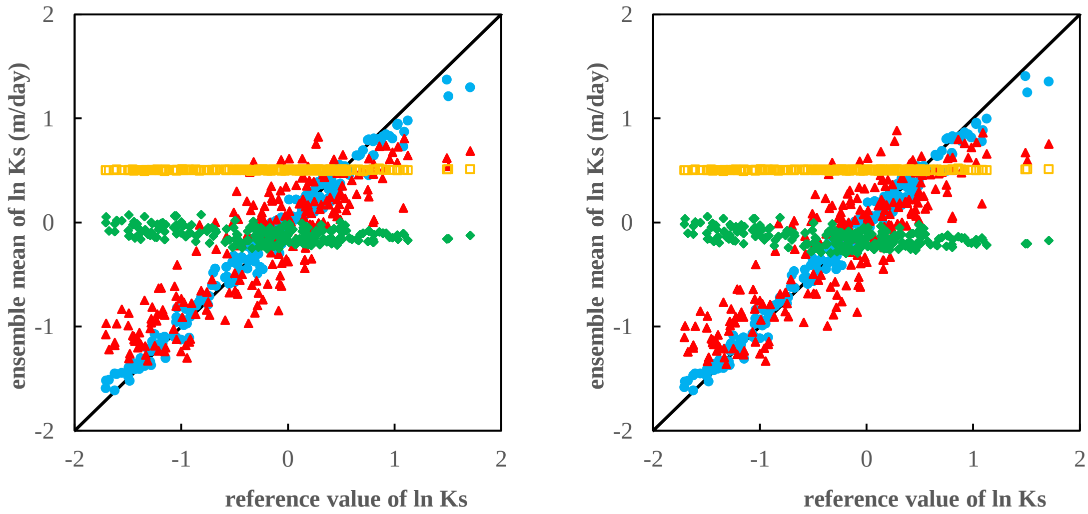


Fig. 2. Illustration of the ln Ks fields in Scenario 1 (Case 1~3): (a) reference filed; (b) ensemble mean of the initial ensemble members; (c) ~ (e) estimated ensemble mean ln Ks fields at the end of the simulation period by 600 m-, 3000 m- and 9000 m-scale soil moisture data, respectively.

Fig. 3.



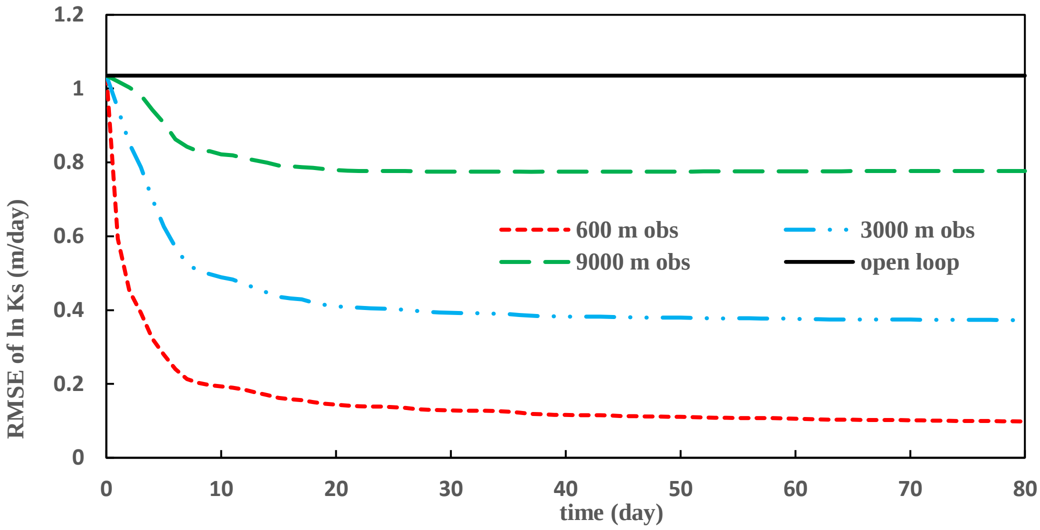
(a) **(b)**



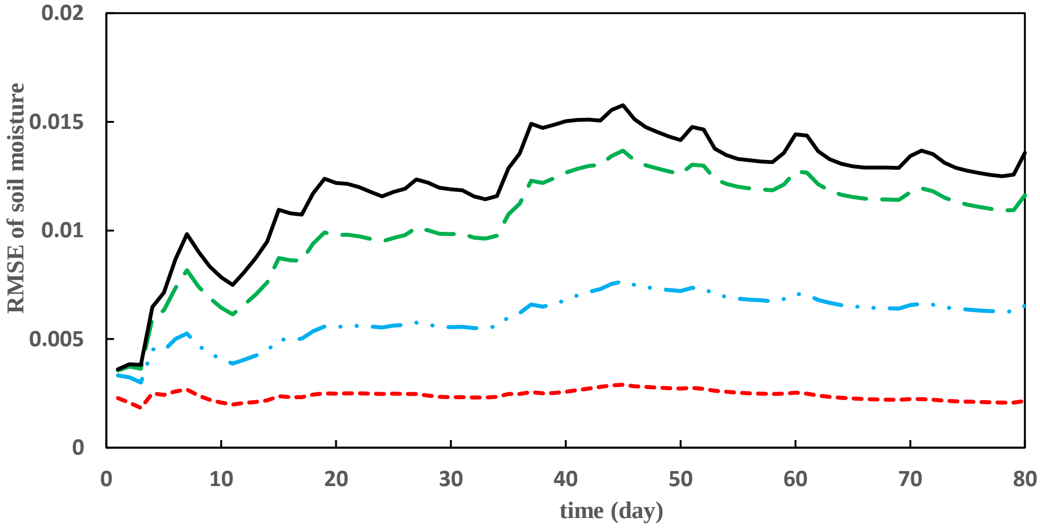
(c) **(d)**

Fig. 3. Estimated ensemble mean values of ln Ks by soil moisture data from 600 m-, 3000 m- and 9000 m- scales in Case 1~3 of Scenario 1. (a) ~ (d) Represent the results of day 1, 10, 50 and 80, respectively. Results corresponding to different observation scales are denoted by different data tags.

Fig. 4.



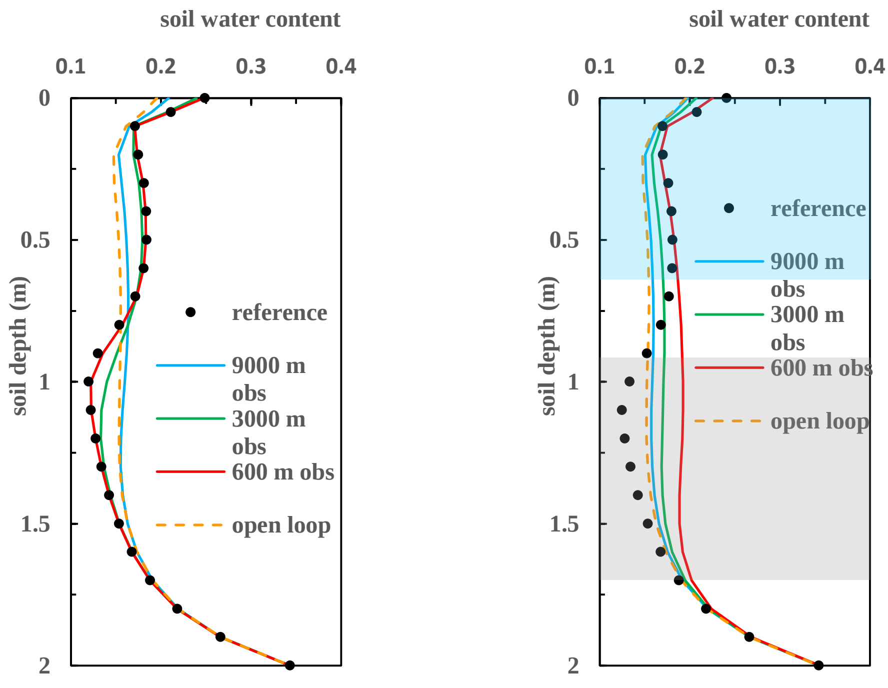
(a)



(b)

Fig. 4. Temporal evolution of RMSEs for the ln Ks fields and profile soil moistures in Scenarios 1 (Case 1~3 and the open-loop run). Different lines represent results by soil moisture data from 600 m, 3000 m, and 9000 m scales, respectively.

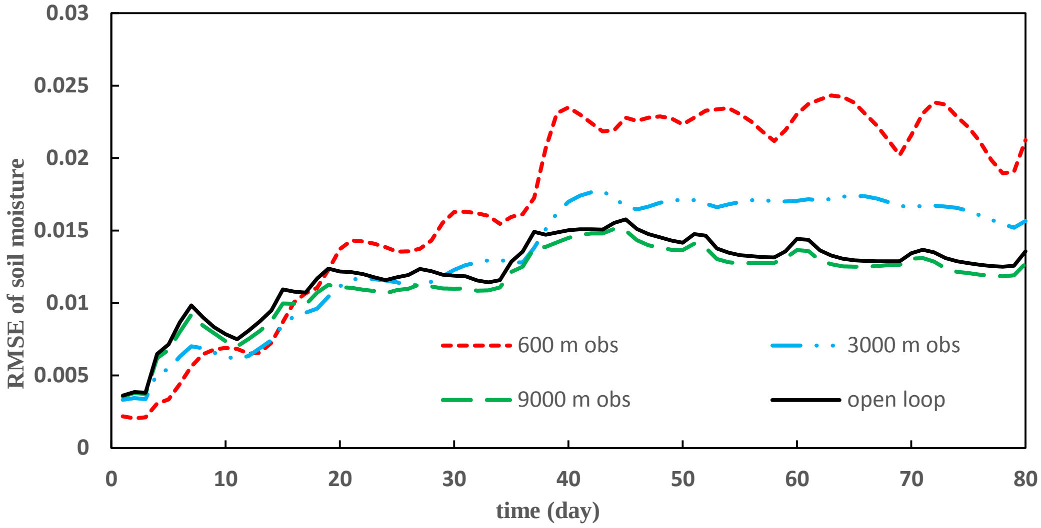
Fig. 5.



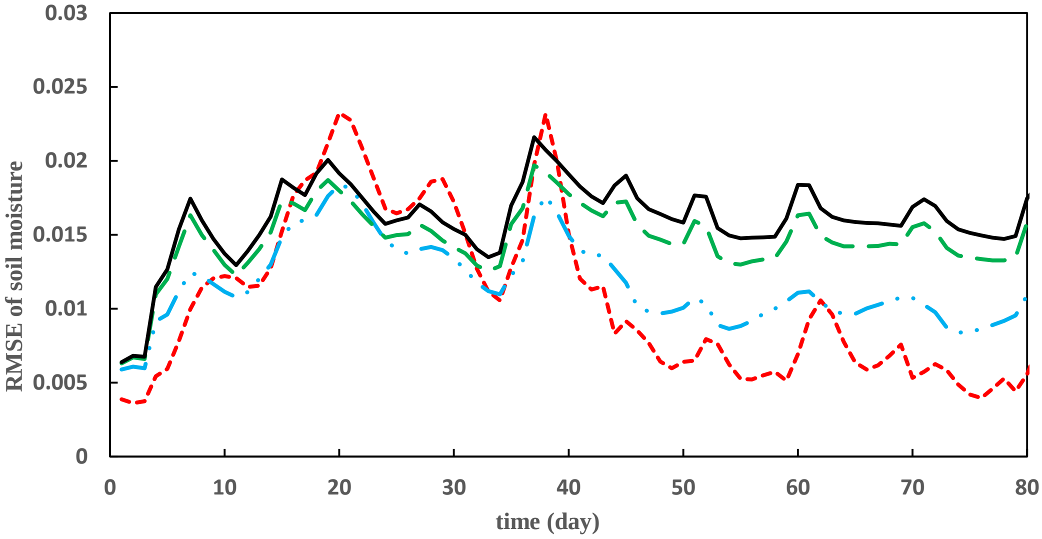
(a) (b)

Fig. 5. Soil moisture profiles of the representative sub-area for the reference modeling, the open-loop run and the EnKF runs in Scenario 1 (Case 1~6 and the open-loop run): (a) parameters are updated; (b) parameters are not updated.

Fig. 6.



(a)



(b)

Fig. 6. Temporal evolution of RMSEs for profile soil moistures in Case 4~6: (a) the whole unsaturated zone; (b) 0~50 cm soil depth.

Fig. 7.

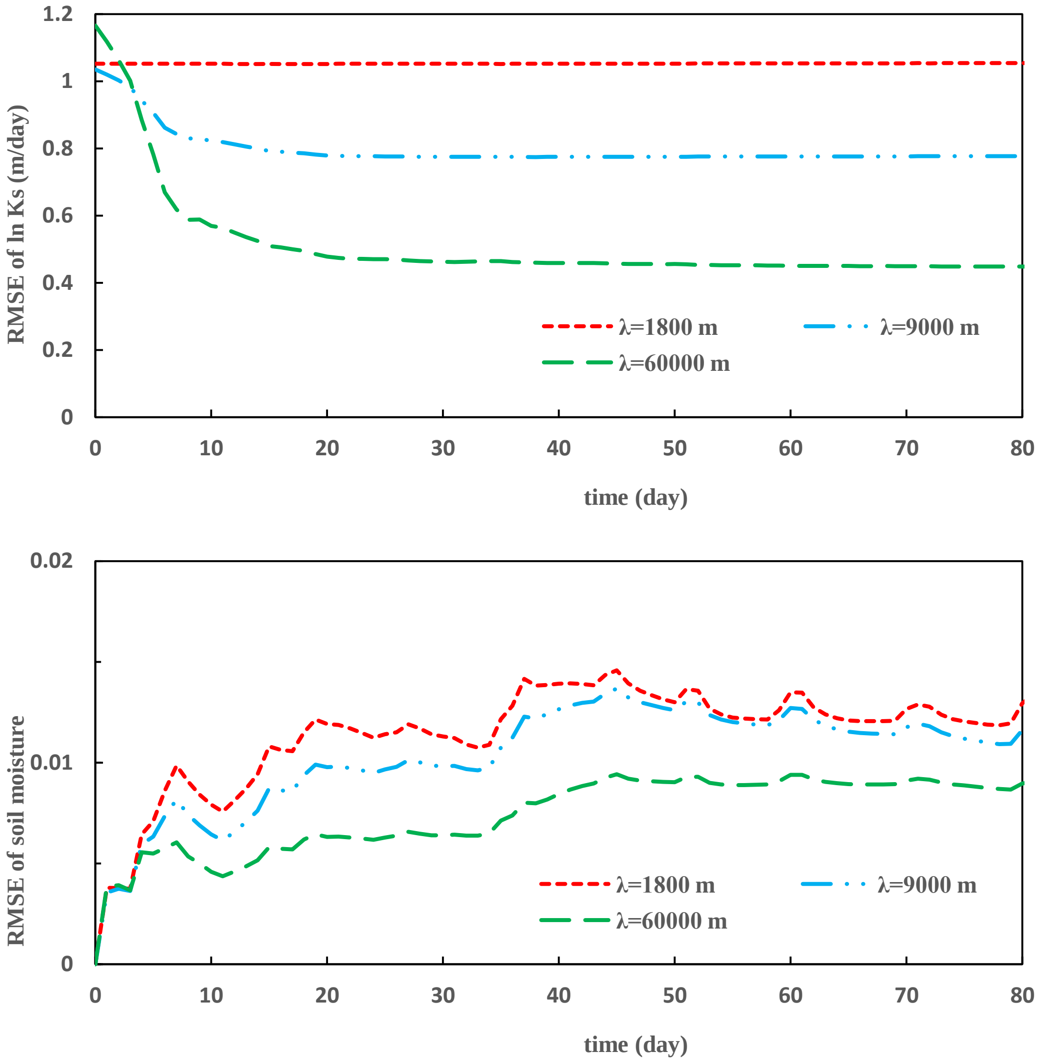


Fig. 7. Temporal evolution of RMSEs for the ln Ks fields and profile soil moisture in Scenario 2 (Case 3, 7 and 8), given different parameter correlation lengths λ.

Fig. 8

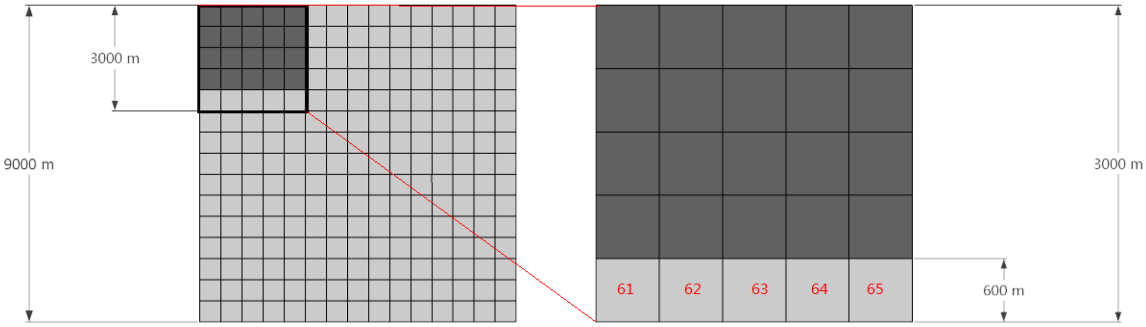


Fig. 8. Illustration of the upper boundary conditions used in Case11~13 of Scenario 3. Most of the sub-areas (shallow grey areas) still receive the precipitation series in Fig. 1 (a), while 20 sub-areas (dark grey areas) in the top left corner of the study domain receive a different precipitation series in Fig. 1 (b). The locations of Sub-area 61~65 are labeled with the red Arabic numerals.

Fig. 9.

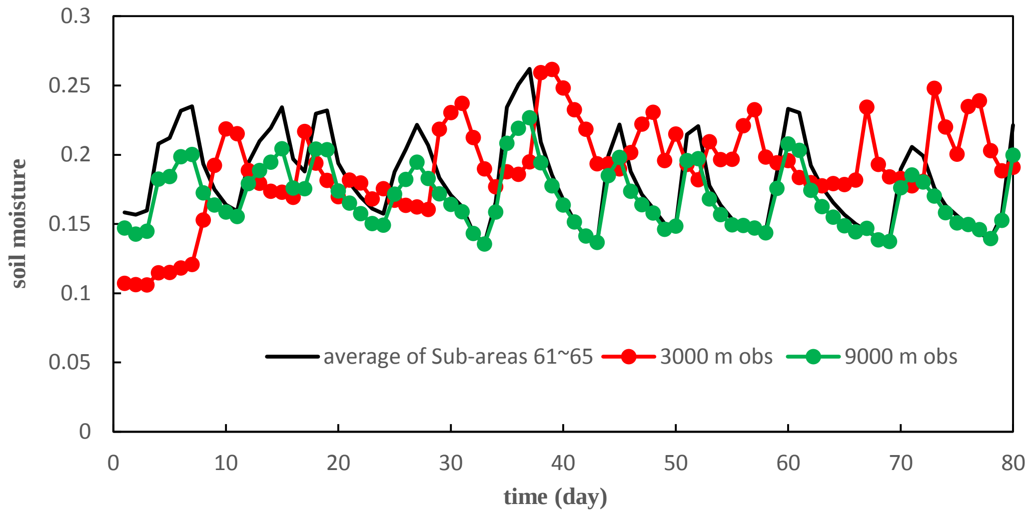


Fig. 9. Temporal trends of the 3000 m-scale (Case 11) and the 9000 m-scale (Case 12) soil moisture data in comparison with the average soil moisture changes of Sub-areas 61~65 in Scenario 3.

Fig.10

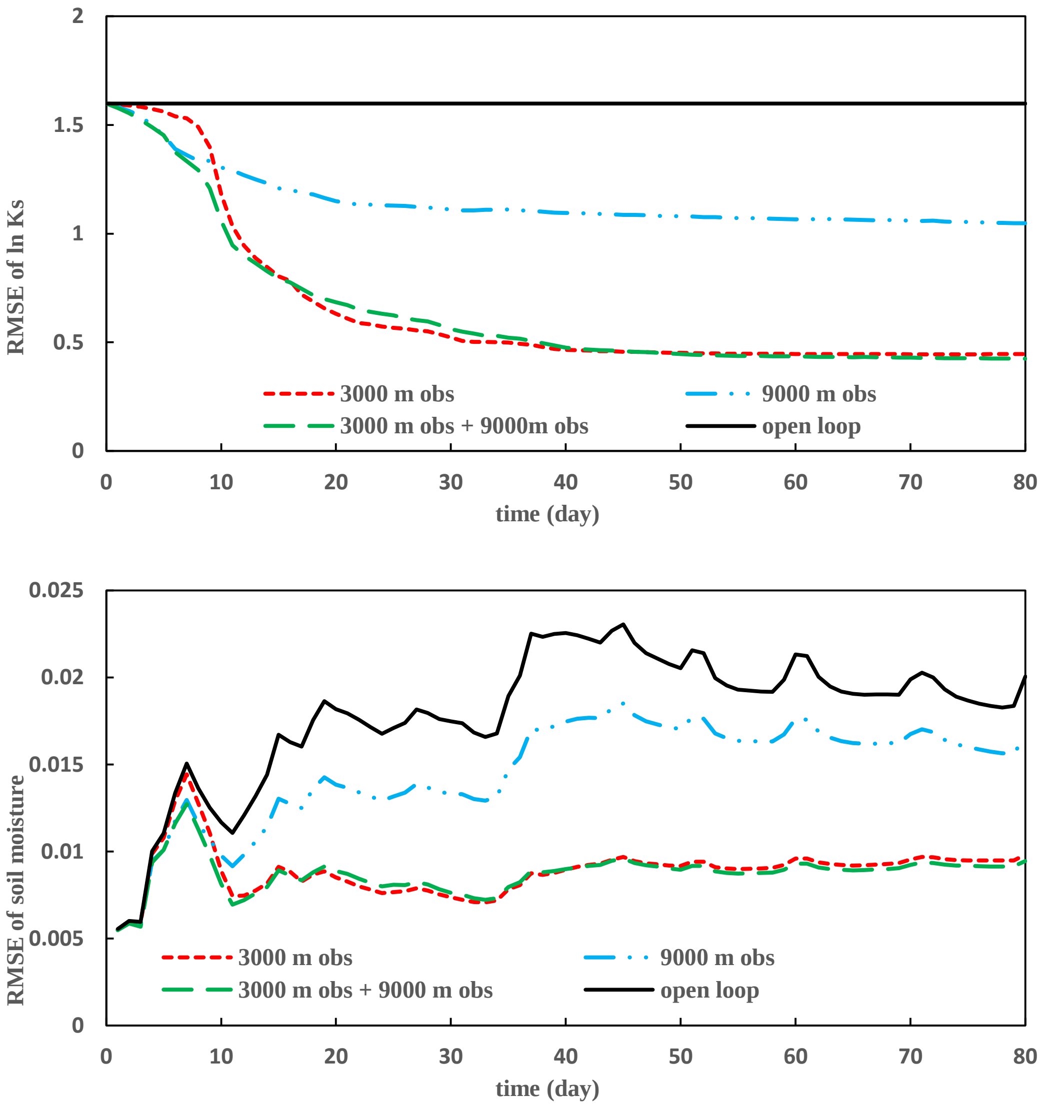


Fig. 10. Temporal evolution of RMSEs for the ln Ks fields and profile soil moistures of Sub-areas 61~65 in Scenarios 3 (Case 11~13 and the open-loop run). Different lines represent results by soil moisture data from a 3000 m- grid, a 9000 m- grid and the combined 3000 m- and 9000 m- grids, respectively.

Fig. 11.

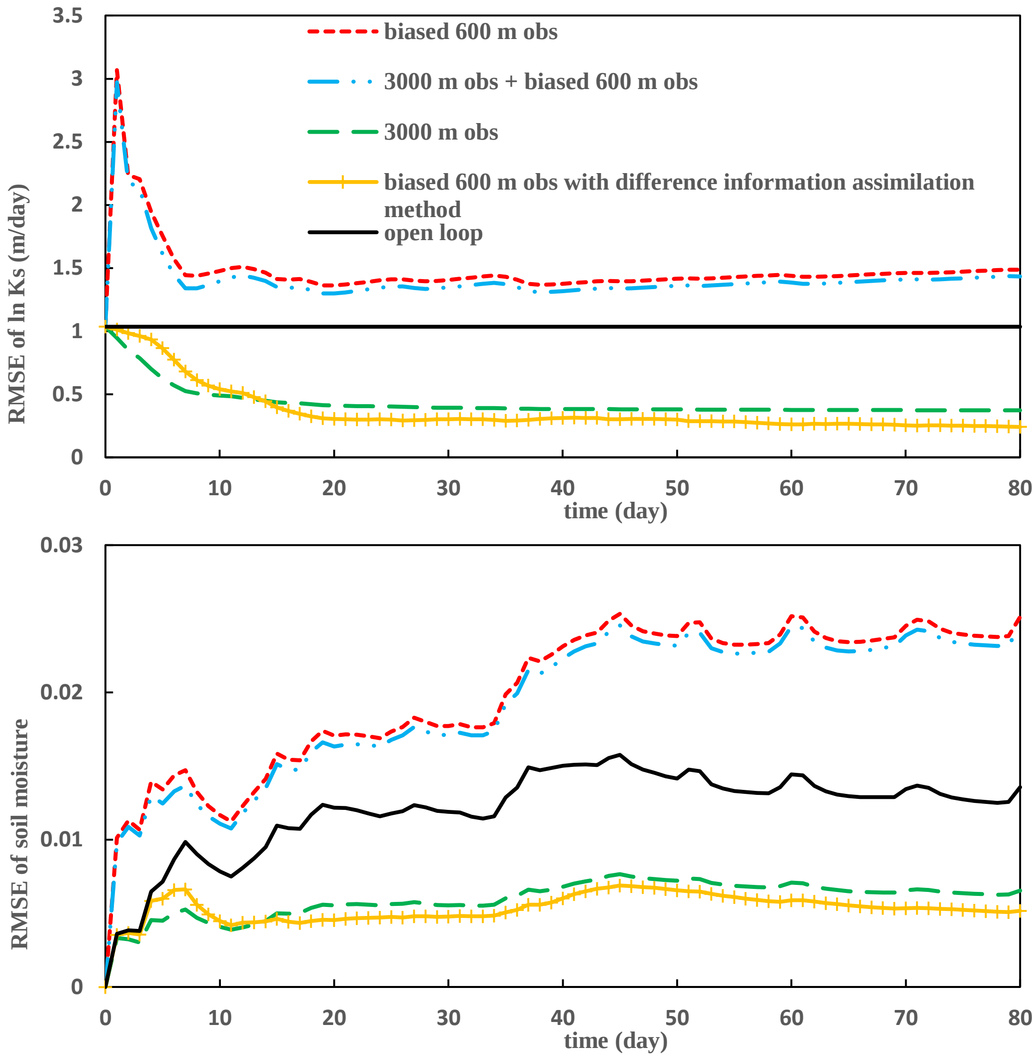


Fig .11. Temporal evolutions of RMSEs for the ln Ks fields and profile soil moistures of Case 2, 14, 15 and 16, as well as the open loop run in Scenario 4.

Table 1

Specifications of all the cases

Scenario	Case	Observation scale (m)	Correlation length of lnKs field (m)	Observation coverage	Parameter update	Systematic measurement error
1	1	600	9000	whole domain	Y	0
	2	3000	9000	whole domain	Y	0
	3	9000	9000	whole domain	Y	0
	4	600	9000	whole domain	N	0
	5	3000	9000	whole domain	N	0
	6	9000	9000	whole domain	N	0
2	3	9000	9000	whole domain	Y	0
	7	9000	1800	whole domain	Y	0
	8	9000	60000	whole domain	Y	0
	2	3000	9000	whole domain	Y	0
	9	3000	1800	whole domain	Y	0
	10	3000	60000	whole domain	Y	0
3	11	3000	9000	25 sub-areas	Y	0
	12	9000	9000	whole domain	Y	0
	13	3000、9000	9000	whole domain	Y	0
4	2	3000	9000	whole domain	Y	0
	14	600	9000	whole domain	Y	0.03
	15	600、3000	9000	whole domain	Y	0.03、0
	16	600	9000	whole domain	Y	0.03 (new method)