Comments by the Associate Editor:

Editor Decision

I received comments from three reviewers, two suggested major revision, and one (reviewer 2) suggested minor revision. All the three reviewers confirmed the contribution of the manuscript, but also pinpointed the problems of the manuscript. After received the comments, I carefully read the manuscript again and concur with the reviewers. Therefore, I suggested "reconsider after major revision". Please provide detailed replies to the comments made by the reviewers and revised your manuscript accordingly.

[Response]

Thanks for handling our manuscript. We have improved our manuscript according to the referees' suggestions.

Reply to comments from Anonymous Referee #1

General Comments

The manuscript investigates the effect of the uncertainty of the initial conditions in the context of soil water movement described by the Richards equation. First the necessary warm-up times for different soils and climates are determined and then the effects of different methods to describe the initial condition on a subsequent data assimilation are compared. The comparison is additionally shown on a real-world case. I think the manuscript is interesting and shows the effects of the uncertain initial conditions nicely. I have few comments that may require some additional investigation or discussion. However, the manuscript is sometimes difficult to read and could be clearer. Therefore, many comments ask for some clarification.

[Response]

Thanks for your positive comments. We have improved the manuscript according to your suggestions.

Major comments:

1. The required computational power varies between the different initial conditions. The most expensive ones (with warm-up period) seem to give the best results in the subsequent data assimilation. I would find it very interesting whether this finding holds if for each method a similar total computation time (computation time for initial condition + computation time for data assimilation) is available. This means that e.g. IC-ObsInt or IC-Flux could afford more ensemble members than IC-WUP. The question is then, if e.g a higher number of ensemble members (in combination with a larger uncertainty in the representation of the initial condition) of IC-ObsInt or IC-Flux could lead to similar, or even better results.

[Response]

Thank you for your valuable suggestion. First of all, we must apologize that we have used mistaken model input folders when analyzing the data from WUE and WUP, and two identical curves for WUE and WUP were generated. Thanks for raising this question, leading to the discovery of this mistake. Nevertheless, the general conclusions from this figure still hold. Several minor modifications are given below (please see our response to the last comment of yours) and will be reflected in the revised manuscript. Then, in order to investigate whether the conclusion for each method holds with a similar computational time and with a large uncertainty of initial state, we have added four cases for lnK_s estimations based on the initial condition IC-HfSatu and IC-ObsInt. In the first two cases, we increase the ensemble size of IC-HfSatu and IC-ObsInt to 500 (hereafter referred to as IC-HfSatu-500 and IC-ObsInt-500 in the manuscript) to explore the influences of ensemble size. The computational costs of them are similar to those of IC-WUP, IC-WUE and IC-Flux. Next, to further explore the effects of the uncertainty of the initial ensemble on the parameter estimations, we add a Gaussian noise (with a standard deviation of 0.017) to both IC-HfSatu-500 and IC-ObsInt-500 (hereafter referred to as IC-HfSatu-500-Un and IC-ObsInt-500-Un). The standard deviation of the Gaussian noise is calculated based on the spreading of IC-WUP initial ensemble, so that we can make sure the initial uncertainties of IC-HfSatu-500-Un and IC-ObsInt-500-Un are similar with IC-WUP.

According to the results of the four cases, we added a new figure (Fig. 7 in the revised manuscript) and corresponding discussion in the manuscript on the effects of ensemble size and the uncertainty of initial state on parameter estimation.

[Changes in the manuscript]

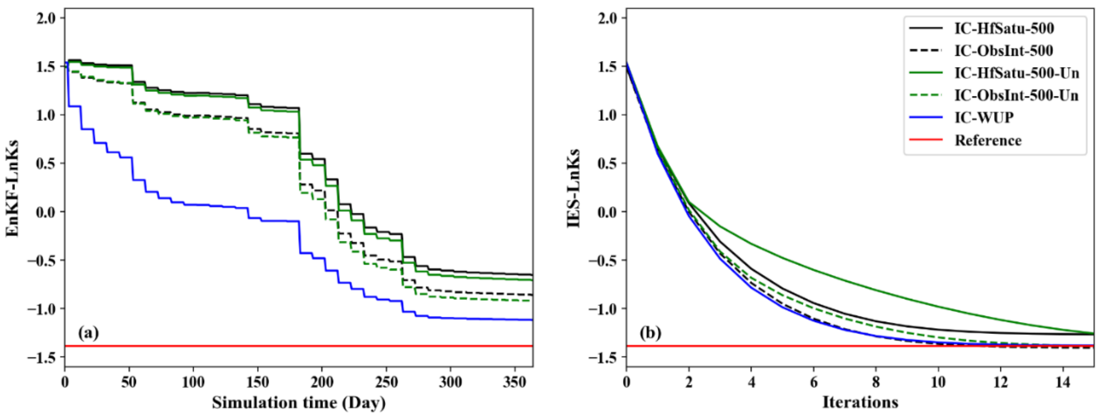


Fig. 7. (revised manuscript) The impacts of ensemble size and the uncertainty of initial state

on the results of $\ln K_s$ estimations using EnKF and IES.

The results of IC-HfSatu-500 and IC-ObsInt-500 with the ensemble size of 500 in Fig. 7 are similar with those of IC-HfSatu and IC-ObsInt (Fig. 6), indicating that the improvement of the parameter estimation result is slight when the ensemble size increases from 300 to 500. Hence, the ensemble size of 300 is sufficient for data assimilation problem in this study. In contrast, the influences of adding the uncertainty to the initial state on parameter estimation are totally different for EnKF and IES. Compared with the results of IC-ObsInt-500 and IC-HfSatu-500, the results of $\ln K_s$ estimation with IC-ObsInt-500-Un and IC-HfSatu-500-Un improve for EnKF (Fig. 7(a)), but deteriorate for IES (Fig. 7(b)). This may attribute to the diversity between two algorithms. EnKF is a sequential algorithm, so the state uncertainty introduced by UIC could decrease over assimilation steps. A larger ensemble state variance implemented at the beginning leads to a larger trust on data and thus a quicker update of parameter to truth, and can prevent EnKF from inbreeding, leading to a better result than that with initial condition of small variance. On the contrary, IES is a batch optimization method. The uncertainty of initial state exists at each iteration and has a negative effect on the model calibration during the whole simulation, worsening the parameter estimation results.

Moreover, the parameter estimation results of IC-WUP are still superior to those of IC-HfSatu-500-Un and IC-ObsInt-500-Un even they all have a similar computational cost, showing the promising performance of warm-up methods. The results are reasonable since all ensemble Kalman filter methods are affected by the quality of the auto-covariance matrix and the mean value of predicted state ensemble (Eqs. (12) and (13) for EnKF; Eqs. (15) and (16) for IES). For WUP method, the initial condition is constructed by warming up the model with the prior parameter, thus IC-WUP contains useful information of prior parameter, even it is biased. Besides, the state covariance matrix is implicitly inflated due to the introduction of uncertain prior parameter ensemble during warming up. These two aspects ensure the robust performance of warm-up methods. However, the initial state ensembles of IC-HfSatu-500-Un and IC-

ObsInt-500-Un are independent from the prior parameter, which introduces additional uncertainties, making the data assimilation results worse. Therefore, even using a larger ensemble size and enlarging the state uncertainty (covariance inflation), warm-up methods are still the optimal choice for both EnKF and IES algorithms.

2. Line 222-226: When the initial condition ensembles are generated for IC-HfSatu, IC-ObsInt and IC-Flux, is uncertainty added? How exactly? The uncertainty in the initial water content must be represented in the initial ensemble. If no uncertainty is added, this could explain partly the inferior result compared to IC-WUP. Please clarify and discuss.

[Response]

Thank you for your careful reading. We are sorry that we did not explain the problem clearly.

(1) The initial conditions of IC-HfSatu and IC-ObsInt were assumed to be deterministic without uncertainty in the original manuscript. In contrast, IC-Flux was conducted by warming up the model for a period (length = "warm-up" time t_{wu}) with a constant infiltration flux until a steady-state soil profile can be obtained. Thus, the uncertainty of parameter is introduced to IC-Flux, IC-WUP and IC-WUE during the construction of initial ensembles.

(2) IC-HfSatu and IC-ObsInt are the most common and convenient methods to initialize the soil water/hydrological model, while most applications of these two methods do not consider including the parameter uncertainty (much larger than the magnitude of observation error) during the construction of initial conditions. We want to know how exactly they affect the data assimilation results and whether we can utilize these first-cut methods to initialize the model within data assimilation framework.

(3) In order to further explore the effects of the uncertainty of the initial ensemble on the parameter estimations for these two approaches, we add a Gaussian noise (with a standard deviation of 0.017, which is calculated based on the spreading of IC-WUP initial ensemble) to both IC-HfSatu and IC-ObsInt (hereafter referred as IC-HfSatu-500-Un and IC-ObsInt-500-Un). We compared the results of parameter estimations

between them and the other initial conditions, as presented in the updated Fig. 7 (revised manuscript).

[Changes in the manuscript]

(1) We have added the content "In addition, IC-HfSatu and IC-ObsInt are assumed to be deterministic without uncertainty, while for the IC-Flux, IC-WUP and IC-WUE, the uncertainty of states are introduced by warming up the model with uncertain parameters." Please see lines 344-346.

(2) We have added the cases and discussions about the effects of the initial state with or without uncertainty. Please see our response to your first comment.

3. Line 247: The spatial resolution of 5cm is rather low for a 1 dimensional case. Is this a computational limitation? Otherwise, I would suggest to reduce the grid size to e.g. 2cm. This is especially relevant for sandy soils where sharp infiltration fronts can develop and require such high resolutions. Could this impact the results?

[Response]

Thank you for your suggestion. To understand the effects of grid size on the results, we compared the temporal change of the soil moisture profiles for the Loam soil (i.e., default soil type) with the grid size of 2 cm and 5 cm, as presented in Figure below.

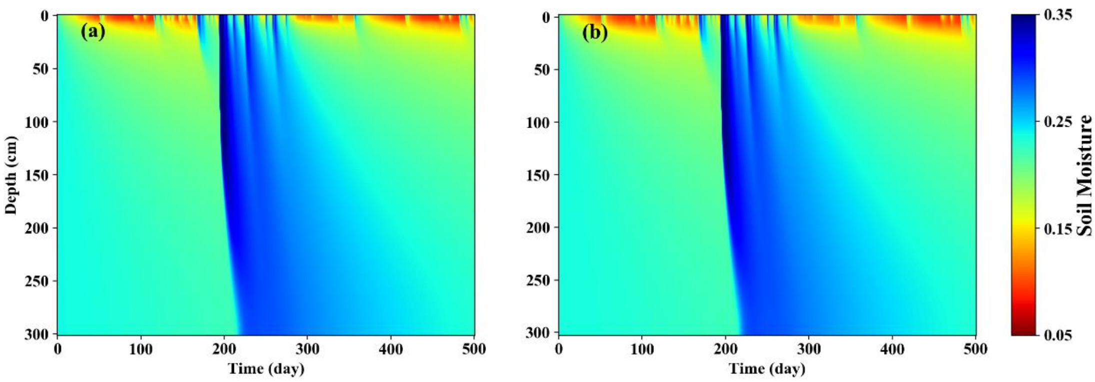


Figure 2. The temporal change of soil moisture profile for the Loam soil with the grid size of 2 cm (a) and 5 cm (b).

We admitted that the grid probably is too coarse to accurately capture the soil moisture dynamics, especially the sharp wetting front. Nevertheless, the difference between the overall results with the grid sizes of 2 cm (Figure 2(a)) and 5 cm (Figure 2(b)) is insignificant. Since our purpose is to explore the temporal change of UIC, which is an overall statistical index of the soil moisture profile, we think this grid is justifiable.

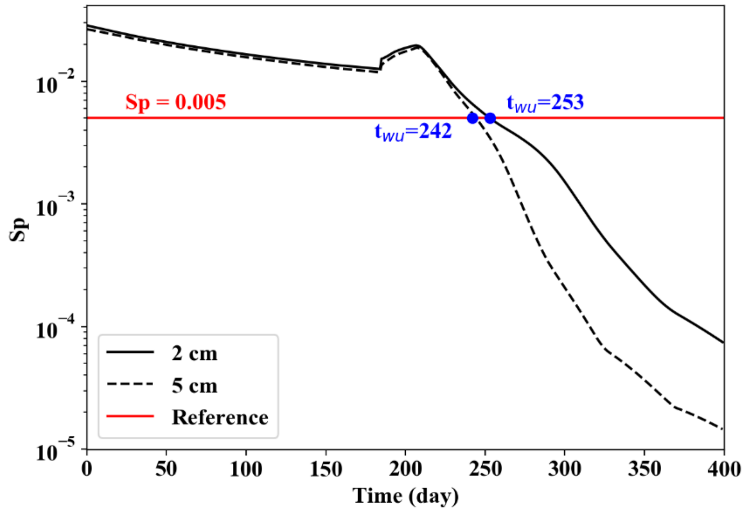


Figure 3. The spread index value for the Loam soil over time with the grid size of 2 cm and 5 cm.

In order to explore the effects of grid size on warm-up time, the spread values over time for the Loam soil with different grids are plotted in Figure 3. The relative difference between the two t_{wu} values is around 4%. Hence, the grid size has insignificant effects on the conclusions of our study.

[Changes in the manuscript]

(*previous manuscript*) The flow domain is discretized into 60 grids with a grid size of 5 cm.

(*revised manuscript*) The flow domain is discretized into 60 grids with a grid size of 5 cm, which has been proved to be sufficient for evaluating UIC in our study (results not shown).

4. Lines 379-383 and Figure 7: The biases of Ks, α, and n as well as their uncertainties differ. Therefore, their RMSEs should not be compared directly. I think it is not a meaningful result that α, which has the largest initial bias and uncertainty, also has the larger RMSE and that n, which has the smallest initial bias and uncertainty, has the smaller RMSE. Their relative improvement might be a better measure.

[Response]

Thank you for your excellent suggestion. To give a fair assessment on the improvement of data assimilation results, we have added the relative error index (*RE*) into the manuscript, which is calculated as,

$$RMSE_m = \sqrt{\frac{1}{N_e}\sum_{j=1}^{N_e}(m_j^E - m^T)^2} \tag{1}$$

$$RE = \frac{RMSE_m^e}{RMSE_m^p} \tag{2}$$

where m_j^E represents the estimated parameter of realization *j* at the last simulation day (EnKF) or the last iteration (IES); m^T represents the true parameter listed in Table 1. N_e is the total number of the realizations. $RMSE_m^e$ and $RMSE_m^p$ represent the $RMSE_m$ of the estimated and prior parameters. Compared with the $RMSE_m$, this index can better present the improvement of parameter estimation during data assimilation. We have modified the figures.

[Changes in the manuscript]

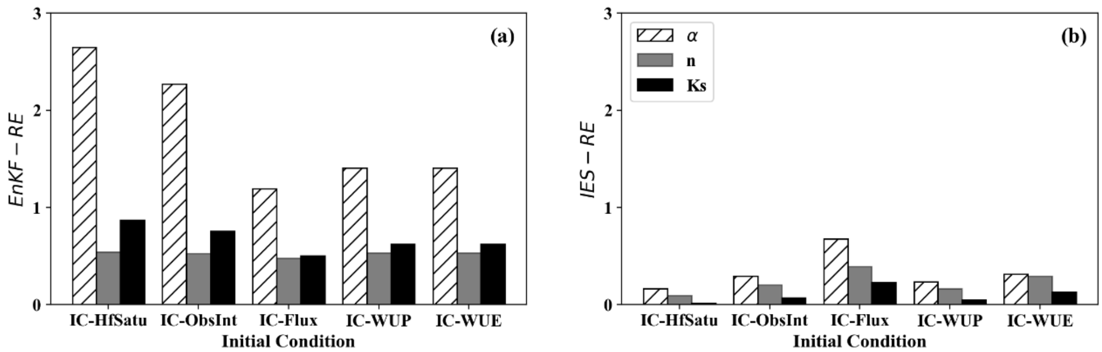


Fig. 8. (revised manuscript) The *RE* results of parameter estimations (*α*, *n* and K_s) under five initialization methods with (a) EnKF and (b) IES.

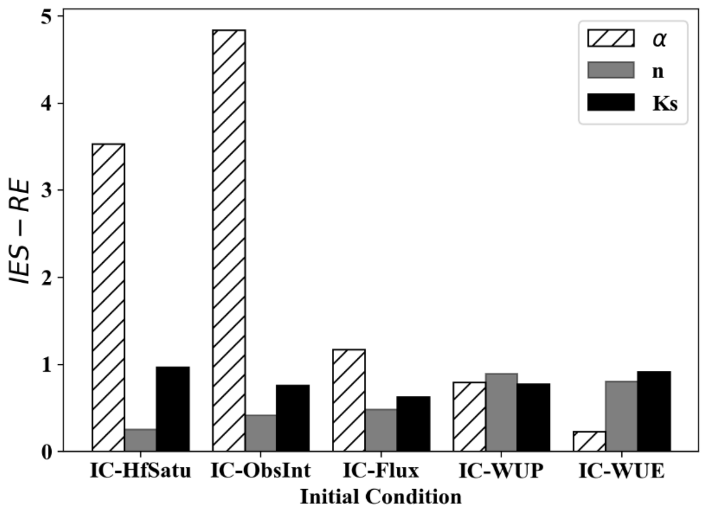


Fig. 9. (revised manuscript) The *RE* results of parameter estimations under five initialization methods with IES when the simulation time is 60 days.

5. Line 286: You find that t_{wu} *is less than one day for sand. I think that this result might be due to the chosen initial condition for the warm-up. It is true, that for the chosen high water contents sand will drain very fast and rapidly approach a similar water content state. However, in case of an initial condition in a very dry state (which should be relevant for the arid climate), the hydraulic conductivity of sand drops to very low values and the initial spread can extend for a very long time, or until a sufficiently large rain event increases the water content and then leads again to the rapid approach of the similar water content. I think it would be interesting to investigate this by choosing a different (dry) initial condition. At least this should be discussed in the manuscript.*

[Response]

Thank you for your valuable comment. To further investigate this problem, we conducted another two Monte-Carlo simulations for sandy soil with wet and dry initial conditions (i.e., the mean values of soil moisture ensemble are 0.2375 and 0.15 respectively with the same standard deviation of 3%). The temporal change of the spread S_p and the corresponding t_{wu} are presented in Figure 4.

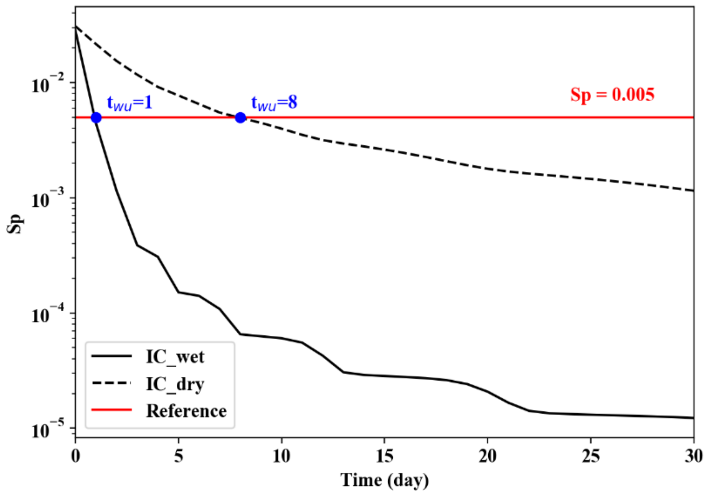


Figure 4. The temporal change of spread for sandy soil with a wet and dry initial condition separately (i.e., the mean of soil moisture ensemble is 0.235 and 0.15 respectively with a standard deviation of 3%).

The results confirm the reviewer's comment: a drier initial condition leads to the increase of warm-up time. Starting with a high soil moisture, the sand drains rapidly; when the soil is very dry, the hydraulic conductivity is extremely small and the initial spread survive for a long time.

[Changes in the manuscript]

We have added the discussion about the effects of the mean value of the initial soil moisture ensemble on the warm-up time in Section 3.2.2: "It should be noted that the t_{wu} is also relevant to the initial state of soil. Regarding the initial condition in an extremely dry state under the arid climate, the hydraulic conductivity is very small, and the initial spread extends for a long time. For instance, t_{wu} of sand increases from 1 day to 8 days when the ensemble mean value of initial soil moisture decreases from 0.2375 to 0.15 (results not shown). Yet, if a sufficiently large rain event takes place, the soil moisture increases and then converges to a similar state rapidly." Please see lines 312-316.

6. Line 273-275 and 477-478: Since you recommend the choice 0.5% as a threshold: Please explain why. What is the advantage? Why should I not choose the other mentioned thresholds (e.g. 1% or 0.1%)?

[Response]

The threshold of 0.5% is recommended due to a reasonable trade-off between the model accuracy and computational cost.

[Changes in the manuscript]

An explanation is added in Section 3.2.2: "As shown in Fig. 3(c), there is a significant diversity of the results between Spin-up and Monte-Carlo methods at index value of 1%, indicating that UIC still plays a significant role. In contrast, the requested t_{wu} is more than 15 months for a value of 0.1%. To balance the estimation accuracy and computational cost, we recommend a threshold of 0.5% for both Spin-up and Monte-Carlo methods, and the corresponding warm-up time t_{wu} is 8 months, which is sufficiently long for UIC to diminish and the difference between PC and S_p is insignificant." Please see lines 275-280.

7. Lines 306-313 and Figure 5: If I understand correctly, you investigate when the uncertainty for the full profile drops below the 0.5% threshold. In addition (or possibly as replacement) I would find it interesting to see the spatially resolved times for each depth for the deepest profile (20 m).

[Response]

Thank you for your valuable comment. We have added a new subfigure (Fig. 5(b)) in the manuscript, which presents the t_{wu} value for each depth with a 20-m soil profile, as presented below. The result still supports our previous conclusion, and more details about UIC along the soil profile have been displayed and analyzed in the revised manuscript.

[Changes in the manuscript]

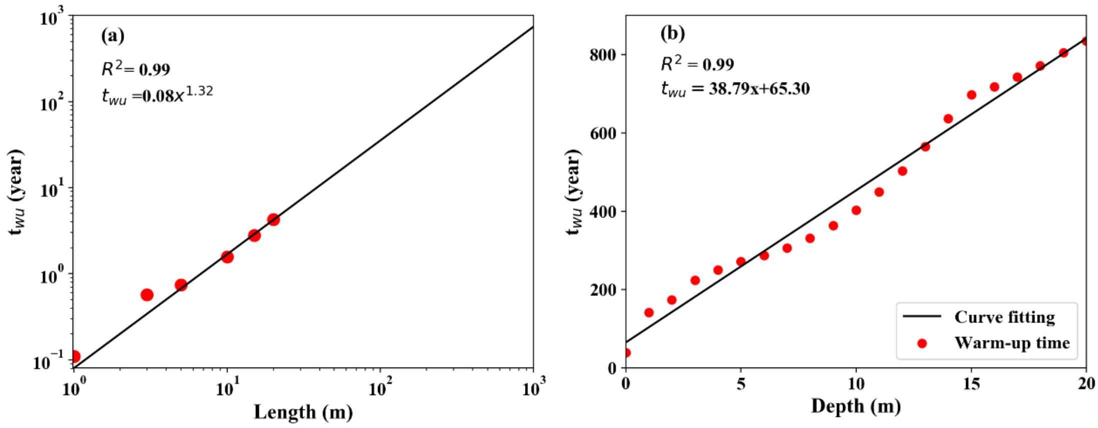


Fig. 5. The value of the warm-up time t_{wu}. (a) The overall profile t_{wu} values versus different soil profile lengths and (b) t_{wu} value as a function of depth z with a 20-m soil profile.

8. Line 152 (Equation 8): If the monthly average of the previous year is required, wouldn't that imply that PC is not defined for the first year? However Fig. 3 shows PC starting from time 0. Please clarify.

[Response]

We are sorry that we did not explain it clearly. *PC* is an index that reflects the deviation of soil moisture between two adjacent years in a recursive run after a period of warm-up time t_{wu}. Following de Goncalves et al. (2006) and Ajami et al. (2014), *PC* at month $t=t_{\mathrm{wu}}$ is calculated by comparing the relative difference of soil moistures at month $t=t_{\mathrm{wu}}$ and month $t=t_{\mathrm{wu}}+12$. As presented in Fig. 3(a) in the manuscript, *PC* at month t=12 is close to 0.

[Changes in the manuscript]

We have updated Equation 8 as

$$PC(t) = 100\left|\frac{M(t) - M(t+12)}{M(t+12)}\right| \tag{8}$$

where $M(t)$ and $M(t+12)$ are the monthly averaged soil moistures after model spin-up for t months and t+12 months. Please see lines 160-162.

9. Figure 3: Why does the water content state after 24 months differ between panel (a) and (b)? Since both are initialized with the same parameter values and the UIC has *already decayed, they should show the same soil moisture. Please clarify.*

[Response]

Good eyes! Thank you for pointing out this problem. We made a mistake when calculating the monthly-average soil moisture at t = 24 month in Fig. 3(b). This error is amended in the updated figure.

[Changes in the manuscript]

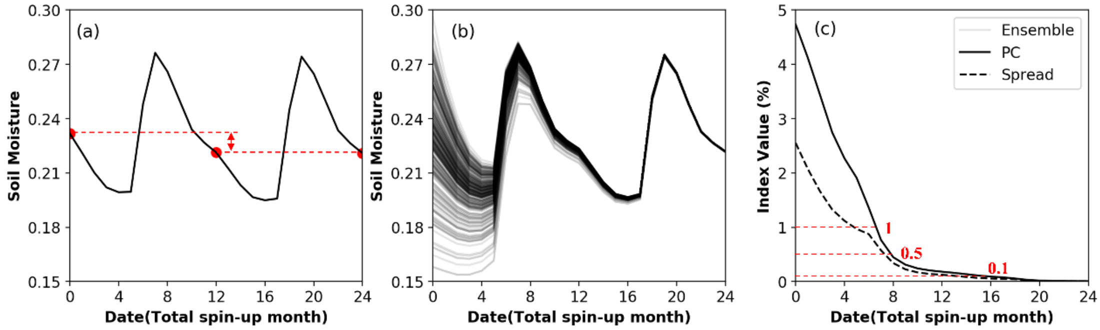


Fig. 3. Comparison of Spin-up and Monte-Carlo methods in determining warm-up time. (a) Spin-up method with monthly-averaged soil moisture versus time by running a simulation recursively for 10 years, (b) Monte-Carlo method with monthly-averaged soil moisture of different realizations versus time based on various initial conditions, and (c) Comparison of *PC* and S_p versus time. For the purpose of demonstration, the parameter uncertainty is not considered and we only show the results of the first two years in the figure.

10. Lines 370-377: Based on Figure 6, I disagree with the statement that filter inbreeding is not a significant issue for the EnKF case. In Figure 6, it seems that the final parameter value does not change any more over time and is over 5 standard deviations away from the truth. This means that the uncertainty is too small. Part of the reason could be that the initial uncertainty is chosen way too small. It is over 9 standard deviations away from the true value. This makes it very difficult for the EnKF to find the true value. I would suggest to repeat the simulations with a larger parameter uncertainty.

[Response]

Thank you for your suggestion. In order to explore the effects of parameter uncertainty on the data assimilation results, we compared the parameter estimations of lnK_s with various initial standard deviations of initial parameters (0.1 and 1 respectively), as presented in Figure 5 below. The results agree with our previous conclusion.

(1) We admit that the data assimilation results could be enhanced with a larger parameter uncertainty for the EnKF case, since the parameter updates more rapidly than the small-variance case (Figure below), and could prevent possible inbreeding problems. As shown in the figure, the data assimilation results with a larger parameter uncertainty (lnK_s-1 (HfSatu) and lnK_s-1 (WUP)) are better than those with a smaller one (lnK_s-0.1 (HfSatu) and lnK_s-0.1 (WUP)).

(2) WUP is still the best initialization method among the five approaches, regardless of lnK_s variance of 0.1 or 1.0. As shown in the figure, the *RMSEs* of lnK_s-1 (WUP) and lnK_s-1 (HfSatu) are 0.13 and 0.36 respectively. Since our topic of this study is to demonstrate the effects of uncertainty of initial condition and initialization methods, we have not revised the model inputs of prior parameters but added the discussions about enlarging the initial parameter uncertainties and filter inbreeding issue in the revised manuscript.

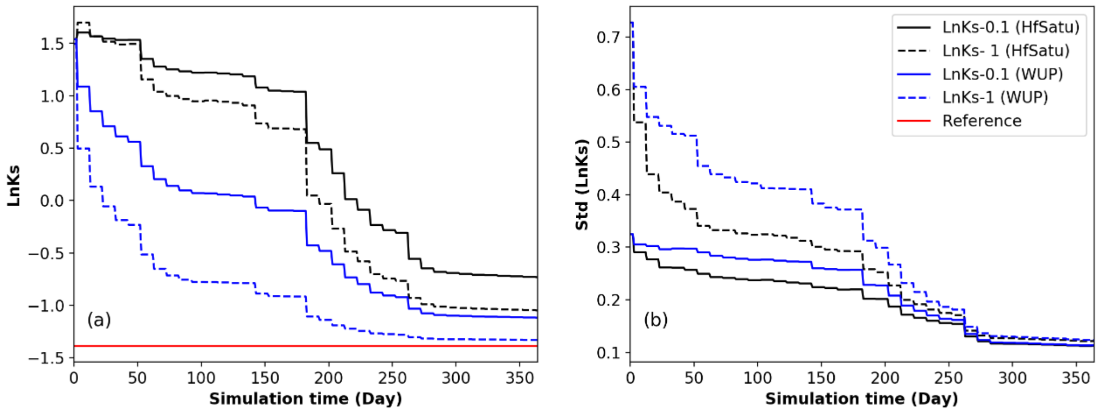


Figure 5. The temporal change of estimated mean and standard deviation of lnK_s with various parameter uncertainties.

[Changes in the manuscript]

We have added our discussion about filter inbreeding and initial parameter uncertainties in Section 3.3.2.

11. Line 177-178: ": : :, uk are state variables (i.e., pressure head and soil *moisture) : : :". Do you update water content and matric potential of the same node simultaneously in the augmented state? Due to their nonlinear relation, the analysis would lead to inconsistencies between water content and matric potential for the analysis. How is this handled in the forecast? Please clarify.*

[Response]

Thank you for your careful reading. In this study, we only update the soil moisture in the simulation and convert the updated soil moisture ensemble into the pressure head to drive the model. The above statement is confusing and we have revised it.

[Changes in the manuscript]

(*Previous manuscript*) $\mathbf{u}_k$ are state variables (i.e., pressure head and soil moisture) at time t_k.

(*Revised manuscript*) $\mathbf{u}_k$ are the state variables (i.e., soil moisture) at time t_k. The updated soil moisture ensemble can be converted to pressure head to drive the model.

Technical comments:

Lines 228-238: This part describes IC-WUP and IC-WUE. However, this is not a general description. In Section 3.2, when the spin up periods are investigated, a different procedure is used. This confused me when reading the paper the first time. Please, only mention the general settings in 3.1 (i.e. climates, soils and model representation), and not specifics that only apply to 3.2 or 3.3. Therefore I would suggest to move this part to Section 3.3. Additionally, here it is not clear how the parameter and initial condition ensembles are exactly generated. Please clarify.

[Response]

Thanks for your valuable suggestions and comments. We are sorry that we did not make the description clear.

IC-HfSatu is a uniform soil moisture profile with the 50% relative saturation (e.g., θ=0.254 loam) of soil; IC-ObsInt is a linear interpolation between observations at the beginning of simulation. The depths of the initial observations are 10 cm, 80 cm, 150 cm, 220 cm and 290 cm; IC-Flux is a steady-state soil moisture profile by warming up the model with a constant infiltration flux (1 mm/d). Besides, the initial conditions of two warm-up methods are given by running the model prior to the beginning of simulation period with available meteorological data (as shown in Fig. 2). If the meteorological data before the simulation period is available, it is used in the warm-up method to obtain the initial condition (IC-WUP); otherwise, we use the meteorological data at the simulation period (IC-WUE) as a surrogate. The length of warm-up time for IC-Flux, IC-WUP and IC-WUE is equal to t_{wu} (242 days) according to the results in Section 3.2.2(a) , so the warming-up period of WUP for these three methods is from day 124 to day 365. In addition, IC-HfSatu and IC-ObsInt are assumed to be deterministic without uncertainty, while for the IC-Flux, IC-WUP and IC-WUE, the uncertainty of states are introduced by warming up the model with uncertain parameters. The initial realizations of soil hydraulic parameters K_s, α and n are generated following logarithm normal distributions, with mean values of 4.7 m d^{-1}, 8.6 m^{-1} and 1.8, and variances (log-transformed) of 0.1, 0.3 and 0.006. The saturated soil moisture θ_s and residual soil moisture θ_r are assumed to be deterministic with the value of 0.43 and 0.078.

[Changes in the manuscript]

We have added the explanation in the manuscript to Section 3.3 as suggested.

Line 222-226: I think this part should be moved to Section 3.3 as well.

Thanks. This part has been moved.

Line 243: “Fig. 1” should be “Fig. 2”.

Thank you. The error has been corrected.

Line 254 and Fig. 3: The text mentions a simulation length of 10 years, the figure shows only 2 years. I would suggest to mention that you only show the first 2 years.

Thanks. This has been revised.

Lines 338-342: How many observations are there? In what depths are the observations? What is the assimilation frequency? Or is only a single observation in the depth of 10 *cm assimilated every 10 days? If that is the case this has to be clarified.*

[Response]

Thank you for your suggestion. The observations are only collected at the depth of 10 cm and assimilated every 10 days, starting from day 3. Unless otherwise specified, the total numbers of the observations are 37 (3rd, 13th, 23th,…, 363th days).

[Changes in the manuscript]

(*Previous manuscript*) In addition, the observation at 10 cm is assimilated into model every 10 days.

(*Revised manuscript*) A total number of 37 observations are assimilated into the model. The observation depth is at z = 10 cm and the observed soil moisture is assimilated every 10 days, starting from day 3.

Lines 343-350. I think this part should be moved to methods in Section 2.

Thanks. Revised.

Line 352-353 and Figure 6: I would mention that this is case 1 and case 2.

Thanks. This has been modified.

Line 399: "Field" instead of "Filed".

Thank you. Revised.

Figure 4: Since essentially the times for sand for all climates as well as silt and clay loam for the M-AC and the M-SC climate can not be properly displayed: Maybe the logarithm of the time could be more meaningful (like in Fig. 5).

[Response]

Thank you. Fig. 4 has been revised in the manuscript according to the suggestions from you and reviewer 2. We added a case to investigate the t_{wu} value in a multiple-layers soil.

[Changes in the manuscript]

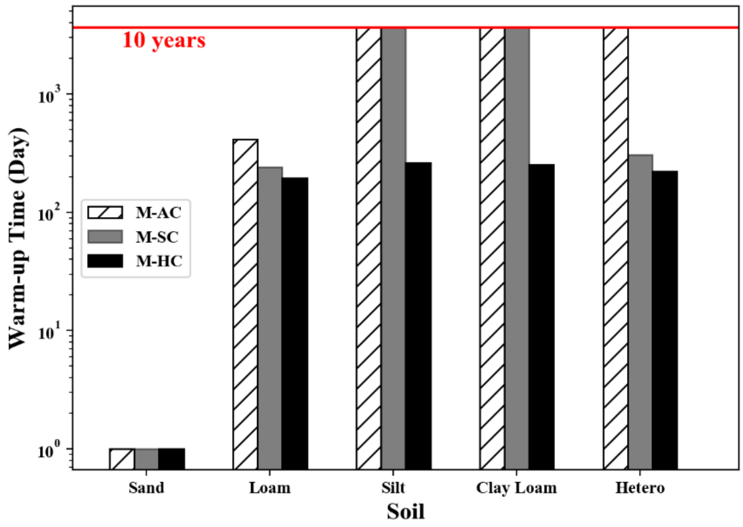


Fig. 4. The length of warm-up time t_{wu} with various soils and meteorological conditions. Note that some of the t_{wu} values are larger than 10 years and are not able to be obtained due to the 10-year simulation time. The heterogeneous soil profile consists of Loam (0-75 cm), Clay loam (75-150 cm), Silt (150-225 cm), and Sand (225-300 cm).

Figure 6: The line for IC-WUE is essentially not visible. Is it below IC-WUP? At least mention this in the caption.

[Response]

Thank you for pointing out this problem. We are sorry that we have used mistaken model input folder when analyzing the data from WUE and WUP, and two identical curves for WUE and WUP were generated. The error has been fixed, and the figure has been updated. Although there is a little difference between Figs. 6 (a) and (b), the general conclusions are consistent with the previous ones.

[Changes in the manuscript]

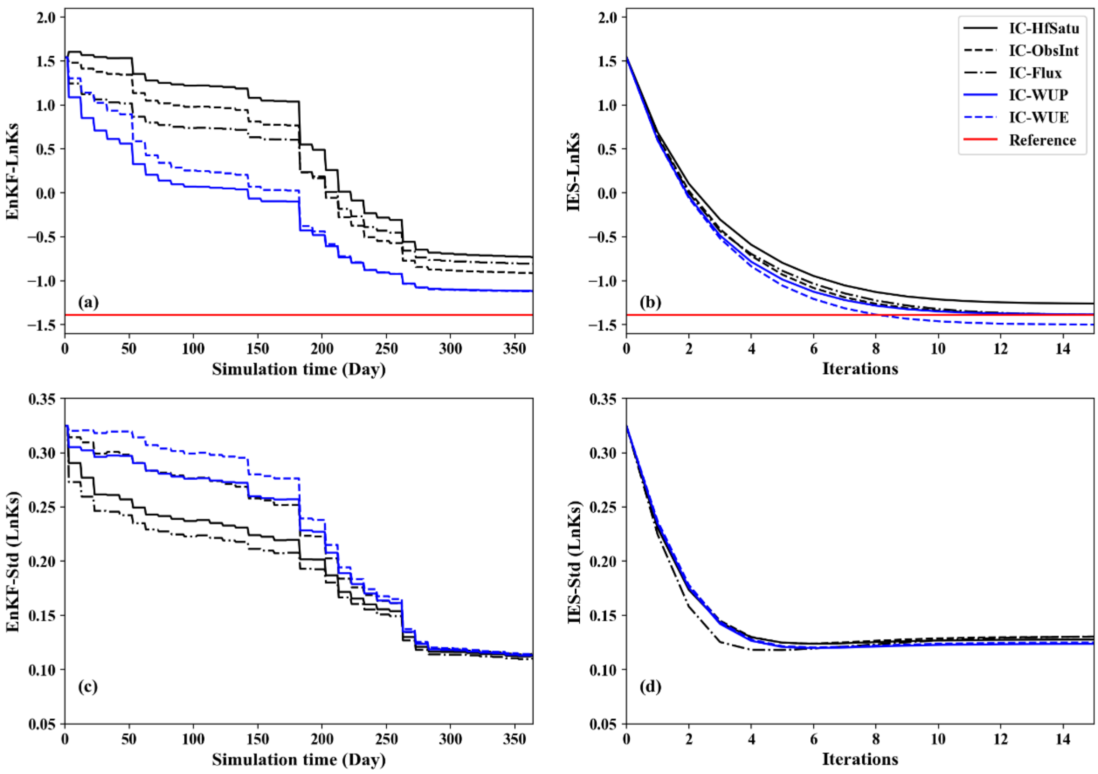


Fig. 6. The results of lnK_s estimations (first row) and their associated standard deviations (second row) within two data assimilation frameworks (left: EnKF; right: IES) under five initialization methods.

Reference

Ajami, H., McCabe, M. F., Evans, J. P. and Stisen, S.: Assessing the impact ofmodel spin-up on surface water-groundwater interactions using an integrated hydrologic model, Water Resour. Res., 50, 1–21, doi:10.1002/2013WR014258.Received, 2014.

De Goncalves, L. G. G., Shuttleworth, W. J., Burke, E. J., Houser, P., Toll, D. L., Rodell, M., and Arsenault, K.: Toward a South America Land Data Assimilation System: Aspects of land surface model warm-up using the Simplified Simple Biosphere, J. Geophys. Res. Atmos., 111(17), 1–13, doi:10.1029/2005JD006297, 2006.

Vereecken, H., Kamai, T., Harter, T., Kasteel, R., Hopmans, J. and Vanderborght, J.: Explaining soil moisture variability as a function of mean soil moisture : A stochastic unsaturated flow perspective, , 34(October), 1–6.

Reply to comments from Anonymous Referee #2.

General Comments

This study investigates the temporal change of the uncertainty of initial condition in variably saturated flow model and assesses the impacts of several commonly-used initializing methods on results within various data assimilation frameworks. The topic is interesting and relevant to the topics of the Hydrology and Earth System Sciences. The manuscript is well-organized and easy to understand, although some of language, may be further refined and improved. The results and discussion are adequate to reach very instructive conclusions for variably saturated flow modeling. Several highlights for this manuscript: compared to previous researches on UIC issue, this study focuses on soil water modeling and makes a comparison between Monte Carlo (preferred by groundwater hydrologist) and Spinning up methods (preferred by surface water hydrologist). The investigation of warm-up time with different soil textures and depths is quite interesting. The study of UIC propagation with data-model interaction is another merit. Therefore, I recommend this paper for publication in the Hydrology and Earth System Sciences, with a few comments.

[Response]

Thank you for your positive comment!

Major comments:

1) Authors have compared the difference of model outputs with various data assimilation framework (i.e., EnKF and IES). As the authors correctly point out, the ensemble size is an important factor for these two algorithms, which need to be discussed further. I encourage the authors to explore the effects of ensemble size on EnKF and IES with multiple test so that a suitable ensemble size for these two assimilation framework can be determined.

[Response]

Thank you for your comment, according to the suggestions from you and reviewer 1, we have added a new figure to explore the effects of ensemble size on the parameter estimations within EnKF and IES.

[Changes in the manuscript]

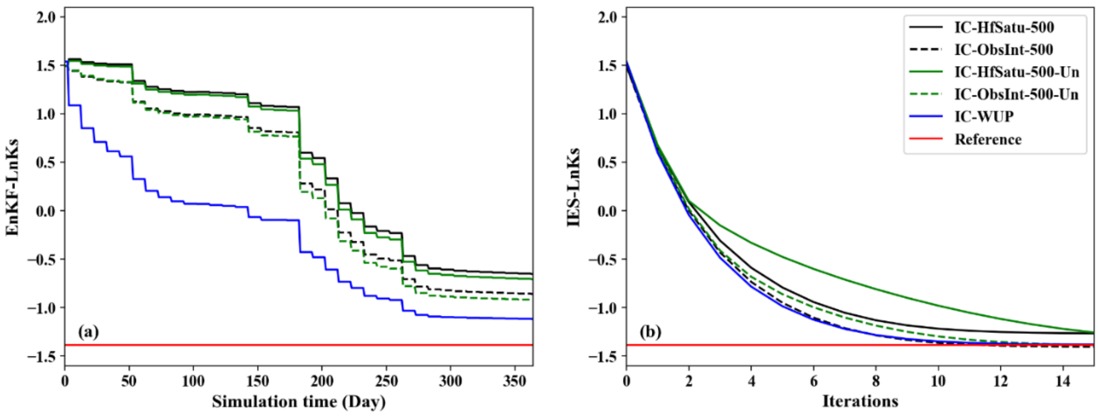


Fig. 7. (revised manuscript) The impacts of increased ensemble size and the uncertainty of initial state on the results of lnK_s estimations within EnKF and IES.

The results of IC-HfSatu-500 and IC-ObsInt-500 with the ensemble size of 500 in Fig. 7 are similar with those of IC-HfSatu and IC-ObsInt (Fig. 6), indicating that the improvement of the parameter estimation result is slight when the ensemble size increases from 300 to 500. Hence, the ensemble size of 300 is sufficient for data assimilation problem in this study.

2) The synthetic case study present the proper warm-up time t_{wu} versus different soil texture, soil depth, and meteorological conditions. While the relationship between t_{wu} and meteorological conditions may be commonsense, the reveal of quantitative relationship between t_{wu} and soil texture and soil depth is surprising and interesting, due to the fact that t_{wu} changes abruptly from sand to finer texture, and it increases nonlinearly with the increase of soil depths. However, the soil is seldom homogeneous in natural conditions, especially for very long soil profile. The authors should at least present one simulating result of t_{wu} for layered soil, which is more applicable for real-world case. I believe this should take too much work since it is one-dimensional model.

[Response]

Thank you for your valuable comment. We have added a case to obtain the t_{wu} for a layered soil profile, which consists of Loam (0 to 75 cm), Clay loam (75 to 150 cm), Silt (150 to 225 cm) and Sand (225 to 300 cm).

[Changes in the manuscript]

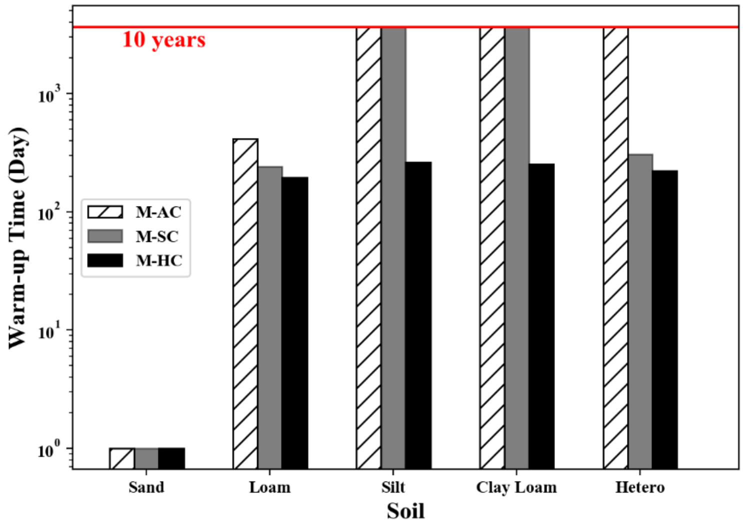


Fig. 4. The length of warm-up time t_{wu} with various soils and meteorological conditions. Note that some of the t_{wu} values are larger than 10 years and are not able to be obtained due to the 10-year simulation time. The heterogeneous soil profile consists of Loam (0-75 cm), Clay loam (75-150 cm), Silt (150-225 cm), and Sand (225-300 cm).

In the revised manuscript, we have expanded the results and presented the t_{wu} value in layered soil: "Moreover, regarding the heterogeneous soil with multiple layers, the t_{wu} under the M-AC is larger than 10 years (similar to Silt and Clay loam), while that under M-SC or M-HC becomes much smaller (higher than that of Loam but they are of the same magnitude). Thus, it is conjectured that t_{wu} is determined by the fine soil texture in the layered profile under dry meteorological condition, but averaged soil hydraulic properties under wet meteorological condition". Please see lines 307-311.

Minor comments:

Line 12: various initial condition »> various initial conditions

Thank you. Revised

Line16: model initializing »> model initializing methods

Thank you. This has been revised.

Line 28: delete in

Thank you. Revised.

Line 48: a space between approaches and comma

Thanks. The error has been corrected.

Line 61: hereafter referred »> hereafter referred to

Thank you. It is revised

Line 77: delete the last the

Thank you. It is modified.

Line 81: initial ensemble are »> initial ensemble is

Thank you. It is rewritten.

Line83: Currenlty »> Currently

Thank you. Revised

Line 110: Richards's »> Richards'

Thanks. Revised

Lines 129-130: as state dependent, atmospheric boundary condition (try to be more concise here and some other statements)

Thanks. It is rewritten.

Line 135: detemined »> determined

Thanks. Revised

Line 141: use UIC instead Eqs. (9-10): try to use one equation instead and shorten the description of the equation.

Thank you. It is revised.

Line 172: assimilation + approach

Thanks. Revised

Line 210: which λ values you use in the simulations?

Thanks. λ is a dynamic stability multiplier during the iterations. The prior value of λ is 10, but the value can be adjusted adaptably according to the data assimilation results at every iteration.

Line 222: perscribe »> prescribe

Thank you. It is revised.

Line 223: availablitity »> availability

Thanks. Revised.

Line 256: be consistent using italic or not for PC.

Thank you. This has been revised.

Line 256: why 3%?

Thank you. The Gaussian noise is determined as 3% according to the observation error of soil moisture since the uncertainty of parameter is not taken into consideration in this part. We have added a sentence to explain it.

Line 335: warms »>warm

Thanks. Revised.

Line 356: delete both

Thanks. This has been revised.

Line 358-359: thus »> and thus

Thanks. Revised.

Line 372: multiple spaces between runs and are.

Thanks. Revised.

Change "than" to that

Thanks. This has been revised.

Line 405: Which evapotranspiration model are you using?

Thank you for your comment. The potential evaporation is calculated by Penman-Monteith's equation. We will add an explanation in the manuscript.

Line 427: needs a space after "part."

Thanks. Revised.

Lines 443-444: "soil moisture profile has large variation, e.g., discontinuous soil moisture in layered soils." — it would be interesting to see an additional case for heterogeneous soils, and this also leads to another interesting question — what will happen if pressure head profile, which is continuous in heterogeneous soil, is used as initial condition. Please add some discussion on this topic.

[Response]

Thank you for your valuable comments. We have added the case about t_{wu} of layered soil in the manuscript, please see Fig. 4 above. Regarding the topic about using initial pressure head as initial condition, we are going to discuss it from three aspects:

(1) It is easier to collect the soil moisture data than soil pressure head, so that we only

use soil moisture as observation in this study. In heterogeneous soil, the pressure head profile is continuous compared to soil moisture profile, which is the reason why head is preferred as the state variable in numerical models.

(2) With respect to t_{wu}, the conversion relationship between the spread of soil moisture and pressure head is deterministic (i.e., the spread as well as the t_{wu} value with pressure head profile can be derived from those with soil moisture profile). Thus, in this study we choose soil moisture in the study of UIC and t_{wu}.

(3) The impact of observation type (i.e., pressure head and soil moisture) on data assimilation results has been widely explored in previous studies (Shi et al. 2015), since these two state variables have different probability distributions, nonlinearity, accessibility, and observation errors. Although choice of head or moisture can affect the data assimilation results, it is not primarily induced by the difference in UIC, which is focus of current study.

[Changes in the manuscript]

We have added some discussions about the effects of initial pressure head profile in heterogeneous soil in Section 5. Please see lines 545-552.

Line 452: atmospheric condition »> atmospheric boundary condition

Thanks. Revised.

Conclusion 2: Please include more details and add quantitative conclusions for this.

[Response]

Thank you. We have modified the conclusion 2 according to your suggestion.

[Changes in the manuscript]

Warm-up time varies nonlinearly with soil textures, meteorological conditions, and soil profile. Under most situations (e.g., Loam with the soil profile length less than 5 m under non-arid climate), one-year warm-up time is sufficient for soil water movement modeling, but an extremely long time (exceeds 10 year) is needed to warm up the model for a long, fine-textured soil profile under an arid meteorological condition.

Errors in references: Line 566, Line 673, Line 610, Line 639.

Thanks. Revised.

Reference

Shi, L., Song, X., Tong, J., Zhu, Y. and Zhang, Q.: Impacts of different types of measurements on estimating unsaturated flow parameters, J. Hydrol., 524, 549–561, doi:10.1016/j.jhydrol.2015.01.078, 2015.

Reply to comments from Anonymous Referee #3

General Comments

This paper studies an important problem of soil water modeling: the uncertainty of initial condition (UIC) through analyzing the effects of different initial conditions on parameter estimation within two data assimilation frameworks. I believe this work provides useful insights to improve our understanding of uncertainty of initial conditions. I would be in favor of publication after the authors addressed the comments given below.

[Response]

Thank you for your positive comment! We have revised our manuscript according to your suggestions.

Comments:

1. The grammar of this paper needs some improvements, some small grammar errors can be found.

Thank you for your suggestion. We have invited a native English speaker to proofread our manuscript.

2. The quantification of initial condition uncertainty (UIC) is unclear, especially for the usage of data assimilation method. I don't follow how two methods combined.

[Response]

Thank you for your valuable comment. We hope our response will make it clearer. At t=0, UIC is introduced when we specify the initial condition for modeling due to our imperfect knowledge of it. In our study, the quantification of UIC is based on a Monte-Carlo simulation method. To be specific, UIC at depth z and time t is measured by the spread (standard deviation) of the soil moisture at depth z. We find UIC decays with simulation time t and a sufficient long warming-up time will let the UIC vanishes. Here warming-up means that we simulate the model prior to the desired simulation period and the result is given as initial condition (see our illustration Fig. 1 in the revised manuscript). Evaluation of UIC and our data assimilation are smoothly combined since we choose EnKF and IES, both of which are also Monte-Carlo method. At t=0, we generate an ensemble of soil moisture profiles based on one initialization method (which introduces UIC), and use this ensemble to initiate the data assimilation (assimilate observations and estimate parameter). Finally, we can evaluate our data assimilation performance based on different initializing methods.

[Changes in the manuscript]

We have added some explanations and modified our manuscript to make the description clearer. Please see lines 143-154 in the revised manuscript.

3. The purposes of using data assimilation method and its relationships to results and conclusions are unclear.

[Response]

On the one hand, data assimilation are widely used in vadoze zone hydrology for accurate estimation of hydraulic parameters/future states, and reduction of uncertainties, using observation. On the other hand, UIC is an important uncertain factor that affect the accuracy of simulation, and some of initialization methods have been proposed in forward simulations. However, to our best knowledge, there is no systematic research on the effects of UIC and different initialization methods on data assimilation. The topic of our research, i.e., combination of UIC and data assimilation, is reasonable since they both deal with the state uncertainty (UIC itself is the state uncertainty at t=0, and is one source of state uncertainty). Some of our results and discussions are on UIC alone, in which neither observation nor parameter uncertainty is involved. Some other discussion and results focus on the effects of UIC and different initialization methods on data assimilation, in which UIC and parameter uncertainty coexist.

[Changes in the manuscript]

We have improved our manuscript according to your suggestions. Some explanations have been added about our purposes of using data assimilation method and we have made its relationships to results and conclusions clearer. Please see lines 75-97 and lines

143-154.

4. Please be more specific about why using both experimental and field model, and how different their results are.

[Response]

In this study, we conducted both synthetic and field experiments. In synthetic experiments the "observed" data are generated by running the forward model with the exactly known parameters, while the field data are collected in the experimental station. The field observations may contain a lot of uncertainties, such as unknown/inaccurate upper and lower boundary conditions, unknown observation error/bias, and unknown parameters. All these unknown uncertainties have impact on the modeling, while the effect from UIC could be overshadowed by those from other unknown uncertainties, and the direct results from field experiments could by inconclusive. By utilizing the synthetic observations, we can separate the effects on modeling from all these uncertainties, since they are all perfectly known. Based on the synthetic case, we can elaborate our conclusions on the temporal evolution of UIC, as well as its effect on data assimilation, which is assessed by the estimated parameter and perfectly-known true parameter. We think this is the logic way: first, a comprehensive investigation on UIC is conducted using synthetic case, then, the field data can be used to validate the applicability of our approaches/results.

The conclusions by using field and synthetic data are similar (the difference of results between various initialization methods are less significant in field case, due to contamination of other unknown uncertainties), indicating a good applicability of our approaches/results.

[Changes in the manuscript]

We have added an explanation "Synthetic observation in previous section is generated by running the model with exactly known uncertainty sources. By conducting synthetic experiments, we can thoroughly analyze the impact of UIC during data assimilation, with scenarios having different numbers of observations/unknown parameters, and more decisive conclusions can be drawn. In contrast, the field observations contain additional uncertainties which are largely unknown (e.g., the calculated evapotranspiration is inaccurate for real-world case). In order to examine the real-world applicability of the conclusions drawn from synthetic case, Field data are necessary to validate our results.” in the manuscript. Please see lines 462-468.

5. Please describe more details about the novelty of this paper, it seems there is no new method involved, and I am not sure how useful and novel the conclusions are.

[Response]

Although the initialization and data assimilation methods used in this study are not new, we claim the innovation of our studies mainly based on that, to the best of our knowledge, this is the first study analyzing the effects of initial conditions and initialization methods within various data assimilation frameworks to date. The novelties include three aspects.

(1) Two common approaches (Spin-up method and Monte Carlo method) for quantifying the temporal evolution of initial condition uncertainty are compared. Spin-up methods are widely used in large-scale hydrological model due to their smaller computational cost. However, Monte Carlo methods have the merit that they can explicitly quantify UIC, which is suitable for data assimilation. Finding an equivalency between these two methods can fill the gap between widely-used initialization methods and data assimilation, both of which are important tools increasing the accuracy of hydrological modeling. Also, new algorithm is not necessary for the combination of UIC/initialization and data assimilation, in both of which we use Monte Carlo method for expressing state uncertainty. This (i.e., no new algorithm) should be regarded as an advantage according to principle of parsimony, since our approach can be easily applied and validated by the readers.

(2) The influences of soil texture, meteorological condition and soil profile length on initial condition uncertainty evolution are exploited. Especially, we propose a warming-up time t_{wu}, which is defined as the time when percentage cutoff PC or ensemble spread S_p is lower that 0.5%, can guide us to select the warming up period and pick up observation at different time in data assimilation.

(3) Different approaches to initialize unsaturated-saturated flow models within two data assimilation frameworks (IES and EnKF) are assessed. Our studies focus on the case when both parameters and initial state are uncertain, and the combination of selected initialization method and data assimilation can be a standard approach for future variably saturated flow modeling.

[Changes in the manuscript]

We have modified our manuscript according to the discussion above to make the novelties of the paper more apparent.

Reference

Ajami, H., McCabe, M. F., Evans, J. P. and Stisen, S.: Assessing the impact ofmodel spin-up on surface water-groundwater interactions using an integrated hydrologic model, Water Resour. Res., 50, 1–21, doi:10.1002/2013WR014258.Received, 2014.

Brandhorst, N., Erdal, D. and Neuweiler, I.: Soil moisture prediction with the ensemble Kalman filter: Handling uncertainty of soil hydraulic parameters, Adv. Water Resour., 110(August), 360–370, doi:10.1016/j.advwatres.2017.10.022, 2017.

Li, C. and Ren, L.: Estimation of Unsaturated Soil Hydraulic Parameters Using the Ensemble Kalman Filter, Vadose Zo. J., 10(4), 1205, doi:10.2136/vzj2010.0159, 2011.

Montzka, C., Moradkhani, H., Weihermüller, L., Franssen, H. J. H., Canty, M. and Vereecken, H.: Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter, J. Hydrol., 399(3–4), 410–421, doi:10.1016/j.jhydrol.2011.01.020, 2011.

Shi, L., Song, X., Tong, J., Zhu, Y. and Zhang, Q.: Impacts of different types of measurements on estimating unsaturated flow parameters, J. Hydrol., 524, 549–561, doi:10.1016/j.jhydrol.2015.01.078, 2015.

Vereecken, H., Huisman, J. A., Bogena, H., Vanderborght, J., Vrugt, J. A. and Hopmans, J. W.: On the value of soil moisture measurements in vadose zone hydrology: A review, Water Resour. Res., 46(4), 1–21, doi:10.1029/2008WR006829, 2010.

Walker, J. P., Willgoose, G. R. and Kalma, J. D.: One-Dimensional Soil Moisture Profile Retrieval by Assimilation of Near-Surface Measurements: A Simplified Soil Moisture Model and Field Application, J. Hydrometeorol., 2(4), 356–373, doi:10.1175/1525-7541(2001)002<0356:ODSMPR>2.0.CO;2, 2001.

Wu, C.-C. and Margulis, S. A.: Real-Time Soil Moisture and Salinity Profile Estimation Using Assimilation of Embedded Sensor Datastreams, Vadose Zo. J., 12(1), doi:10.2136/vzj2011.0176, 2013.

On the uncertainty of initial condition and initialization approaches in variably saturated flow modeling

Danyang Yu[1], Jinzhong Yang[1*], Liangsheng Shi[1], Qiuru Zhang[1], Kai Huang[2], Yuanhao Fang[3], Yuanyuan Zha[1*]

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

[2] Guangxi Hydraulic Research Institute, Nanning, Guangxi 530023, China

[3] School of Earth Sciences and Engineering, Hohai University, Nanjing 210098, China

* *Corresponding author*: Jinzhong Yang (jzyang@whu.edu.cn); Yuanyuan Zha (zhayuan87@whu.edu.cn)

Abstract:

Soil water movement has direct effects on environment, agriculture and hydrology. Simulation of soil water movement requires accurate determination of model parameters as well as initial and boundary conditions. However, it is difficult to obtain the accurate initial soil moisture/matric potential profile at the beginning of simulation time, making it necessary to run the simulation model from arbitrary initial condition until the uncertainty of initial condition (UIC) diminishes~~. The behavior of this process~~, which is ~~usually defined~~often known as “warming up”. In this paper, we compare two commonly used methods for quantifying the UIC (one is based on running a single simulation recursively across multiple hydrological years, and the other is based on Monte-Carlo simulations with realization of various initial ~~condition~~conditions) and identify the “warm-up” time ~~t_{wo}~~t_{wu} (minimum time required ~~for the model to warm up~~ to eliminate the UIC by warming up the model) required with different soil textures, meteorological conditions, and soil profile lengths. Then we analyze the effects of different initial conditions on parameter estimation within two data assimilation frameworks (i.e, ensemble Kalman filter and iterative ensemble smoother) and assess several existing model initializing methods that uses available data to retrieve initial soil moisture profile. Our results reveal that Monte-Carlo simulations and the recursive simulation over many years can both demonstrate the temporal behavior of UIC and a common threshold is recommended to determine ~~the warm-up time for both methods.~~t_{wu}. Moreover, the relationship between ~~warm-up time~~t_{wu} for variably saturated flow modeling and the model settings (soil textures, meteorological conditions and soil profile length) are quantitatively identified. In addition, we propose a “warm-up” period before assimilating data in order to obtain a better performance for parameter and state estimation.

Key words: Variably saturated flow; Initialization ~~methods~~method; Initial condition uncertainty; Data assimilation; Soil moisture

1. Introduction

Understanding the movement of soil water is of great importance due to its direct effects across different disciplines, such as ~~in~~ environment, agriculture and hydrology (Doussan et al., 2002). However, ~~the~~ modeling of flow in variably saturated soil is complicated by many difficulties, including highly variable and nonlinear physical processes, as well as limited information about the soil hydraulic properties, initial conditions, and boundary conditions (DeChant, 2014; Rodell et al., 2005; Seck et al., 2014; Bauser et al., 2016; Li et al., 2012). The soil hydraulic parameter uncertainty is identified as ~~one of the~~a major uncertainty ~~sources~~source in vadose zone hydrology and many studies have been focused on this topic. A highly relevant research area, inverse ~~problem~~modeling, has been developed to reduce the uncertainty of parameter by incorporating observational data (Erdal et al., 2014; Montzka et al., 2011; Wu and Margulis, 2013; Wu and Margulis, 2011). ~~Initial and boundary~~Boundary conditions also introduce uncertainty during the simulation of soil water flow (Ataie-Ashtiani et al., 1999; Forsyth et al., 1995; Szomolay, 2008). For instance, the uncertainty ~~of boundary conditions, due to~~introduced by flawed/noise-contaminated meteorological data or fluctuating groundwater table, has been investigated in the past (Freeze, 1969; French et al., 1999; van Genuchten and Parker, 1984; Ji and Unger, 2001; Xie et al., 2011).

Many publications have addressed the issue of the uncertainty of initial condition (UIC) in modeling soil water movement. For example, Walker and Houser (2001) compared the simulation with degraded soil moisture initial condition to that with true initial condition and found the discrepancy did not fade away even after one month. Then, Mumen (2006) concluded that the initial soil water state was one of the most important factors for estimating soil moisture in the case of bare soil. Chanzy et al. (2008) tested three initial water potential profiles and found that initialization had a strong impact on the soil moisture prediction. ~~All these~~These studies showed that the incorrect initial condition may lead to false results. Based on the availability of information, ~~initial condition can be obtained from~~ different initialization approaches can be used for constructing initial conditions, e.g., an arbitrary uniform profile (Chanzy et al., 2008; Das and Mohanty, 2006; Varado et al, 2006), a linear interpolation with in situ observation (Bauser et al., 2016), a steady-state soil moisture profile induced with a constant infiltration flux (Freeze, 1969). All of the approaches involve great uncertainties due to nonlinearity of soil moisture profile, observation error, or inaccurate boundary condition. As a result, it is crucial to explore the effects of UIC on model outputs and compare the uncertainties inherited from various initialization approaches.

~~To minimize~~Besides the ~~bias introduced by initial states and attain more accurate model results in earlier run, a~~simple initialization methods referred above, another common approach~~ of model initialization~~ is to obtain initial condition inherited from the warm-up model with preceding meteorological data. Starting from an arbitrary initial condition, this approach runs the model using a certain period (i.e., warm-up time t_{wu}) of meteorological data~~ (t_{wu})~~ until the model state (e.g., soil moisture) reaches an equilibrium state, which is defined as the state when the uncertainty of state originated from UIC is negligible during simulation. The equilibrium state can be obtained by either running Monte-Carlo simulations until the states from different initial conditions converge to the same value (hereafter referred to as Monte-Carlo method) (Chanzy et al., 2008), or running a single simulation for several years by repeating one-year or multiple-year meteorological condition until the state at an arbitrary date ~~cease~~ceases to vary from year to year (Spin-up method) (Dechant and Moradkhani, 2011; Seck et al., 2014). Spin-up method is widely used in large-scale hydrological model due to its smaller computational cost, while the less-common Monte-Carlo method has the merit of quantifying UIC explicitly at arbitrary time, which can be potentially used to construct state covariance matrix for data assimilation. To the best of our knowledge, there is no comparison made between these two methods to date. Finding an equivalency between these two methods is beneficial for linking initialization methods and data assimilation techniques. Moreover, the determination of warm-up time t_{wu} is ~~key~~crucial to the success of this approach (Ajami et al., 2014; Rahman and Lu, 2015). An underestimation of t_{wu} may bring uncertainty from arbitrarily-specified initial condition prior to initialization, while a large t_{wu}~~ for initialization~~ leads to higher computational demands (Rodell et al., 2005). A variety of modeling settings, such as soil hydraulic properties, meteorological conditions, and soil profile lengths, have strong influences on t_{wu} (Ajami et al., 2014; Cosgrove et al., 2003; Lim et al., 2012a; Walker and Houser, 2001). Thus, the determination of t_{wu} should be investigated thoroughly with different settings.

~~Data assimilation~~ As well as model predictions, UIC also has ~~become a popular tool~~considerable effects on parameter estimation. One of the commonly-used inverse methods in the field of vadose zone hydrology is data assimilation approach (Vereecken et al., 2010; Chirico et al., 2014; Medina et al., 2014a, 2014b). ~~Some~~Previous studies showed that ~~the~~a poor initial soil moisture profile can be corrected by assimilating near-surface measurements (Galantowicz et al., 1999; Walker and Houser, 2001; Das and Mohanty, 2006). Oliver and Chen (2009) discussed several possible approaches to improve the performance of data assimilation through improved sampling of the initial ensemble, and suggested the use of the pseudo-data. Recently, Tran et al. (2013) found that ~~the~~ decreasing assimilation interval could improve the soil moisture profile results induced by wrong initial condition and Bauser et al. (2016) has addressed the importance of UIC in data assimilation framework. ~~Nevertheless, investigation~~However, these preliminary investigations of the influence of UIC on data assimilation results ~~(i.e.,~~ are degraded by the narrow choice of initialization and data assimilation methods, and the lack of comprehensive assessment of the temporal evolution of state/parameter ~~and state estimation) is limited.~~uncertainty when UIC and the parameter uncertainty coexist. For instance, during data assimilation, the initial ensemble ~~are~~is often assumed to be known without uncertainty (Shi et al., 2015) or created by adding Gaussian noise to the initial estimate (Huang et al., 2008), both of which may result in false outputs. ~~Currenlty, related~~The researches mentioned above are all based on a sequential data assimilation approach (i.e., ensemble Kalman filter, or EnKF (Walker and Houser, 2001; Oliver and Chen, 2009)), which incorporates observation in a sequential fashion, so the effect of UIC can be eliminated quickly. Compared to EnKF, an iterative ensemble smoother (IES), which assimilates all data available simultaneously, can obtain reasonably good history-matching results and performs better in strongly nonlinear problems (Chen and Oliver, 2013). However, IES utilize all the observation simultaneously at every iteration and UIC may have a more persistent effect on IES. Thus, ~~it is important to understand~~a systematical analysis for the ~~propagation process~~effects of UIC ~~during variably saturated flow modeling, to identify the warm-up time~~ ~~t_{wu} under a variety of scenarios, and to compare different existing~~and initialization methods within various data assimilation frameworks is necessary and obliged.

The objectives of this paper, therefore, are to: a) compare the temporal evolution of UIC with two common methods (Spin-up method and Monte-Carlo method) and identify the warm-up time t_{wu} under different soil hydraulic parameters, meteorological conditions and soil profile lengths; b) analyze the effects of different initial conditions on parameter estimation during data assimilation with EnKF or IES, and c) propose a selection scheme for choosing a suitable approach of initializing variably saturated flow models within different data assimilation frameworks to minimize the influence of UIC. We first summarize the governing equations of variably saturated flow and method of UIC quantification in Section 2. Then we present results of synthetic simulations designed to investigate the propagation of UIC under different scenarios in Section 3, which is complemented by the results for field data in Section 4. Finally, we present our conclusions in Section 5.

2. Method

2.1 One-dimensional soil water movement

Richards' equation can be used to describe the one-dimensional, vertical soil water movement, which is given as:

$$\cancel{\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K\left(\frac{\partial h}{\partial t}+1\right)\right]} \quad \frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K\left(\frac{\partial h}{\partial t}+1\right)\right] \tag{1}$$

where h [L] represents the pressure head; ~~θ~~ θ [-] denotes volumetric soil moisture; t [T] indicates the time; z [L] is the spatial coordinate taken positive upward; ~~K~~ K [LT^{-1}] denotes the unsaturated hydraulic conductivity. The solution of one-dimensional Richards'~~s~~ equation is numerically solved by a noniterative numerical scheme, which was originally proposed in ~~(Ross, 2003; Ross, 2006).~~Ross (2003) and Ross (2006). By using the primary variable switching scheme, this scheme uses the soil moisture as the unknown variable for unsaturated nodes and pressure head for saturated nodes (Zha et al., 2013). It can greatly reduce the ~~computation~~computational cost of variably saturated flow modeling in soils under atmospheric boundary condition, where alternative drying-wetting conditions are often encountered.

To obtain the solution of Eq. (1), the knowledge of functions K and θ versus h must be required. In this study, we use the van Genuchten-Mualem model (van Genuchten, 1980; Mualem, 1976) to describe ~~the~~these relationships,

$$\cancel{\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{\left[1+|\alpha h|^n\right]^m}} \quad \theta(h) = \theta_r + \frac{\theta_s - \theta_r}{\left[1+|\alpha h|^n\right]^m} \tag{2}$$

$$K(\theta) = K_s S_e^{0.5}[1-(1-S_e^{1/m})^m]^2 \tag{3}$$

where K_s [LT^{-1}] denotes the saturated hydraulic conductivity; ~~θ_s~~ θ_s and ~~θ_r~~ θ_r represent the saturated and residual soil moistures, respectively; parameters ~~α~~ α [L^{-1}] and n are related to the measure of the pore-size density functions and m=1-1/n (n>1); the effective saturation degree ~~S_e~~ S_e is defined as ~~$S_e=(\theta-\theta_r)/(\theta_s-\theta_r)$.~~ $S_e=(\theta-\theta_r)/(\theta_s-\theta_r)$.

Initial and boundary conditions are needed to solve the one-dimensional Richards' equation. The initial condition could be the states of soil moisture

$$\cancel{\theta(z,t)\big|_{t=0}=\theta_0(z)}\ \theta(z,t)\big|_{t=0}=\theta_0(z) \tag{4}$$

where ~~$\theta_0(z)$~~ $\theta_0(z)$ is the initial soil moisture profile.

~~The upper boundary condition is specified as~~ State-dependent, atmospheric boundary ~~(state-dependent boundary~~ condition can be described as (Šimůnek et al., 2013)~~) in this study,~~

$$\cancel{|q|=\left|-K\frac{\partial h}{\partial z}-K\right|\le\left|E_p-P_p\right|}:$$

$$|q|=\left|-K\frac{\partial h}{\partial z}-K\right|\le\left|E_p-P_p\right| \tag{5}$$

$$\cancel{h_m>h>h_c}\ h_m>h>h_c \tag{6}$$

where q [LT^{-1}] is the Darcian flux at the soil surface; E_p [LT^{-1}] denotes the potential evaporation; P_p [LT^{-1}] represents the precipitation intensity; h_m [L] and h_c [L] are maximum and minimum pressure heads allowed at the soil surface, respectively. The value of h_m is set to 0, whereas ~~h_s~~ h_c is ~~deteminedd~~ determined as -100 m.

The bottom boundary condition is the free drainage boundary:

$$\cancel{\frac{\partial h}{\partial z}\Big|_{z=z_N}=0}\ \frac{\partial h}{\partial z}\Big|_{z=z_N}=0 \tag{7}$$

where ~~z_N~~ z_N is the depth of bottom boundary.

2.2 UIC quantification

The investigation of uncertainty in this study includes model states (e.g., soil moisture) and model parameters, where ~~the uncertainty of initial condition (state at t=0)~~ UIC is a special case of state uncertainty~~. We consider two cases in our~~ at t=0. The analysis is twofold. First, we consider a particular situation when

UIC is the only uncertain source and all the model parameters are known. Thus, the choice of initial conditions is solely responsible for the accuracy of the model outputs. In this case, the temporal decay of UIC ~~(either with or without observation)~~ can be clearly demonstrated~~.~~ by utilizing Spin-up or Monte-Carlo methods. Second, a more complex and realistic situation, including both uncertain initial condition and model parameters, is considered during the data assimilation of soil moisture observation. ~~Data assimilation methods, i.e., EnKF and IES, are used to update the model parameters and state simultaneously when observation data are available~~UIC and data assimilation are smoothly combined in our approach since we choose Monte-Carlo-based methods (EnKF and IES). At t=0, we generate an ensemble of soil moisture profiles based on one initialization method (which introduces UIC), and use this ensemble to initiate the data assimilation (assimilate observations and estimate parameter). Finally, we can evaluate our data assimilation performance based on different initializing methods.

2.2.1 The indexes of Spin-up and Monte-Carlo methods

The uncertainty of initial condition can be measured by the percent change PC for Spin-up method (Ajami et al., 2014; Seck et al., 2014) or the ensemble spread S_p for Monte-Carlo method (Reichle and Koster, 2003). ~~Percent change is~~Percent change is an index that reflects the deviation of soil moisture between two adjacent years in a recursive run after a period of warm-up time t_{wu}, which could be calculated as:

$$\cancel{PC = 100\left|\frac{M_1 - M_2}{M_2}\right|} \quad PC(t) = 100\left|\frac{M(t) - M(t+12)}{M(t+12)}\right| \tag{8}$$

where ~~M_1 is~~$M(t)$ and $M(t+12)$ are the monthly ~~mean of~~averaged soil ~~moisture from the previous year and M_2 is the monthly average~~moistures after model spin-up for ~~the current year~~t months and t+12 months (de Goncalves et al., 2006).

The ensemble spread (S_p~~)~~), calculated as a square root of averaged variance over all interested nodes, is an ~~useful~~ index to quantify the difference ~~of~~among various realizations in Monte Carlo simulation, and it is given as:

$$\cancel{S_p(k) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} Var(y_{i,k})}} \quad S_p(k) = \sqrt{\frac{1}{N(N_e-1)}\sum_{i=1}^{N}\sum_{j=1}^{N_e}\left(y^a_{i,j,k} - \langle y^a_{i,k}\rangle\right)^2} \tag{9}$$

~~with~~

$$\cancel{Var(y_{i,k}) = \frac{1}{N_e - 1}\sum_{j=1}^{N_e}(\mathbf{y}^a_{i,j,k} - \langle \mathbf{y}^a_{i,k} \rangle)^2} \qquad \cancel{(10)}$$

where ~~$Var(y_{i,k})$ denotes the nodal ensemble variance at time t_k;~~ ~~$\mathbf{y}^a_{i,j,k}$~~ $y^a_{i,j,k}$ is nodal soil moisture value; ~~$\langle \mathbf{y}^a_{i,k} \rangle$~~ $\langle y^a_{i,k} \rangle$ is the ensemble mean of ~~$\mathbf{y}^a_{i,j,k}$~~ $y^a_{i,j,k}$; i = 1, 2, …, N are the nodes of interest (can be part of the profile); j=1, 2, …, N_e is the ensemble number index; N_e is the ensemble size, which is taken as 300 in this study based on sensitivity analysis of the ensemble size on the calculated results. ~~S_p is the ensemble spread, which can be thought as a square root of averaged variance over all interested nodes.~~ When N = 1, the concept of $S_p(k)$ is equivalent to the standard deviation of ~~$\mathbf{y}^a_k$~~ y^a_k at one location and time t_k.

2.2.2 Data assimilation approaches

We employ EnKF and IES ~~as the~~for data assimilation ~~approaches~~ in this study. Fig. 1 illustrates the basic ideas and differences of the two methods.

EnKF approach was first proposed by Evensen (1994) and has been widely used in variably saturated flow problems (Huang et al., 2008; De Lannoy et al., 2007). This approach is a sequential data assimilation method (as shown in Fig. 1(a)) which ~~incorporate~~incorporates observations into the model in order.

In this ~~work~~part, we assume that hydraulic parameters K_s, α, and n are unknown, while the other parameters θ_r and θ_s are ~~supposed to be~~ deterministic. The vector of parameter and state ~~can be~~is described as,

$$\cancel{\mathbf{y}_k = [\mathbf{m}_k, \mathbf{u}_k]^T} \qquad \cancel{(11)}$$

$$\mathbf{y}_k = [\mathbf{m}_k, \mathbf{u}_k]^T \qquad (10)$$

where ~~$\mathbf{m}_k$~~ $\mathbf{m}_k$ is the parameter vector (i.e., K_s, α, and n), ~~$\mathbf{u}_k$~~ $\mathbf{u}_k$ are state variables (i.e., ~~pressure head and~~ soil moisture) at time t_k, the dimension of y_k is N_y: N_y= N_m+~~2~~ N_d, where N_m indicates the amount of the parameters to be estimated; N_d are the number of nodes of the numerical model. The updated soil moisture ensemble can be converted to pressure head to drive the model. The observation vector can be defined as,

$$\cancel{\mathbf{d}_{j,k} = \mathbf{d}_k + \boldsymbol{\varepsilon}_{jk}} \qquad \text{~~(12)~~}$$

$$\mathbf{d}_{j,k} = \mathbf{d}_k + \boldsymbol{\varepsilon}_{j,k} \qquad (11)$$

where ~~$\mathbf{d}_k$~~ $\mathbf{d}_k$ denotes the observation at time t_k; ~~$\boldsymbol{\varepsilon}_{jk}$~~ $\boldsymbol{\varepsilon}_{j,k}$ (j=1, 2, …, N_e) are independent Gaussian noises added to the observations; ~~$\mathbf{d}_{j,k}$~~ $\mathbf{d}_{j,k}$ is the observation vector for ensemble index j at time t_k. Based on the differences of model forecast and observations, the state-parameter vector can be updated as:

$$\cancel{\mathbf{y}^a_{j,k} = \mathbf{y}^f_{j,k} + \mathbf{K}_k(\mathbf{d}_{j,k} - \mathbf{H}\mathbf{y}^f_{j,k})} \qquad \text{~~(13)~~}$$

$$\mathbf{y}^a_{j,k} = \mathbf{y}^f_{j,k} + \mathbf{K}_k(\mathbf{d}_{j,k} - \mathbf{H}\mathbf{y}^f_{j,k}) \qquad (12)$$

where ~~$\mathbf{y}^f_{j,k}$~~ $\mathbf{y}^f_{j,k}$ denotes the estimated or initially guessed values of parameter and state, while ~~$\mathbf{y}^a_{j,k}$~~ $\mathbf{y}^a_{j,k}$ is the updated estimates; $\mathbf{H}$ is an observation operator, linking the relationship between the state-parameter vector and the observation vector. ~~$\mathbf{K}$~~ $\mathbf{K}$ represents the Kalman gain matrix, which can be calculated as,

$$\mathbf{K}_K = \mathbf{C}^f_k\mathbf{H}^T[\mathbf{H}\mathbf{C}^f_k\mathbf{H}^T + \mathbf{C}_{D_K}]^{-1} \qquad (\cancel{14}13)$$

where ~~$\mathbf{C}_{D_k}$~~ $\mathbf{C}_{D_k}$ indicates the covariance matrix of observed data errors, while ~~$\mathbf{C}^f_k$~~ $\mathbf{C}^f_k$ is the error covariance matrix of forecast ensemble, given by

$$\cancel{\mathbf{C}^f_k \approx \frac{1}{N_e - 1}\sum_{j=1}^{N_e}\left\{\left[\mathbf{y}^f_{j,k} - \left\langle \mathbf{y}^f_k \right\rangle\right]\left[\mathbf{y}^f_{j,k} - \left\langle \mathbf{y}^f_k \right\rangle\right]^T\right\}} \qquad \text{~~(15)~~}$$

$$\mathbf{C}^f_k \approx \frac{1}{N_e - 1}\sum_{j=1}^{N_e}\left\{\left[\mathbf{y}^f_{j,k} - \left\langle \mathbf{y}^f_k \right\rangle\right]\left[\mathbf{y}^f_{j,k} - \left\langle \mathbf{y}^f_k \right\rangle\right]^T\right\} \qquad (14)$$

where ~~$\left\langle \mathbf{y}^f_k \right\rangle$~~ $\left\langle \mathbf{y}^f_k \right\rangle$ is the ensemble mean of ~~$\mathbf{y}^f_k$~~ $\mathbf{y}^f_k$.

Compared to EnKF, IES gives a better estimate by taking all the available observation into consideration (van Leeuwen and Evensen, 1996), as presented in Fig. 1(b). Thus, it can keep the overall consistency of parameters and state variables over time effectively and has been increasingly used to solve the parameter estimation problem in hydrology (Crestani et al., 2013; Emerick and Reynolds, 2013). By calculating iteratively, the nonlinear relationship between observation and parameter is linearized and the information content of the observations can be fully utilized (Chen and Oliver, 2013). In this case, we write the analyzed vector of model parameters $\cancel{\mathbf{m}_j^r}\,\mathbf{m}_j^r$, as

$$\cancel{\mathbf{m}_j^{r+1} = \mathbf{m}_j^r + \mathbf{K}^r(\mathbf{d}_j^r - \mathbf{H}\mathbf{m}_j^r)} \qquad \cancel{(16)}$$

$$\mathbf{m}_j^{r+1} = \mathbf{m}_j^r + \mathbf{K}^r(\mathbf{d}_j^r - \mathbf{H}\mathbf{m}_j^r) \qquad (15)$$

The notation is similar to the one presented for EnKF, where r is the iteration index; $\cancel{\mathbf{m}_j^r}\,\mathbf{m}_j^r$ is the initially guessed or estimated parameters for realization j at iteration r; $\cancel{\mathbf{m}_j^{r+1}}\,\mathbf{m}_j^{r+1}$ is the updated estimates for realization j by conditioning on the observed information at iteration r. It should be noted that the $\cancel{\mathbf{d}_j^r}$ $\mathbf{d}_j^r$ and $\cancel{\mathbf{H}\mathbf{m}_j^r}\,\mathbf{H}\mathbf{m}_j^r$ denotes the total number of ~~observation~~observations and predicted data at iteration r, which is different from EnKF. The Kalman gain $\cancel{\mathbf{K}}\,\mathbf{K}$ is defined as,

$$\mathbf{K}^r = \mathbf{C}_r^f\mathbf{H}^T[\mathbf{H}\mathbf{C}_r^f\mathbf{H}^T + \mathbf{C}_D + \lambda\mathrm{diag}(\mathbf{H}\mathbf{C}_r^f\mathbf{H}^T)]^{-1} \qquad (\cancel{17}16)$$

where $\mathbf{C}_r^f\mathbf{H}^T$ is the cross-covariance matrix between the prior vector of model and the vector of predicted data at iteration r; $\mathbf{H}\mathbf{C}_r^f\mathbf{H}^T$ is the auto-covariance matrix of predicted data at iteration r and $\cancel{\mathbf{C}_D}\,\mathbf{C}_D$ is the covariance matrix of observed data errors. λ donates a dynamic stability multiplier, which is set as 10 initially, and can be adjusted adaptively according to the data misfit at every iteration. diag $(\mathbf{H}\mathbf{C}_r^f\mathbf{H}^T)$ is a diagonal matrix with the same diagonal elements as $\mathbf{H}\mathbf{C}_r^f\mathbf{H}^T$. Mathematically, the dynamic stabilizer term facilitates the solution switching between the Gauss-Newton solution and the steepest-descent method, which is known as the Levenberg-Marquardt approach (Pujol, 2007).

2.3.3 Quantitative index for data assimilation

To assess model parameter and state estimations, root mean square of estimated parameters ($RMSE_m$) and soil moisture ($RMSE_{obs}$), and the relative error index (RE) are computed as follows:

$$RMSE_m = \sqrt{\frac{1}{N_e}\sum_{j=1}^{N_e}(m_j^E - m^T)^2} \qquad (17)$$

$$RMSE_{obs} = \sqrt{\frac{1}{N_{obs}}\sum_{n=1}^{N_{obs}}(d_n^e - d_n^{obs})^2} \qquad (18)$$

$$RE = \frac{RMSE_m^e}{RMSE_m^p} \tag{19}$$

where m_j^E represents the estimated parameter of realization j at the last simulation day (EnKF) or the last iteration (IES); m^T represents the true parameter listed in Table 1. d_n^e and d_n^{obs} indicate the predicted and measured soil moistures, respectively. N_{obs} is the amount of observations. $RMSE_m^e$ and $RMSE_m^p$ represent the *RMSE* of the estimated and prior parameters. *RE* varies from 0 to positive infinity. As *RE* approaches to 0, the analysis result is close to the truth, but a large value of *RE* (more than 1) indicates a bad parameter estimation. Compared with the $RMSE_m$, this index can better present the improvement of parameter estimation during data assimilation.

3. Numerical examples

A series of synthetic numerical experiments are performed in this section. In Section 3.1, we give a general description of the numerical experiments. In order to gain a better understanding of the propagation of the UIC, all the hydraulic parameters (i.e., K_s, α and n) are ~~assumed to be~~ deterministic and the UIC is the only uncertainty source in Section 3.2. Finally, the numerical cases are designed to evaluate performances of data assimilation algorithms ~~initialized by~~combined with various ~~initial conditions~~initialization methods in Section 3.3, in which the parameter uncertainty is taken into consideration in conjunction with UIC.

3.1 General description of model inputs

~~As shown in Table 1, there are five common methods to perscribe initial conditions in variably saturated flow model based on the availablitity of information, including a uniform 50% relative saturation over the entire soil profile (hereafter abbreviated as IC-HfSatu) (Margulis et al., 2002), a linear interpolation between observations at the beginning of simulation (IC-ObsInt) and a steady-state soil moisture profile with a constant infiltration flux (IC-Flux). In this study, the flux is set as 1 mm/d. Besides, we also employ two warm-up methods, which give initial conditions by running the model prior to the beginning of simulation period with available meteorological data (as shown in Fig. 1). If the~~

~~meteorological data before the simulation period is available, it is used in the warm-up method to obtain the initial condition (IC-WUP); otherwise, we use the meteorological data at the simulation period (IC-WUE) as a surrogate. Starting with guessed parameter and initial condition (i.e., we generate two ensembles with guessed means and variances, whereas the means may be biased from the true values), IC-WUP or IC-WUE first warms up the model with available meteorological data, and then uses the ensemble of soil moisture profiles (still uncertain due to the uncertainty of parameters prior to data assimilation, while the bias and uncertainty introduced by state at t_{pre} are reduced) on the last warm-up day (or the beginning of the simulation time t_0) as the initial condition of the formal simulation/data assimilation (see~~ As shown in Table 1, four soils~~Fig.~~ ~~1). The length of warm-up time t_{wu} is determined based on the conclusion in Section 3.2.2(a).~~

~~Four typical types of soil~~ (Sand, Loam, Silt and Clay loam) are chosen in this study to explore the impacts of soil hydraulic property on UIC. ~~Table 2 lists the mean~~The values of ~~uncertain~~ hydraulic parameters ~~(i.e., K_s, α and n)~~are determined according to Carsel and Parrish (1988).

~~To investigate~~ Besides, the ~~impact~~effects of ~~model settings on the temporal evolution of UIC, three different~~ meteorological ~~conditions~~condition are ~~employed.~~also considered: M-AC, M-SC and M-HC in Fig. ~~1~~2 represent three sets of precipitation and potential evaporation data from three different regions (arid region, semi-arid region and humid region) in China.

Unless otherwise specified, a uniform soil profile with the 50% relative saturation (a value of 0.254 for Loam) is chosen as the initial condition (IC-HfSatu). The soil profile is set to be 300~~-~~ cm thick and is filled with Loam. The flow domain is discretized into 60 grids with a grid size of 5 cm~~.~~ which has been proved to be sufficient for evaluating UIC in our study (results not shown). Besides, the total simulation time during the synthetic simulation is one year (365 days). In addition, the default ~~value of the~~ upper ~~boundary condition is~~ and bottom boundaries are set to be M-SC~~.~~ and free drainage boundary, respectively. Other specifications and assumptions for our model simulation runs are given in Table ~~1~~2.

3.2 The temporal evolution of UIC

3.2.1 Comparison of UIC quantification methods

A synthetic experiment is conducted to compare two methods (i.e., Spin-up method and Monte-Carlo method) in quantifying UIC. Using the ~~spin~~Spin-up method, the first case runs a single simulation for 10 years by repeating the preceding meteorological condition starting with IC-HfSatu (Fig. 3(a)), and the percentage cutoff *PC* is calculated. In the second case, the Gaussian noise with a standard deviation of 3% (determined according to the observation error of soil moisture) is added to the IC-HfSatu to generate an ensemble with different initial soil moisture profiles. Then we run ~~the~~different model ~~until the ensemble spread S_p is small~~realizations (Fig. 3(b)). Finally, the *PC* and S_p values of the two cases versus time are compared in Fig. 3(c).

As shown in Fig. 3(a), there is a visible difference between the monthly-averaged soil moistures at the beginning and the 12th months, while the difference is much smaller for θ at the 12th and 24th months, indicating the decay of UIC. Similarly, the soil moistures ~~for~~from different realizations gradually get closer to each other~~, implying the decay of UIC over simulation time~~. As shown in Fig. 3(c), *PC* and S_p values gradually decrease with the simulation time, and their values are approximately the same after t>6 months. The significant difference at the beginning (*PC* of 4.7% and S_p of 2.6%) is due to different initial soil moistures given by the Spin-up and Monte-Carlo methods. The result indicates that the widely-used Spin-up method and Monte-Carlo method are equivalent ~~to one another~~ in terms of quantifying UIC. We will use Monte-Carlo method for the rest of the study since it is consistent with the data assimilation approaches used in this study.

The determination of the threshold value when UIC is regarded to have negligible ~~effect~~effects on ~~prediction~~modeling has been discussed in previous studies (Ajami et al., 2014; Lim et al., 2012; Seck et al., 2014). ~~In general,~~ *PC* or S_p values of 1% (Yang et al., 1995), 0.1% (de Goncalves et al., 2006), or 0.01 % (Henderson-Sellers et al., 1993) have been used~~as threshold for the model. In this study~~. As shown in Fig. 3(c), there is a significant diversity of the results between Spin-up and Monte-Carlo methods at index value of 1%, indicating that UIC still plays a significant role. In contrast, the requested t_{wu} is more than 15 months for a value of 0.1%. To balance the estimation accuracy and computational cost, we recommend a threshold of 0.5% for both Spin-up and Monte-Carlo methods, and the corresponding warm-up time t_{wu} is 8 months, which is sufficiently long for UIC to diminish and the difference between *PC* and S_p ~~to become small~~is insignificant.

3.2.2 The influencing factors on UIC

The Monte-Carlo method is used in this part to obtain the warm-up time t_{wu} and a number of scenarios are constructed under a variety of conditions (different soils, ~~different~~ meteorological conditions, and soil profile lengths). First, the influence of soil texture and meteorological condition on t_{wu} are examined. Four different types of homogeneous soils (Sand, Loam, Silt and Clay loam listed in Table ~~2)~~1) and a heterogeneous soil with multiple layers (consists of Loam (0-75 cm), Clay loam (75-150 cm), Silt (150-225 cm), and Sand (225-300 cm)) under three typical meteorological conditions (M-AC, M-SC and M-HC) are employed in these scenarios, while the other model inputs use the default values (see Table ~~1~~2). Besides, the influence of different soil profile lengths (1 m, 3 m, 5 m, 10 m, 15 m, and 20 m) on UIC is also investigated.

a. The influences of soil texture and meteorological condition

Fig. 4 plots t_{wu} with ~~four~~five different soils under three typical meteorological conditions. The computational times vary greatly according to soil property. We find that t_{wu} of Sand are all less than one day, whereas t_{wu} of Loam are 412 days, 242 days, and 195 days respectively. In addition, the warm-up times of Silt and Clay loam with M-AC and M-SC exceed 10 years, while those with M-HC are 264 days and 253 days. The results imply that the warm-up time t_{wu} for the fine-textured soil is larger than that for coarse-textured soil. ~~To further explore the reason behind, results of soil moisture profiles at the given days (e.g., days 1st and 30th) for four kinds of soils are manually inspected (results not shown).~~This may attribute to the diversity of the drainage property for different soils. For Sand, due to its fast drainage property, the soil moisture ensemble converges extremely quickly and most of the values at the profile are maintained as residual soil moisture. Thus, the UIC of ~~sand~~Sand disappears very fast. ~~On the other hand~~In contrast, the soil moisture ~~results~~states for Silt and Clay loam ~~show slow convergence. For Loam, the convergence speed is smaller~~change more slowly than ~~that of Sand but larger than that of Clay loam.~~Sand during the simulation. Therefore, faster drainage property leads to a smaller warm-up time.

In addition, the meteorological condition has a strong impact on UIC ~~and t_{wu} decreases with the increase of precipitation.~~. For example, with soil Loam, the order of t_{wu} is M-HC<M-SC<M-AC. ~~Regarding~~For Silt and Clay loam, t_{wu} of M-AC and M-SC decrease from more than 10 years to 264 days and 253 days with a humid climate M-HC, respectively. With intensive and excessive rainfall events, θ approaches to the saturated soil moisture, leading to a sudden drop of S_p. Thus, the meteorological condition, especially the precipitation, plays an important role in the propagation of UIC. ~~Excessive and~~

~~intensive rainfall could greatly reduce the impact on simulation solely induced by UIC (i.e., no parameter uncertainty).~~Moreover, regarding the heterogeneous soil with multiple layers, the t_{wu} under the M-AC is larger than 10 years (similar to Silt and Clay loam), while that under M-SC or M-HC becomes much smaller (higher than that of Loam but they are of the same magnitude). Thus, it is conjectured that t_{wu} is determined by the fine soil texture in the layered profile under dry meteorological condition, but averaged soil hydraulic properties under wet meteorological condition.

It should be noted that the t_{wu} is also relevant to the initial state of soil. Regarding the initial condition in an extremely dry state under the arid climate, the hydraulic conductivity is very small, and the initial spread extends for a long time. For instance, t_{wu} of sand increases from 1 day to 8 days when the ensemble mean value of initial soil moisture decreases from 0.2375 to 0.15 (results not shown). Yet, if a sufficiently large rain event takes place, the soil moisture increases and then converges to a similar state rapidly.

b. The influence of soil profile length

To investigate the effects of soil profile length on warm-up time, we investigate the t_{wu} values for simulations with various soil profile lengths. As presented in Fig. 5~~,~~(a), the t_{wu} for soil lengths of 1 m, 3 m, 5 m, 10 m, 15 m and 20 m are 0.11 year, 0.57 year, 0.74 year, 1.57 years, 2.78 years and 4.3 years respectively, indicating that the warm-up time increases with increasing depth of soil column. Fig. ~~This can be attributed to a larger portion of soil is not affected by boundary conditions and UIC decays more slowly.~~5(b) plots the t_{wu} value for each depth with the profile length of 20 m, showing that a longer warm-up time is needed if the soil layer is deeper. Both subfigures imply that UIC decays more slowly if the effects of boundary condition become less important. We also examine the case for substituting free drainage boundary for a prescribed groundwater table. The results indicate that the t_{wu} is further shortened due to the ~~restriction~~influence of bottom saturation condition (not shown).

In addition, t_{wu} in homogeneous loam reveals a power law relationship with the length of soil profile. According to the fitted curve in Fig. 5~~,~~(a), the warm-up time t_{wu} is more than seven years for a depth d of 30 m (e.g., North China Plain, (Huo et al., 2014)) and 700 years for d=1000 m (e.g., Yucca Mountain Site, (Flint et al., 2001)) with loam soil. This result suggests that we should be very careful to deal with simulation with a long unsaturated profile, where the UIC ~~can exist~~lasts for an extremely long time and influence the simulation/data assimilation results.

3.3. Initialization of data assimilation

Besides IC-HfSatu, two other common methods to prescribe initial conditions in variably saturated flow model based on the availability of information are also considered in this study, including a linear interpolation between observations (at depths of 10 cm, 80 cm, 150 cm, 220 cm and 290 cm) at the beginning of simulation (IC-ObsInt) and a steady-state soil moisture profile by warming up the model with a constant infiltration flux of 1 mm/d (IC-Flux). Moreover, we employ two warm-up methods, which give initial conditions by running the model prior to the beginning of simulation period with available meteorological data (as shown in Fig. 2). If the previous meteorological data before the simulation period is available, it is used in the warm-up method (IC-WUP); otherwise, we use the meteorological data at the experimental period as a surrogate (IC-WUE). The length of warm-up time for IC-Flux, IC-WUP and IC-WUE is equal to t_{wu} (242 days) based on the results in Section 3.2.2(a), so the warming-up period of WUP for these three methods is from day 124 to day 365. In addition, IC-HfSatu and IC-ObsInt are assumed to be deterministic without uncertainty, while for the IC-Flux, IC-WUP and IC-WUE, the uncertainty of states are introduced by warming up the model with uncertain parameters.

Thus, a total of~~**3.3. Initialization of data assimilation**~~

~~In this sub-section,~~ five initialization methods (IC-HfSatu, IC-ObsInt, IC-NetFlux, IC-WUP and IC-WUE) are assessed to investigate the effect of ~~initial condition~~UIC on model state and parameter estimations within two data assimilation frameworks (EnKF and IES). The initial realizations of soil hydraulic parameters K_s, α and n for all data assimilation models as well as the warming-up models IC-Flux, IC-WUP and IC-WUE are generated following logarithm normal distributions, with mean values of 4.7 md^{-1}, 8.6 m^{-1} and 1.8, and variances (log-transformed) of 0.1, 0.3 and 0.006. The saturated soil moisture θ_s and residual soil moisture θ_r are assumed to be deterministic with the value of 0.43 and 0.078. Compared with the reference values (K_s, α and n for Loam are 0.2496 md^{-1}, 3.6 m^{-1} and 1.56) listed in Table ~~2~~1, the prior means of unknown parameters are biased.

3.3.1 General description ~~of the cases~~ for various ~~initialization methods~~data assimilation cases

Several test cases are conducted to explore the effects of initialization on parameter estimation under various data assimilation frameworks. ~~Cases~~Case 1 ~~and 2 investigate~~investigates the effects of five initialization methods ~~(Table 1)~~ on individual parameter estimation with EnKF and IES, respectively.

Then, we increase the ensemble size of IC-HfSatu and IC-ObsInt to 500 (hereafter referred to as IC-HfSatu-500 and IC-ObsInt-500) in Case 2 to demonstrate the impacts of ensemble size. Case 3 explores the effects of the uncertainty magnitude of the initial ensemble on the parameter estimations. A Gaussian noise with a standard deviation of 0.017 (counted from IC-WUP) is added to both IC-HfSatu-500 and IC-ObsInt-500 (hereafter referred to as IC-HfSatu-500-Un and IC-ObsInt-500-Un). Furthermore, to find out the role of initial condition in multi-parameter inverse problems, Case ~~3 and Case~~ 4 ~~are~~is conducted to estimate K_s, α and n simultaneously. ~~To~~Case 5 is implemented with a simulation time of 60 days to explore the ~~impact~~influence of assimilation time on multiple parameter estimation with IES~~, Case 5 is implemented with a simulation time of 60 days.~~. It should be noted that the warm-up methods (IC-WUP and IC-WUE) used in IES ~~warms~~warm up model before every iteration (as presented in Fig. 1(b)), since the initialization of IES by warming up the model for only the first iteration leads to poor assimilation results.

The synthetic observations used for data assimilation are generated by running the model with “true” parameter (Loam) and “true” initial condition (produced by warming up model with a sufficient long ~~t_{wu} =~~time of 10 years). The generated observations are perturbed by ~~observation errors (~~a Gaussian noise with a standard deviation of ~~1%), which~~0.01. A total number of 37 observations are ~~assumed to be Gaussian. In addition, the observation at 10 cm is~~ assimilated into the model. The observation depth is at z = 10 cm and the observed soil moisture is assimilated every 10 days~~.~~, starting from day 3. The details of the model inputs for Case 1 to Case 5 are listed in Table 3.

3.3.2 ~~Quantitative index~~

~~To quantify the quality of model parameter and state estimations, root mean square of estimated parameters ($RMSE_m$) and soil moisture ($RMSE_{obs}$) are computed as follows:~~

$$RMSE_m = \sqrt{\frac{1}{N_e}\sum_{j=1}^{N_e}(P_j^E - P^T)^2} \tag{18}$$

$$RMSE_{obs} = \sqrt{\frac{1}{N_{obs}}\sum_{n=1}^{N_{obs}}(x_n - x_n^{obs})^2} \tag{19}$$

~~where P_j^E represents the parameter estimation of realization j on the last simulation day; P^T represents~~

~~the true parameter listed in Table 2. N_e is the total number of realizations. x_n and x_n^{obs} indicate the predicted and measured soil moistures, respectively. N_{obs} is the amount of observations.~~

~~3.3.3~~ Result ~~comparison~~

The results for parameter estimation (lnK_s) using the two data assimilation frameworks ~~and under the various~~with different initialization methods (Case 1) are compared in Fig. 6. In Fig. 6(a), the estimated lnK_s values of EnKF are presented. In general, the lnK_s estimations under different initial conditions all gradually approach the "true" values ~~with~~over assimilation time, but the final assimilation results are different. For IC-HfSatu, because the initial profile is uniform and arbitrarily specified, the assimilation results are ~~both~~ affected by the parameter uncertainty and UIC simultaneously. Thus, the decreasing of $RMSE_{\mathrm{m}}$ is the slowest and the final parameter estimation result is the worst. In contrast, the initial conditions generated by warm-up methods (IC-WUP and IC-WUE) can eliminate the UIC in advance, and thus data assimilation can handle parameter uncertainty more efficiently, leading to the best results among the five. The data assimilation results of IC-WUE are a little worse than those of IC-WUP owing to the diversity of meteorological condition. Since IC-ObsInt and IC-Flux are created by adding observation information or simple infiltration information, they perform better than that with IC-HfSatu but worse than warm-up methods. Similarly, the assimilation results for IES with IC-WUP ~~and IC-WUE~~ are also the best, while those with IC-HfSatu have the worst parameter estimation in the five initialization methods (Fig. 6(b)). ~~Moreover, the parameter estimations with IC-WUP and IC-WUE require much fewer iteration steps (at about 7th iterations) than the other methods.~~ In addition, by comparing Figs. 6(a) and 6(b), the cases using IES shows ~~superior~~better results than those using EnKF, indicating a ~~better~~superior ability for IES to estimate individual parameter in variably saturated model. However, since IES estimates parameter iteratively, it has a much larger computational cost than EnKF when using warm-up methods.

~~An important issue in~~For data assimilation ~~based on Monte Carlo method~~ problem, the ensemble variance is increasingly underestimated over time/iteration, which may cause the filter inbreeding problem (Hendricks Franssen and Kinzelbach, 2008~~), which underestimates the ensemble variance over time/iteration, which can lead to poor performance of parameter updating.~~). To determine if our data assimilation runs are affected by filter inbreeding, the temporal change of the standard deviation of estimated lnK_s are plotted in Figs. 6(c) and 6(d). In general, the standard deviation of estimated lnK_s decline gradually with assimilation steps (EnKF) or iteration steps (~~ES), which agree well with the differences between the estimates and the true values (Figs. 6(a) and 6(b)). Thus, filter inbreeding problem is not significant here, and the data assimilation results are reliable.~~IES). As given in Figs. 6(a) and 6(c), the filter inbreeding might take place after 280th days for EnKF, since the standard deviation of ensemble all approach to 0.1 and the estimated parameters stay constant over time. However, with the help of a damping parameter, the filter inbreeding problem for IES could be reduced significantly. This partly explains the inferior result of EnKF compared to IES.

Increasing the ensemble size and model uncertainty is an efficient approach to reduce the filter inbreeding (Hendricks Franssen and Kinzelbach, 2008). To demonstrate the impacts of ensemble size and initial uncertainty on data assimilation results, the results of lnK_s estimations utilizing the initial condition IC-HfSatu and IC-ObsInt with the ensemble size of 500 (Case 2) and a Gaussian noise (Case 3) are plotted in the Fig. 7.

The results of IC-HfSatu-500 and IC-ObsInt-500 with the ensemble size of 500 in Fig. 7 are similar with those of IC-HfSatu and IC-ObsInt (Fig. 6), indicating that the improvement of the parameter estimation result is slight when the ensemble size increases from 300 to 500. Hence, the ensemble size of 300 is sufficient for data assimilation problem in this study. In contrast, the influences of adding the uncertainty to the initial state on parameter estimation are totally different for EnKF and IES. Compared with the results of IC-ObsInt-500 and IC-HfSatu-500, the results of lnK_s estimation with IC-ObsInt-500-Un and IC-HfSatu-500-Un improve for EnKF (Fig. 7(a)), but deteriorate for IES (Fig. 7(b)). This may attribute to the diversity between two algorithms. EnKF is a sequential algorithm, so the state uncertainty introduced by UIC could decrease over assimilation steps. A larger ensemble state variance implemented at the beginning leads to a larger trust on data and thus a quicker update of parameter to truth, and can prevent EnKF from inbreeding, leading to a better result than that with initial condition of small variance. On the contrary, IES is a batch optimization method. The uncertainty of initial state exists at each iteration and has a negative effect on the model calibration during the whole simulation, worsening the parameter estimation results.

Moreover, the parameter estimation results of IC-WUP are still superior to those of IC-HfSatu-500-

Un and IC-ObsInt-500-Un even they all have a similar computational cost, showing the promising performance of warm-up methods. The results are reasonable since all ensemble Kalman filter methods are affected by the quality of the auto-covariance matrix and the mean value of predicted state ensemble (Eqs. (12) and (13) for EnKF; Eqs. (15) and (16) for IES). For WUP method, the initial condition is constructed by warming up the model with the prior parameter, thus IC-WUP contains useful information of prior parameter, even it is biased. Besides, the state covariance matrix is implicitly inflated due to the introduction of uncertain prior parameter ensemble during warming up. These two aspects ensure the robust performance of warm-up methods. However, the initial state ensembles of IC-HfSatu-500-Un and IC-ObsInt-500-Un are independent from the prior parameter, which introduces additional uncertainties, making the data assimilation results worse. Therefore, even using a larger ensemble size and enlarging the state uncertainty (covariance inflation), warm-up methods are still the optimal choice for both EnKF and IES algorithms. We also construct another case with a larger parameter uncertainty to alleviate filter inbreeding problem and the data assimilation for all cases are improved (not shown). The results also agree with our conclusion that WUP performs the best among the five initialization methods.

To evaluate the effects of UIC in multi-parameter inverse problem, the ~~$RMSE_m$~~RE results of K_s, α, and n estimates of Case ~~3 and Case~~ 4 are presented in Fig. ~~7~~8. In general, the ~~$RMSE_m$~~RE results of n and K_s are ~~the smallest~~small no matter using EnKF or IES, while the ~~$RMSE_m$~~RE of α is the largest. A cross-correlation analysis indicates that soil moisture observations are~~ very~~ insensitive to parameter α with a free drainage boundary condition, which agrees with the results of Hu et al., (2017). In Fig. ~~7~~8(a), similar to the conclusion of one-parameter inverse problem, the parameter estimation results of K_s and α with IC-HfSatu and IC-ObsInt are worse than those of IC-WUP and IC-WUE. There is not much difference between the n estimates under various initial conditions, implying that n is less affected by UIC when estimating K_s, α and n simultaneously. Compared with EnKF, IES shows a smaller ~~$RMSE_m$~~RE (Fig. ~~7~~8(b)) ~~of below 5~~ for all parameters, indicating IES can also perform better in multi-parameter inverse problem. However, the assimilation results with various initialization methods do not show much difference, implying that the final ~~$RMSE_m$~~RE values are not significantly affected by UIC, possibly due to abundant observations available over one year. ~~However~~Nevertheless, long-term observation data may not be available in many cases.

To examine the impact of assimilation time on parameter estimation with IES, ~~a case~~Case 5 with shorter assimilation period (60 days) and a fewer number of observations (i.e., 6) is conducted. Fig. ~~8~~9 shows the ~~$RMSE_m$~~RE results and it is inferior to ~~than~~those in ~~-~~ Case 4, where the simulation time is one year (Fig. 8(b)). Nevertheless, the effects of assimilation time on parameter estimation are different for different parameters. For instance, parameter n can still be estimated ~~quite~~ well in ~~this case.~~the most of the situations. In addition, though the assimilation results of K_s degraded with a 60-~~days~~day simulation, the variation of K_s estimation values among different initialization methods is small. The number of observation can greatly affect the estimation of parameter α, since ~~$RMSE_m$~~RE of α in Case 5 (~~25.64, 35.06,~~ 3.5, 4.8~~.52, 5.76~~, 1.17, 0.79, and ~~5.76~~0.23) is much larger than those in Case 4 (~~1.19, 2.12, 4.19, 2.81~~0.16, 0.29, 0.68, 0.24, and ~~2.39).~~0.31). Furthermore, the warm-up methods show the best data assimilation results among the five when the observations are insufficient.

4. ~~Filed~~Field validation

~~In order to examine the applicability of the conclusions drawn from synthetic case, a~~Synthetic observation in previous section is generated by running the model with exactly known uncertainty sources. By conducting synthetic experiments, we can thoroughly analyze the impact of UIC during data assimilation, with scenarios having different numbers of observations/unknown parameters, and more decisive conclusions can be drawn. In contrast, the field observations contain additional uncertainties which are largely unknown (e.g., the calculated evapotranspiration is inaccurate for real-world case). In order to examine the real-world applicability of the conclusions drawn from synthetic case, Field data are necessary to validate our results. A field experiment is conducted in the irrigation-drainage experimental site of Wuhan University (Li et al., 2018) (Fig. ~~9~~10(a)). Meteorological data, including air temperature, relative humidity, atmospheric pressure, incident solar radiation, and precipitation, is continuously monitored by an automatic weather station (LoggerNet 4.0), which can be used as upper boundary condition after the calculation of the potential evaporation (Penman-Monteith's equation) on the bare soil (see Fig. ~~10~~11(a)). A vertically-inserted frequency domain reflectometry (FDR) tube was used to monitor soil moisture (Fig. ~~9~~10(b)). The in-situ soil moisture observation was measured every 3 days. The tube gave soil moisture measurements at the depths of 10, 20 and 30 cm. During 18th April 2017 to 30th May 2017, the measurements were repeated 14 times and 42 soil moisture data were collected (see Fig.

~~10~~11(b)). Besides, the soil moisture at the depth of 10 cm, 20 cm, 30 cm, 40 cm, 60 cm and 80 cm at the beginning of the simulation time is also available to construct an initial profile via ~~interpolation with~~ IC-ObsInt.

4.1 General description of the experimental case

To analyze the experimental data, the 1-D numerical domain is set as 2 m and discretized in 50 grids. The top 40 grids have a size of 2.5 cm and the rest has a size of 10 cm. The upper boundary is set as an atmospheric boundary using the data shown in Fig ~~10~~11(a) and the bottom boundary is set to be free drainage since the groundwater table is much deeper than the bottom of the domain.

The prior parameter distributions follows the study of Li et al. (2018). The saturated soil moisture $\cancel{\theta_s}$ θ_s and residual soil moisture $\cancel{\theta_r}$ θ_r are given as 0.43 and 0.078, while the other hydraulic parameters are to be estimated. The initial means of K_s, α and n are set as 1 md^{-1}, 5 m^{-1} and 2, and the initial natural logarithmic variances ~~of the natural logarithm~~ of them are set as 0.22, 0.16 and 0.003. The data from 18th April through 27th April are used for calibration, while the remaining data are reserved for validation.

~~In addition, according to our prior knowledge, the meteorology~~ The climate of Wuhan is semi-arid conditions and the soil of experimental site is sandy loam. We use a warm-up time of 242 days based on our investigation in Section 3.2.2.

4.2 Results

The assimilation results with four different initialization results (IC-HfSatu, IC-ObsInt, IC-Flux and IC-WUP) are presented in this part. Since the true hydraulic parameters at the experimental site are unknown, we assess the estimation by comparing the predicted (using estimated parameters) and observed soil moistures during the validation period. The $RMSE_{obs}$ for soil moisture predictions under different assimilation scenarios are listed in Table 4. Generally speaking, $RMSE_{obs}$ with IC-WUP are again the smallest, while IC-HfSatu has the largest $RMSE_{obs}$ values.

In order to evaluate the overall performances of the four initialization methods, the soil moisture observations and predictions at all depths are plotted in Fig. ~~11~~12. The coefficients of determination under the four scenarios are 0.033, 0.599, 0.083 and 0.553, and the $RMSE_{obs}$ are 0.045, 0.037, 0.036, and 0.028 respectively. As shown in Fig. ~~11~~12(a) and Fig. ~~11~~12(c), IC-HfSatu and IC-Flux show very large scattering, indicating a bad prediction performance. A significant improvement is found in IC-WUP with a large R^2 and the smallest $RMSE_{obs}$ value, as shown in Fig. ~~11~~12(d). Surprisingly, IC-ObsInt has the largest R^2 among the four ~~in this case~~methods, though its $RMSE_{obs}$ value is bigger than that of IC-WUP. The simulation of real-world problems may have uncertainties that are not considered in data assimilation. For instance, the meteorological data prior to the simulation for warming up ~~may not be accurate.~~is not precise from the weather-station instrument error and calculation of evapotranspiration, which has a detrimental effect on IC-WUP. IC-ObsInt, on the other hand, ~~consistently utilize~~takes the advantage that it utilizes the soil moisture observations for both initialization and predictions. ~~It should be noted that~~ However, IC-ObsInt may not be applicable ~~in the case~~ when soil moisture observations at t=0 are not available or the soil moisture profile ~~has large variation, e.g.,~~is discontinuous ~~soil moisture~~ in layered soils~~.~~, leading to a large interpolation error. In summary, ~~as also demonstrated in the numerical case studies (Section 3.3), the model with initial condition~~for both the synthetic and field cases, models initialized using the warm-up method ~~results~~result in low uncertainty and superior soil moisture predictions even if the calibration data are insufficient.

5. Discussion and Conclusions

The study investigates the effects of UIC on variably saturated flow simulations subject to different soil hydraulic parameters, meteorological conditions and soil profile lengths. Two common approaches (Spin-up and Monte-Carlo methods) are applied to explore the required warm-up time t_{wu} and temporal behavior of UIC. In addition, the data assimilation performances with five common initialization approaches are compared using synthetic experiments and a field soil moisture dataset.

Under atmospheric boundary condition, the soil moisture value near the upper boundary could approach its upper and lower bounds (saturated water content and residual water content) due to rainfall and evaporation. This ~~will~~ significantly ~~reduce~~reduces the UIC of soil moisture profile near the soil surface. ~~Moreover, the wetter (drier) soil drains more (less) water and evaporates more (less) water, making UIC of soil moisture dissipates rapidly.~~ Our investigation ~~of the influence of soil texture and boundary condition on UIC~~ shows~~, as expected,~~ that the coarse-textured soil results in faster reduction of soil moisture UIC because of fast redistribution of water in sandy soil. Regarding the influence of boundary conditions, we ~~found~~find that heavy rainfall ~~and long-term evaporation~~ can reduce UIC significantly, while an initial condition in a drier status leads to a growth of t_{wu}, since a drier soil drains and evaporates less water, making UIC of soil moisture dissipates slowly. The conclusion agrees with the conclusions reported by Castillo et al., (2003) and Seck et al., (2014). Although t_{wu} for sandy soil is very small, ~~they~~it could be very large for other soils (less than one day versus more than 10 years in Fig. 4). The length of soil profile plays an important role in UIC since UIC decays from the boundaries. As a result,~~ the~~ UIC could exist persistently in a very thick vadose zone. Our findings imply that UIC dissipation depends nonlinearly on soil type, meteorological condition, and soil profile lengths~~ care~~, and special attention should be ~~taken when handling UIC in~~paid to during vadose zone modeling.

Ideally, the initial ensemble should represent the error statistics of the initial guess for the model state during data assimilation (Evensen, 2003). Thus, effort should be invested to reduce the impact of UIC on data assimilation. Methods which do not consider the UIC (i.e., assuming an initial ensemble arbitrarily without uncertainty, which was used in some studies, e.g., Shi et al., 2015) can induce significant bias according to our data assimilation results. By constructing initial condition using the information of observations or boundary condition (averaged flux), the data assimilation results can be improved. However, these two initialization methods are also suboptimal, due to ~~simplification of~~the oversimplification to the complex initial condition. By warming up model with available meteorological data, the initialization methods can improve data assimilation results. Moreover, EnKF is more sensitive to filter inbreeding problem than IES. The initial condition with a larger state uncertainty gains better performance than that without covariance inflation for EnKF. While for IES, this inflated uncertainty cannot decrease over iterations, making the results inferior.

In this study, we only use the soil moisture observations rather than pressure head to construct the initial profile. For homogeneous soil column, there is a one-to-one relationship between the spread of soil moisture and pressure head (i.e., UIC in terms of pressure head can be converted from that of soil moisture). The situation will be much more complex if the soil is heterogeneous, since a large number of unknown hydraulic parameters may introduce significant nonlinearity during the transformation between head and soil moisture. For instance, the soil moisture profile is discontinuous in layered soils. The use of pressure head instead of soil moisture as initial condition for heterogeneous soils deserves investigation in our future work.

Our work leads to the following major conclusions:

1. Spin-up method and Monte-Carlo method can both quantify UIC and they agree well with each other after a sufficiently long simulation. A threshold of 0.5% for percentage cutoff *PC* or ensemble spread S_p is recommended to determine the warm-up time.

2. Warm-up time varies nonlinearly with soil textures, meteorological conditions, and soil profile length. Under most situations (e.g., Loam with the soil profile length less than 5 m under non-arid climate), one-year warm-up time is sufficient for soil water movement modeling, but an extremely long time (exceeds 10 year) is needed to warm up the model for a long, fine-textured soil profile under an arid meteorological condition.

3. IES shows better performance than EnKF in the strongly ~~non-linear~~nonlinear problem and is affected less by the UIC ~~if warm-up method is implemented at~~with a long-period of observations. In addition, covariance inflation of initial condition could improve the ~~beginning of the simulation~~data assimilation results for ~~every iteration. For both algorithms, the estimation of α is the most difficult while the parameter n can be estimated more easily in the multi-parameter inverse problem.~~EnKF, but deteriorate them for IES.

4. The following procedure is recommended to initialize soil water ~~model if meteorological data are available~~modeling: 1) Evaluate the approximate warm-up time based on the model settings; 2) Initialize the model using the method of WUP (if meteorological data are available) and make sure the warming up time is larger than the required t_{wu}; 3) Run the simulation with the initial condition obtained in step 2. ~~If~~WUE is an alternative to WUP if the preceding meteorological data are not available~~, WUE is an alternative to obtain initial condition~~. ObsInt is also a practical choice if dense soil moisture observations at the beginning of simulation are available.

Further research may examine the performance of these initialization methods in two- or three-dimensional variably saturated flow conditions~~ and for large-scale problems.~~. Our approach can also be ~~applied~~extended to ~~models with multiple~~other modeling and data assimilation problems in other disciplines (e.g., groundwater flow and solute transport modeling, and soil~~ layers for the parameter estimation and to identify the warm-up time.~~-water-crop modeling).

Data/code availability. All the data used in this study can be requested by email to the corresponding author Yuanyuan Zha at zhayuan87@gmail.com.

Author contribution: ~~Danyang Yu,~~ Yuanyuan Zha and Jinzhong Yang developed the new package for soil water ~~movement~~flow modeling based ~~on a~~ switching the primary variable of numerical Richards' equation; Danyang Yu and Yuanyuan Zha developed ~~the~~ EnKF and IES codes for data assimilation ~~methods of EnKF and IES~~of variably saturated flow, and designed and conducted the numerical cases and field data validation for this study. Seven of the co-authors made non-negligible efforts preparing the manuscript.

Competing interests: The authors declare that they have no conflict of interest.

~~Acknowledgements.~~ Acknowledgments

This work is supported by Natural Science Foundation of China through grants No. 51609173~~,~~ and 51779179~~, 51479144, 51522904~~. The authors appreciate Michael Tso (Lancaster University, U.K.) for editing the manuscript.

We thank the three reviewers for their constructive comments and suggestions.

References

Ajami, H., McCabe, M. F., Evans, J. P., and Stisen, S: Assessing the impact of model warm-up on surface water-groundwater interactions using an integrated hydrologic model. Water Resour. Res., 50, 1–21, doi:10.1002/2013WR014258, 2014.

Ataie-Ashtiani, B., Volker, R. E. and Lockington, D. A.: Numerical and experimental study of seepage in unconfined aquifers with a periodic boundary condition, J. Hydrol., 222(1-4), 165–184, doi:10.1016/S0022-1694(99)00105-5, 1999.

Bauser, H. H., Jaumann, S., Berg, D., and Roth, K.: EnKF with closed-eye period - Towards a consistent aggregation of information in soil hydrology, Hydrol. Earth Syst. Sci., 20(12), 4999–5014, doi:10.5194/hess-20-4999-2016, 2016.

Carsel, R. F. and Parrish, R. S.: Developing joint probability distributions of soil water retention characteristics, Water Resour. Res., 24(5), 755–769, doi:10.1029/WR024i005p00755, 1988.

Castillo, V. M., Gómez-Plaza, A. and Martínez-Mena, M.: The role of antecedent soil water content in the runoff response of semiarid catchments: A simulation approach, J. Hydrol., 284(1-4), 114–130, doi:10.1016/S0022-1694(03)00264-6, 2003.

Chanzy, A., Mumen, M. and Richard, G.: Accuracy of top soil moisture simulation using a mechanistic model with limited soil characterization, Water Resour. Res., 44(3), 1–16, doi:10.1029/2006WR005765, 2008.

Chen, Y. and Oliver, D. S.: Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification, Comput. Geosci., 17(4), 689–703, doi:10.1007/s10596-013-9351-5, 2013.

Chirico, G. B., 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., 18(7), 2503–2520, doi:10.5194/hess-18-2503-2014, 2014.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Sheffield, J., Duan, Q. Y., Luo, L. F., Higgins, R. W., Pinker, R. T. and Tarpley, J. D.: Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS), J. Geophys. Res., 108, D22, doi:10.1029/2002JD003316, 2003.

Crestani, E., Camporese, M., Baú, D. and Salandin, P.: Ensemble Kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation, Hydrol. Earth Syst. Sci., 17(4), 1517–1531, doi:10.5194/hess-17-1517-2013, 2013.

Das, N. N. and Mohanty, B. P.: Root Zone Soil Moisture Assessment Using Remote Sensing and Vadose Zone Modeling, Vadose Zo. J., 5(1), 296, doi:10.2136/vzj2005.0033, 2006.

DeChant, C. M.: Quantifying the Impacts of Initial Condition and Model Uncertainty on Hydrological Forecasts. Ph.D thesis, Portland State University, Portland, United States, 2014.

Dechant, C. M. and Moradkhani, H.: Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation, Hydrol. Earth Syst. Sci., 15(11), 3399–3410, doi:10.5194/hess-15-3399-2011, 2011.

~~de~~De Goncalves, L. G. G., Shuttleworth, W. J., Burke, E. J., Houser, P., Toll, D. L., Rodell, M., and Arsenault, K.: Toward a South America Land Data Assimilation System: Aspects of land surface model warm-up using the Simplified Simple Biosphere, J. Geophys. Res. Atmos., 111(17), 1–13, doi:10.1029/2005JD006297, 2006.

De Lannoy, G. J., Reichle, R. H., Houser, P. R., Pauwels, V., and Verhoest, N. E.: Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. Water Resour. Res., 43(9). doi:10.1029/2006WR005449, 2007.

Doussan, C., Jouniaux, L. and Thony, J. L.: Variations of self-potential and unsatureated water flow with time in sandy loam and clay loam soils, J. Hydrol., 267, 173–185, doi:10.1016/~~S0022-1694~~S00221694(02)00148-8, 2002

Emerick, A. A. and Reynolds, A. C.: Ensemble smoother with multiple data assimilation, Comput. Geosci., 55, 3–15, doi:10.1016/j.cageo.2012.03.011, 2013.

Erdal, D., Neuweiler, I. and Wollschläger, U.: Using a bias aware EnKF to account for unresolved structure in an unsaturated zone model, Water Resour. Res., 50(1), 132–147, doi:10.1002/2012WR013443, 2014.

Evensen, G.: The Ensemble Kalman Filter: Theoretical formulation and practical implementation, Ocean Dyn., 53(4), 343–367, doi:10.1007/s10236-003-0036-9, 2003.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res. Oceans., 99(C5), 10143-10162. doi:10.1029/94JC00572, 1994.

Flint, A L., Flint, L. E., Bodvarsson, G. S., Kwicklis, E. M. and Fabryka-Martin, J.: Development of the conceptual model of unsaturated zone hydrology at Yucca Mountain, Nevada, Concept. Model. Flow Transp. Fract. Vadose Zo., ~~247,~~ 47–85, doi:10.1016/S0022-1694(01)00358-4, 2001.

Forsyth, P. A., Wu, Y. S. and Pruess, K.: Robust numerical methods for saturated-unsaturated flow with dry initial conditions in heterogeneous media, Adv. Water Resour., 18(1), 25–38, doi:10.1016/0309-1708(95)00020-J, 1995.

Freeze, R. A.: The Mechanism of Natural Ground~~ - ~~-Water Recharge and Discharge: 1. One~~ - ~~-dimensional, Vertical, Unsteady, Unsaturated Flow above a Recharging or Discharging Ground~~ - ~~-Water Flow System, Water Resour. Res., 5(1), 153 – 171, doi:10.1029/WR005i001p00153, 1969.

French, H. K., Van Der Zee, S. E. A. T. M. and Leijnse, A.: Differences in gravity-dominated unsaturated flow during autumn rains and snowmelt, Hydrol. Process., 13(17), 2783–2800, doi:10.1002/(SICI)1099-1085(19991215)13:17<2783::AID-HYP899>3.0.CO;2-9, 1999.

Galantowicz, J. F., Entekhabi, D., and Njoku, E. G.: Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness, IEEE. T. Geosci. Remote., 37(4), 1860-1870, doi:10.1109/36.774699, 1999.

Henderson-Sellers, A., Yang, Z.-L., Dickinson, R. E., Henderson-Sellers, A., Yang, Z.-L. and Dickinson, R. E.: The Project for Intercomparison of Land-surface Parameterization Schemes, Bull. Am. Meteorol. Soc., 74(7), 1335–1349, doi:10.1175/~~1520-0477~~15200477(1993)074<1335:TPFIOL>2.0.CO;2,~~-~~1993.

Hu, S., Shi, L., Zha, Y., Williams, M. and Lin, L.: Simultaneous state-parameter estimation supports the evaluation of data assimilation performance and measurement design for soil-water-atmosphere-plant system, J. Hydrol., 555, 812–831, doi:10.1016/j.jhydrol.2017.10.061, 2017.

Hendricks Franssen, H. J. and Kinzelbach, W.: Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem, Water Resour. Res., 44(9), 1–21, doi:10.1029/2007WR006505, 2008.

Huang, C., Li, X., Lu, L. and Gu, J.: Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter, Remote Sens. Environ., 112(3), 888–900, doi:10.1016/j.rse.2007.06.026, 2008.

Huo, S., Jin, M., Liang, X. and Lin, D.: Changes of vertical groundwater recharge with increase in thickness of vadose zone simulated by one-dimensional variably saturated flow model, J. Earth Sci., 25(6), 1043–1050, doi:10.1007/s12583-014-0486-7, 2014.

Ji, S. and Unger, P. W.: Soil Water Accumulation under Different Precipitation, Potential Evaporation, and Straw Mulch Conditions, Soil Sci. Soc. Am. J., 65(2), 442, doi:10.2136/sssaj2001.652442x, 2001.

Li, B., Toll, D., Zhan, X. and Cosgrove, B.: Improving estimated soil moisture fields through assimilation of AMSR-E soil moisture retrievals with an ensemble Kalman filter and a mass conservation constraint, Hydrol. Earth Syst. Sci., 16(1), 105–119, doi:10.5194/hess-16-105-2012, 2012.

Li, C. and Ren, L.: Estimation of Unsaturated Soil Hydraulic Parameters Using the Ensemble Kalman Filter, Vadose Zo. J., 10(4), 1205, doi:10.2136/vzj2010.0159, 2011.

Li, X., Shi, L., Zha, Y., Wang, Y. and Hu, S.: Data assimilation of soil water flow by considering multiple uncertainty sources and spatial–temporal features: a field-scale real case study, Stoch. Environ. Res. Risk Assess., 32(9), 2477–2493, doi:10.1007/s00477-018-1541-1, 2018.

Lim, Y.-J., Hong, J. and Lee, T.-Y.: Spin-up behavior of soil moisture content over East Asia in a land surface model, Meteorol. Atmos. Phys., 118(3-4), 151–161, doi:10.1007/s00703-012-0212-x, 2012.

Margulis, S. ~~a~~A., Mclaughlin, D. B., Entekhabi, D. and Dunne, S.: Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment, Water Resour. Res., 38(12), 1–18, doi:10.1029/2001WR001114, 2002.

Medina, H., Romano, N. and Chirico, G. B.: Kalman filters for assimilating near-surface observations into the Richards equation – Part 2: A dual filter approach for simultaneous retrieval of states and parameters, Hydrol. Earth Syst. Sci., 18(7), 2521–2541, doi:10.5194/hess-18-2521-2014, 2014a.

Medina, H., Romano, N. and Chirico, G. B.: Kalman filters for assimilating near-surface observations into the Richards equation - Part 3: Retrieving states and parameters from laboratory evaporation experiments, Hydrol. Earth Syst. Sci., 18(7), 2543–2557, doi:10.5194/hess-18-2543-2014, 2014b.

Mohamed Mumen: Caractérisation du fonctionnement hydrique des sols à l’aide d’un modèle mécaniste de transferts d’eau et de chaleur mis en ~~œuvre~~oeuvre en ~~fonction~~fonctions des informations disponibles sur le sol, Ph.D. thesis, University of Avignon, Avignon, France, 2006.

Montzka, C., Moradkhani, H., Weihermüller, L., Franssen, H. J. H., Canty, M. and Vereecken, H.: Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter, J. Hydrol., 399(3-4), 410–421, doi:10.1016/j.jhydrol.2011.01.020, 2011.

Mualem, Y.: Wetting front pressure head in the infiltration model of Green and Ampt, Water Resour. Res., 12(3), 564–566, doi:10.1029/WR012i003p00564, 1976.

Oliver, D. S. and Chen, Y.: Improved initial sampling for the ensemble Kalman filter, Comput. Geosci., 13(1), 13–27, doi:10.1007/s10596-008-9101-2, 2009.

Pujol, J.: The solution of nonlinear inverse problems and the Levenberg-Marquardt method, Geophysics, 72(4), doi:10.1190/1.2732552, 2007.

Rahman, M. M. and Lu, M.: Model spin-up behavior for wet and dry basins: A case study using the xinanjiang model, Water (Switzerland), 7(8), 4256–4273, doi:10.3390/w7084256, 2015.

Reichle, R. H. and Koster, R. D.: Assessing the Impact of Horizontal Error Correlations in Background Fields on Soil Moisture Estimation, J. Hydrometeorol., 4(6), 1229–1242, doi:10.1175/1525-7541(2003)004<1229:ATIOHE>2.0.CO;2, 2003.

Rodell, M., Houser, P. R., Berg, A. A. and Famiglietti, J. S.: Evaluation of 10 Methods for Initializing a Land Surface Model, J. Hydrometeorol., 6(2), 146–155, doi:10.1175/JHM414.1, 2005.

Ross, P. J.: Modeling Soil Water and Solute Transport~~—~~ Fast, Simplified Numerical Solutions, Agron. J., 95, 1352~~–1361~~, https://doi.org/10.2134/agronj2003.1352, 2003.

Ross, P. J.: Fast Solution of Richards ’ Equation for Flexible Soil Hydraulic Property Descriptions. Land and Water Technical Report, CSIRO, 39(06), 2006.

Seck, A., Welty, C. and Maxwell, R. M.: Spin-up behavior and effects of initial conditions for an integrated hydrologic model, Water Resour. Res., 51, 2188–2210, doi:10.1002/2014WR016371, 2014.

Shi, L., Song, X., Tong, J., Zhu, Y. and Zhang, Q.: Impacts of different types of measurements on estimating unsaturated flow parameters, J. Hydrol., 524, 549–561, doi:10.1016/j.jhydrol.2015.01.078, 2015.

~~Šim˚unek~~Šimůnek, J., Šejna, M., Saito, H., Sakai, M., and van Genuchten, M. T.: The Hydrus-1D software package for simulating the movement of water, heat, and multiple solutes in variably saturated media, version 4.16, HYDRUS software series 3 Department of Environmental Sciences, University of California Riverside, Riverside, California, USA, 308 pp., 2013.

Szomolay, B.: Analysis of a moving boundary value problem arising in biofilm modelling, Math. ~~method~~Method Appl. Sci., 31(15), 1835–1859, doi:10.1002/mma.1000, 2008.

Tran, A. P., Vanclooster, M. and Lambot, S.: Improving soil moisture profile reconstruction from ground-penetrating radar data: A maximum likelihood ensemble filter approach, Hydrol. Earth Syst. Sci., 17(7), 2543–2556, doi:10.5194/hess-17-2543-2013, 2013.

~~van~~Van Genuchten, M. T., and Parker, J. C.: Boundary Conditions for Displacement Experiments Through Short Laboratory Soil Columns, Soil Sci. Soc. Am. J., 48(4), 703–708, doi:10.2136/sssaj1984.03615995004800040002x, 1984.

Varado, N., Braud, I., Ross, P. J., & Haverkamp, R.: Assessment of an efficient numerical solution of the 1D Richards' equation on bare soil, J. Hydrol., 323(1-4), 244–257, https://doi.org/10.1016/j.jhydrol.2005.07.052, 2006.

Vereecken, H., Huisman, J. A., Bogena, H., Vanderborght, J., Vrugt, J. A., & Hopmans, J. W.: On the value of soil moisture measurements in vadose zone hydrology: A review, Water Resour. Res., 46(4), 1–21. https://doi.org/10.1029/2008WR006829, 2010.

Walker, J. P. and Houser, P. R.: A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations, J. Geophys. Res. Atmos., 106(D11), 11761–11774, doi:10.1029/2001JD900149, 2001.

Wu, C. C. and Margulis, S. A.: Feasibility of real-time soil state and flux characterization for wastewater reuse using an embedded sensor network data assimilation approach, J. Hydrol., 399(3-4), 313–325, doi:10.1016/j.jhydrol.2011.01.011, 2011.

Wu, C~~.~~. C. and Margulis, S. A.: Real-Time Soil Moisture and Salinity Profile Estimation Using Assimilation of Embedded Sensor Datastreams, Vadose Zo. J., 12(1), doi:10.2136/vzj2011.0176, 2013.

Xie, T., Liu, X. and Sun, T.: The effects of groundwater table and flood irrigation strategies on soil water and salt dynamics and reed water use in the Yellow River Delta, China, Ecol. Modell., 222(2), 241–252, doi:10.1016/j.ecolmodel.2010.01.012, 2011.

Yang, Z. L., Dickinson, R. E., Henderson-Sellers, A. and Pitman, A. J.: Preliminary study of spin-up processes in land surface models with the first stage data of Project for Intercomparison of Land Surface Parameterization Schemes Phase 1(a), J. Geophys. Res., 100(D8), 16553, doi:10.1029/95JD01076, 1995.

Zha, Y., Shi, L., Ye, M. and Yang, J.: A generalized Ross method for two- and three-dimensional variably saturated flow, Adv. Water Resour., 54, 67–77, doi:10.1016/j.advwatres.2013.01.002, 2013.

Table 1. Soil hydraulic parameters used in simulation.

Soil	θ_s	θ_r	K_s/md^{-1}	α/m^{-1}	n
Sand	0.43	0.045	7.128	14.5	2.68
Loam	0.43	0.078	0.2496	3.6	1.56
Silt	0.46	0.034	0.06	1.6	1.37
Clay loam	0.41	0.095	0.062	1.9	1.31

Table 2. The ~~candidates of initialization scheme and~~ default model settings used in the simulations.

Parameter definition	Value or type
Initial condition	a uniform 50% relative saturation over the soil profile (IC-HfSatu)
Number of soil layers	1
Thickness of soil zone	3 m
Soil hydraulic properties	Loam
Upper boundary	M-SC
Bottom boundary	Free drainage
Number of grids	60 (with the size of 5 cm)
Simulation time	365 days

~~Soil hydraulic parameters used in simulation.~~

~~Soil~~	~~θ_s~~	~~θ_r~~	~~K_s/md^{-1}~~	~~α/m^{-1}~~	~~n~~
~~Sand~~	~~0.43~~	~~0.045~~	~~7.128~~	~~14.5~~	~~2.68~~
~~Loam~~	~~0.43~~	~~0.078~~	~~0.2496~~	~~3.6~~	~~1.56~~
~~Silt~~	~~0.46~~	~~0.034~~	~~0.06~~	~~1.6~~	~~1.37~~
~~Clay loam~~	~~0.41~~	~~0.095~~	~~0.062~~	~~1.9~~	~~1.31~~

Table 3. Case summary for parameter estimation within EnKF and IES.

Case	Description	Ensemble Size	Initial Uncertainty	Simulation Time	Framework
Case 1	Individual parameter estimation	-	-	-	EnKF/IES
Case 2		500	-	-	EnKF/IES
Case 3		500	0.017	-	EnKF/IES
Case 4	Multiple parameter estimation	-	-	-	EnKF/IES
Case 5		-	-	60	IES

Note: Ungiven values use the default values. The default value of initial uncertainty for IC-ObsInt and IC-HfSatu is 0.

Table 4. *$RMSE_{obs}$* results for the soil moisture predictions at observation points with different initial conditions in the experimental case.

Initial condition	**10cm**	**20cm**	**30cm**
IC-HfSatu	0.0232	0.0271	0.0280
IC-ObsInt	0.0286	0.0187	0.0134
IC-Flux	0.0198	0.0222	0.0206
IC-WUP	0.0180	0.0153	0.0155

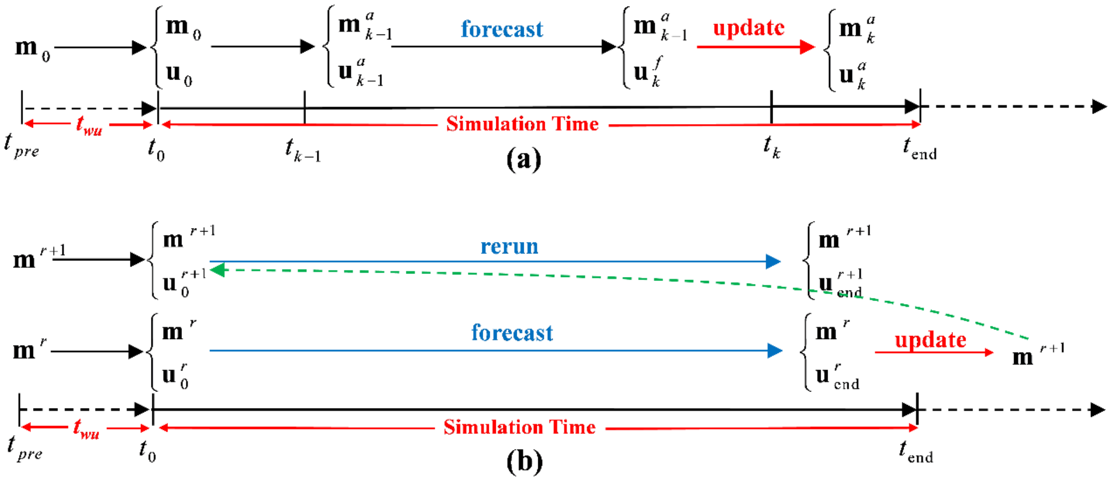
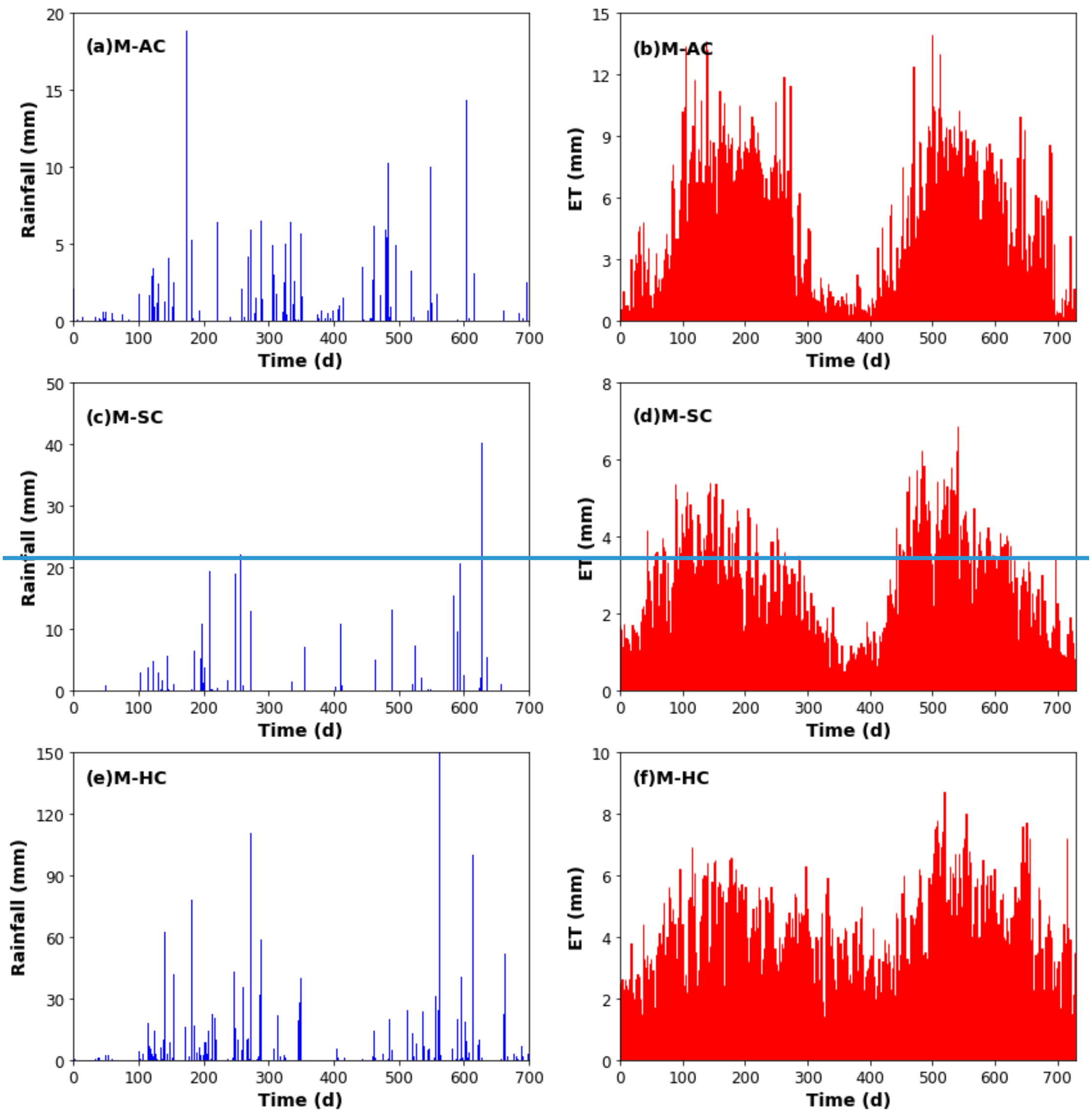
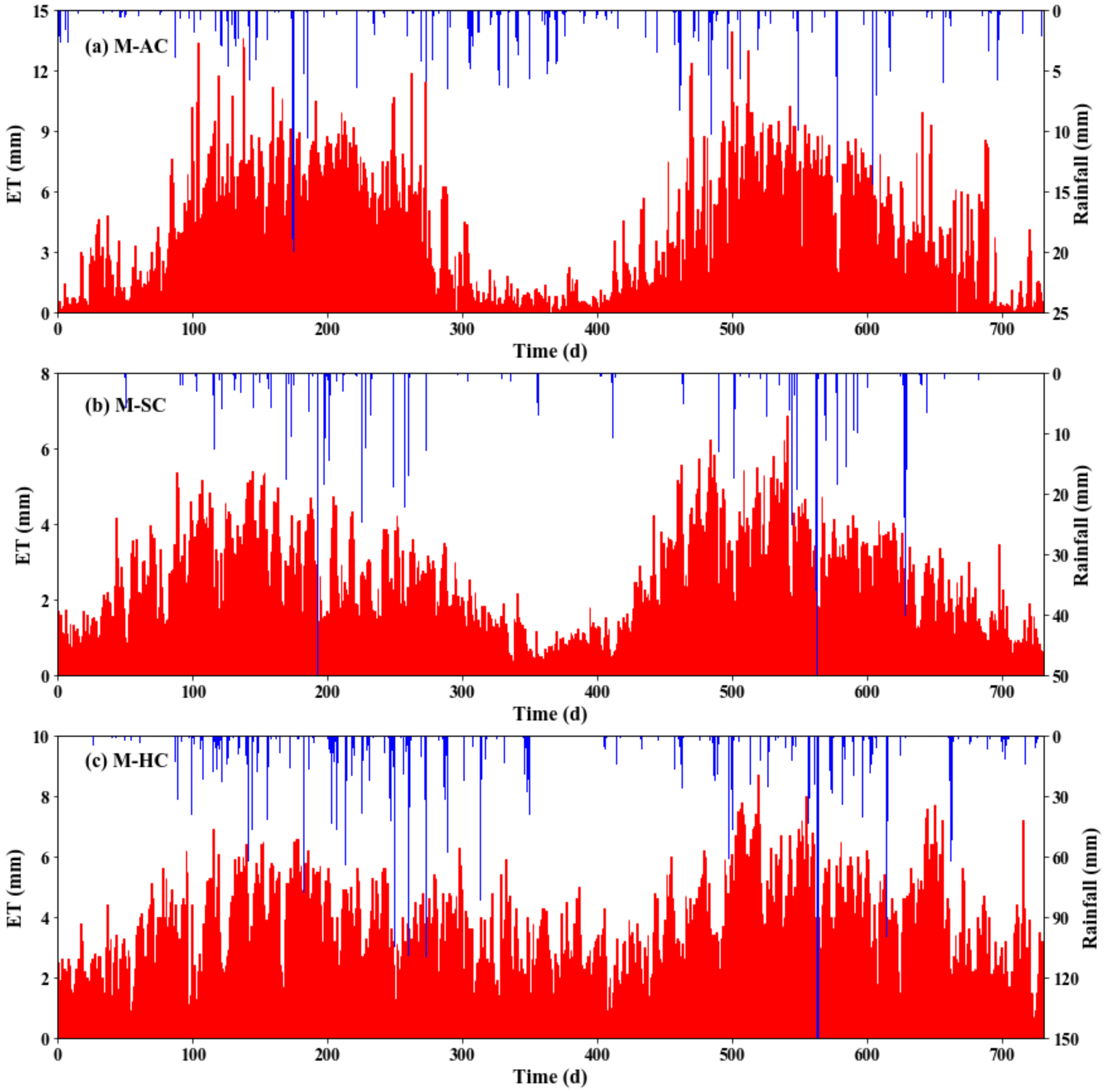


Fig. 1. Flowcharts of simulation period (or data assimilation period with (a) ensemble Kalman filter (EnKF) and (b) iterative ensemble smoother (IES)) and warming up period. t_0 is the initial time and t_{end} is the end ~~time~~ of simulation time. $\mathbf{m}_k$ and $\mathbf{u}_k$ are the vectors of model parameters (e.g., hydraulic conductivity) and state variables (e.g., soil moisture), respectively, at time t_k, while ~~$\mathbf{m}^r$~~ $\mathbf{m}^r$ and ~~$\mathbf{u}^r$~~ $\mathbf{u}^r$ are the vectors at iteration r; the superscripts a and f refer to model analysis and forecast (or initial guess). Besides, the period between t_{pre} and t_0 donates the process of warming up, and t_{wu} is the required warm-up time.


(a)M-AC
Rainfall (mm)
Time (d)
(b)M-AC
ET (mm)
Time (d)
(c)M-SC
Rainfall (mm)
Time (d)
(d)M-SC
ET (mm)
Time (d)
(e)M-HC
Rainfall (mm)
Time (d)
(f)M-HC
ET (mm)
Time (d)



E

Fig. 2. Synthetic rainfall (blue bars) and potential evaporation (red bars) of three typical climates: (a) ~~synthetic rainfall of~~ arid climate, (b) ~~synthetic potential evaporation of arid climate, (c) synthetic rainfall of~~ semi-arid climate, ~~(d) synthetic potential evaporation of semi-arid climate, (e) synthetic rainfall of~~ and (c) humid climate~~, (f) synthetic potential evaporation of humid climate~~. It should be noted that the meteorological data of simulation period is from day 366 to day 730.

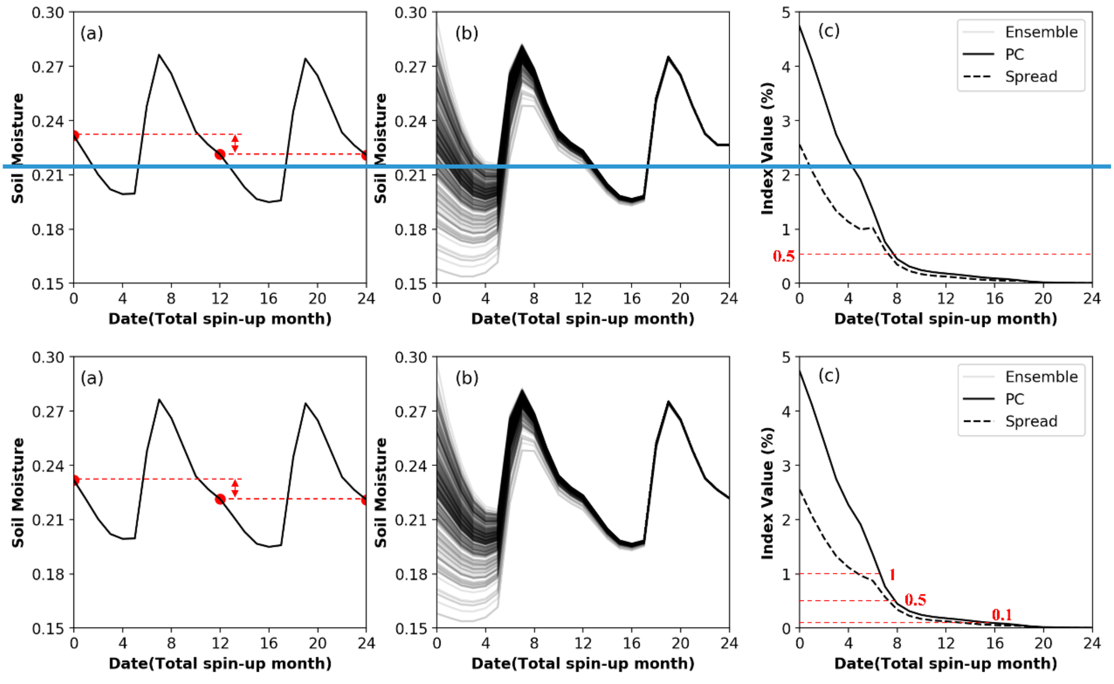


Fig. 3. Comparison of Spin-up and Monte-Carlo methods in determining warm-up time. (a) Spin-up method with monthly-averaged soil moisture versus time by running a simulation recursively for ~~several~~10 years, (b) Monte-Carlo method with monthly-averaged soil moisture of different realizations versus time based on various initial conditions, and (c) Comparison of *PC* and S_p versus time. For the purpose of demonstration, the parameter uncertainty is not considered and we only show the results of the first two years in the figure.

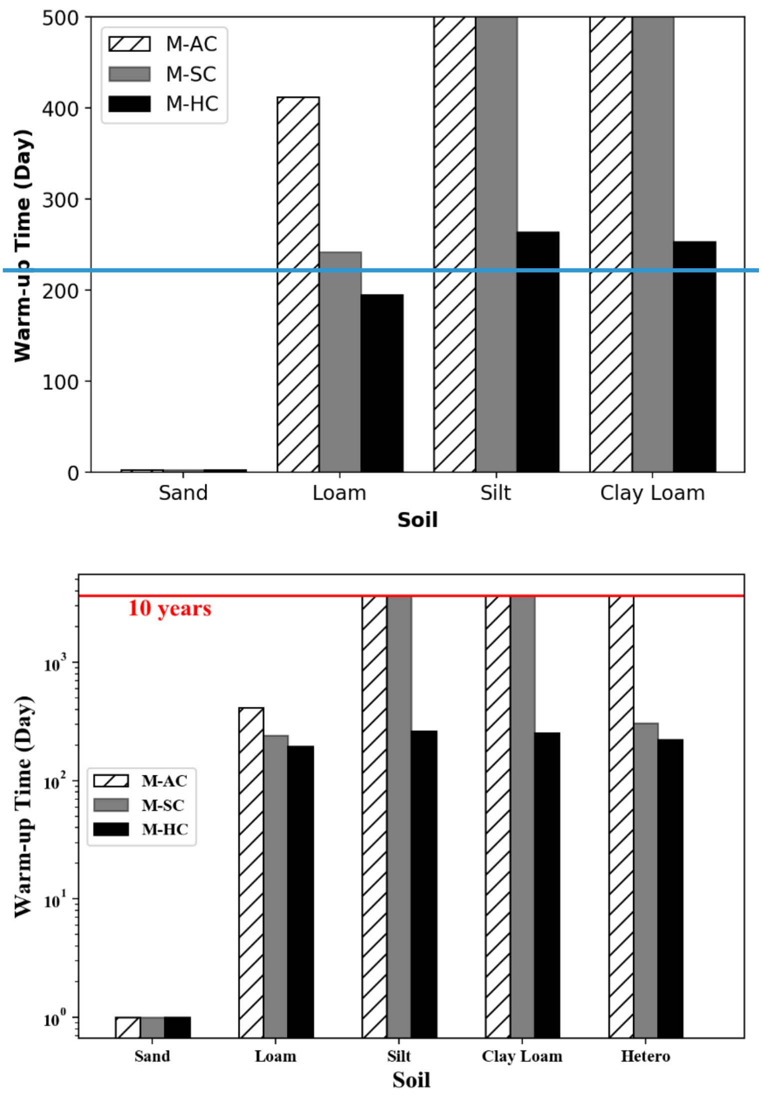
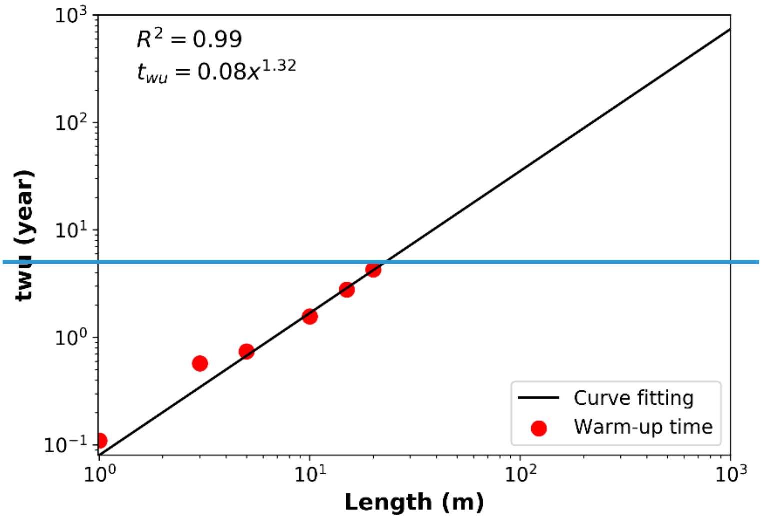


Fig. 4. The length of warm-up time t_{wu} with various soils and meteorological conditions. Note that some of the t_{wu} ~~of Silt~~ values are larger than 10 years and are not able to be obtained due to the 10-year simulation time. The heterogeneous soil profile consists of Loam (0-75 cm), Clay loam ~~with M-AC~~(75-150 cm), Silt (150-225 cm), and ~~M-SC exceed 10 years, and they are trimmed for visualization purpose.~~Sand (225-300 cm).



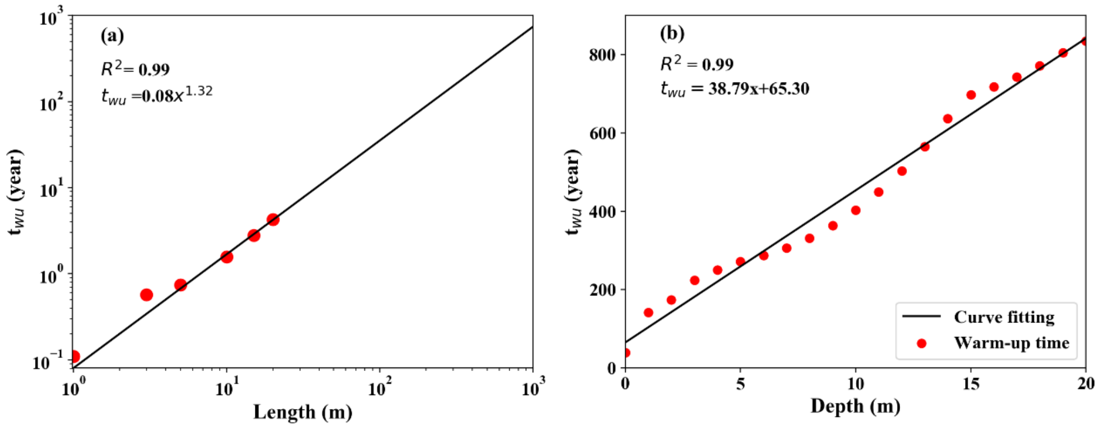
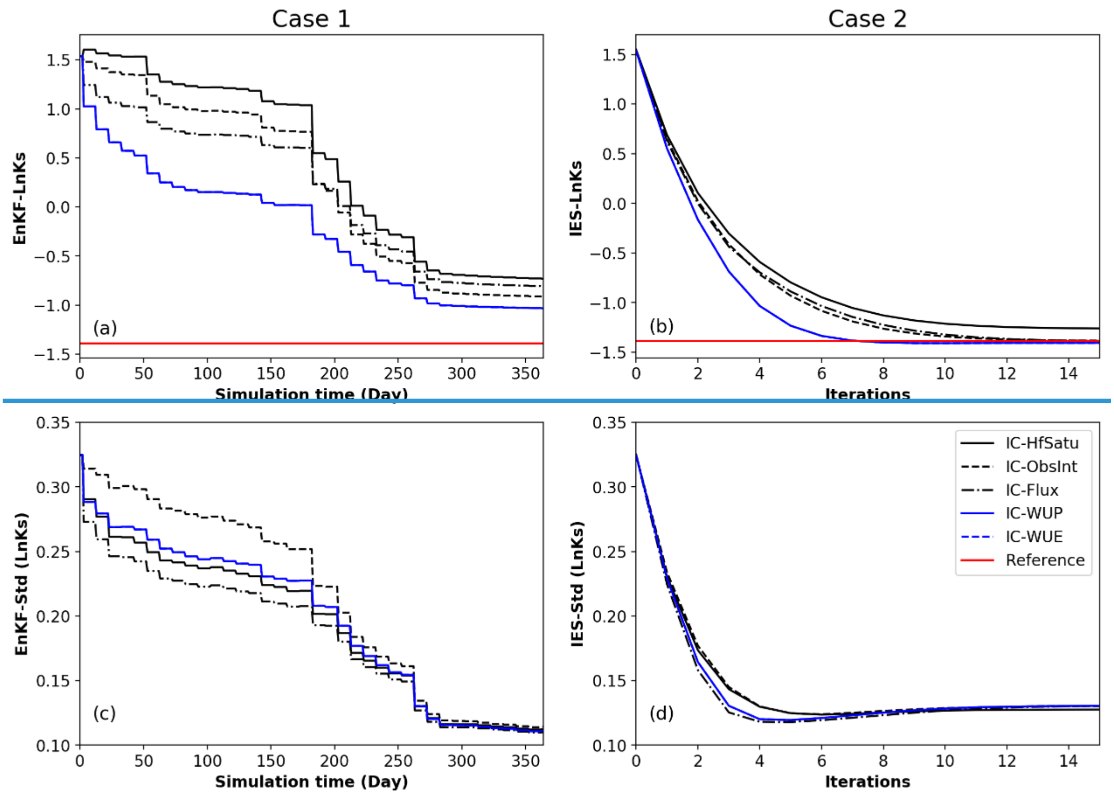


Fig. 5. The ~~relationship between~~ value of the ~~length of soil profile and~~ warm-up time t_{wu}. (a) The overall profile t_{wu} values versus different soil profile lengths and (b) t_{wu} value as a function of depth z with a 20-m soil profile.



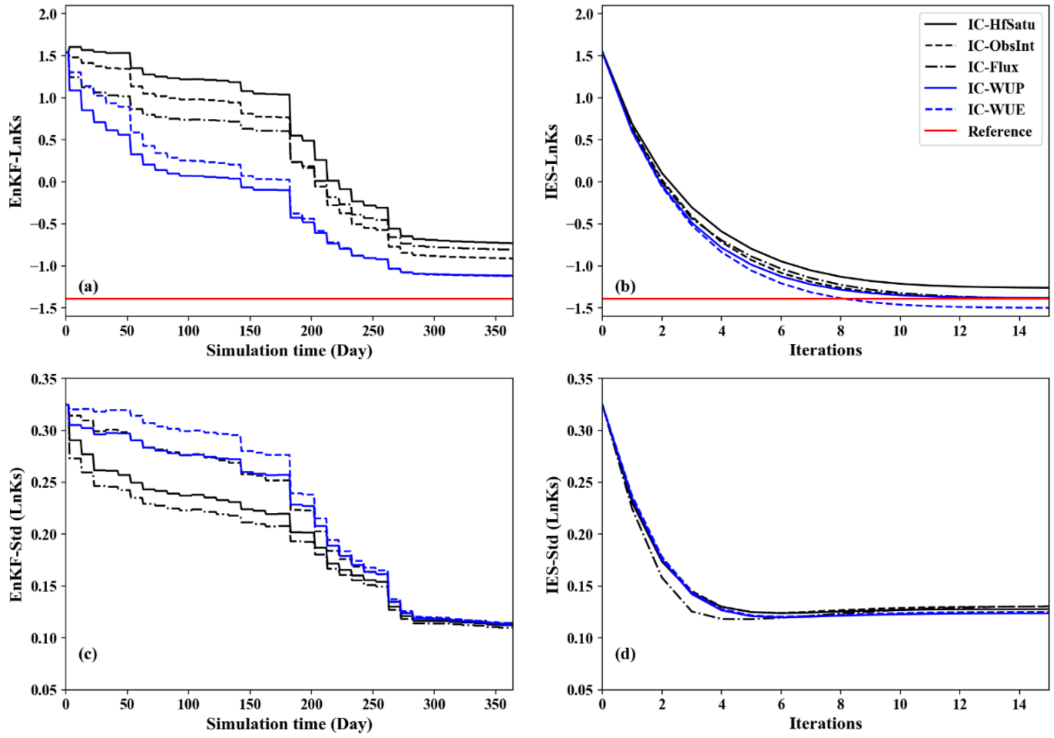
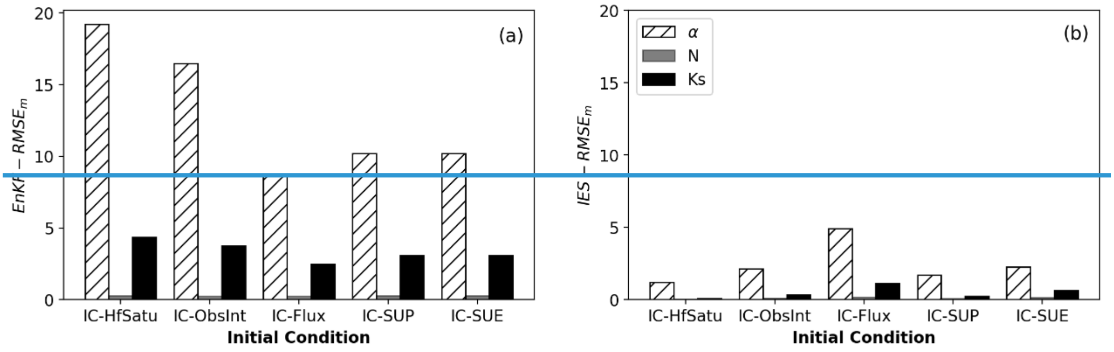


Fig. 6. The results of lnK_s estimations (first row) and their associated standard deviations (second row) within two data assimilation frameworks (left: EnKF; right: IES) under five initialization methods~~.~~ (Case 1).



~~**Fig. 7.** The RMSE~~$\sout{_m}$

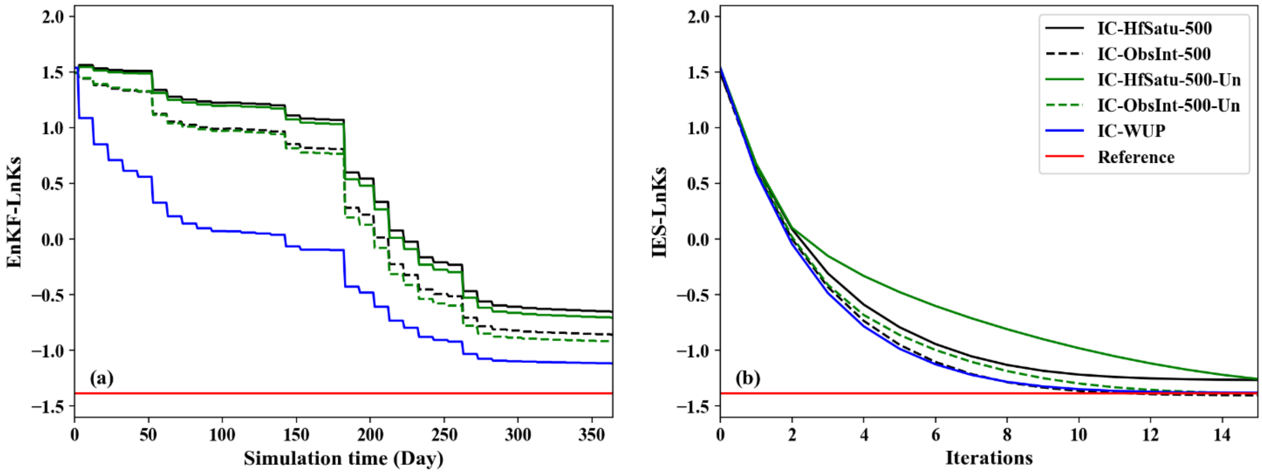


Fig. 7. The impacts of increased ensemble size (Case 2) and uncertainty of initial state (Case 3) on the results of lnK_s estimations within EnKF and IES.

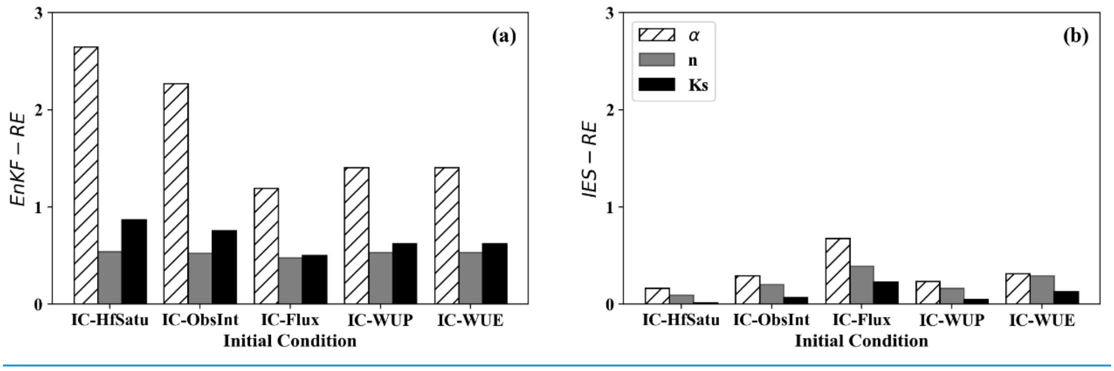


Fig. 8. The *RE* results of parameter estimations (α, n and K_s) under five initialization methods with (a) EnKF and (b) IES. (Case 4).

920

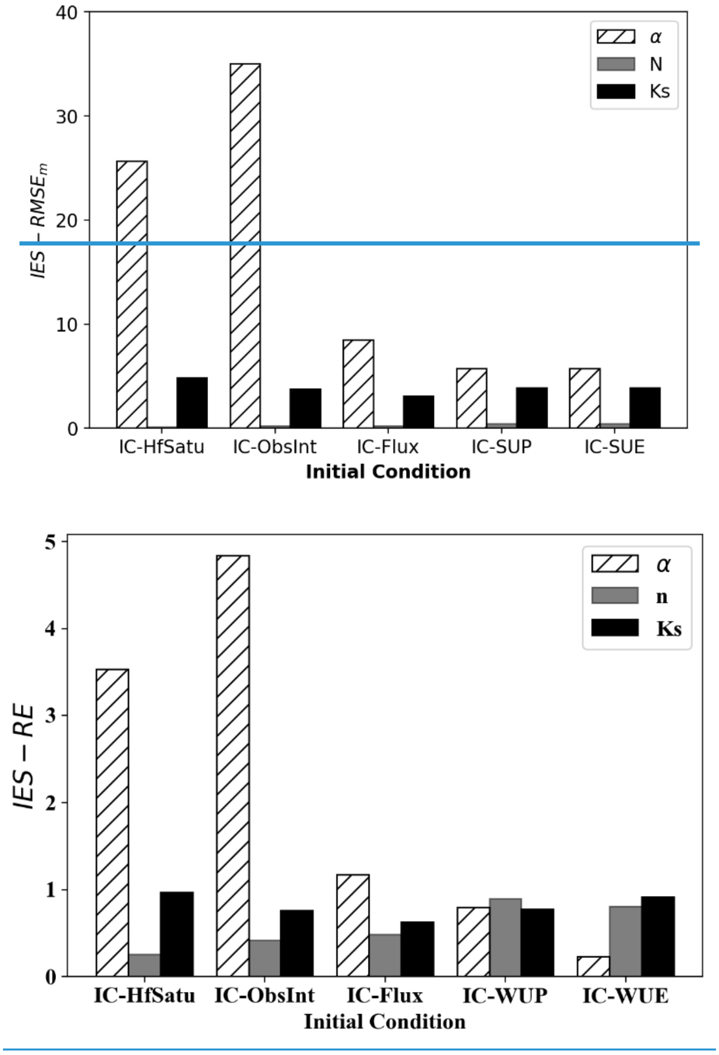


Fig. ~~8~~9. The ~~$RMSE_m$~~ RE results of parameter estimations under five initialization methods with IES when the simulation time is 60 days~~.~~ (Case 5).

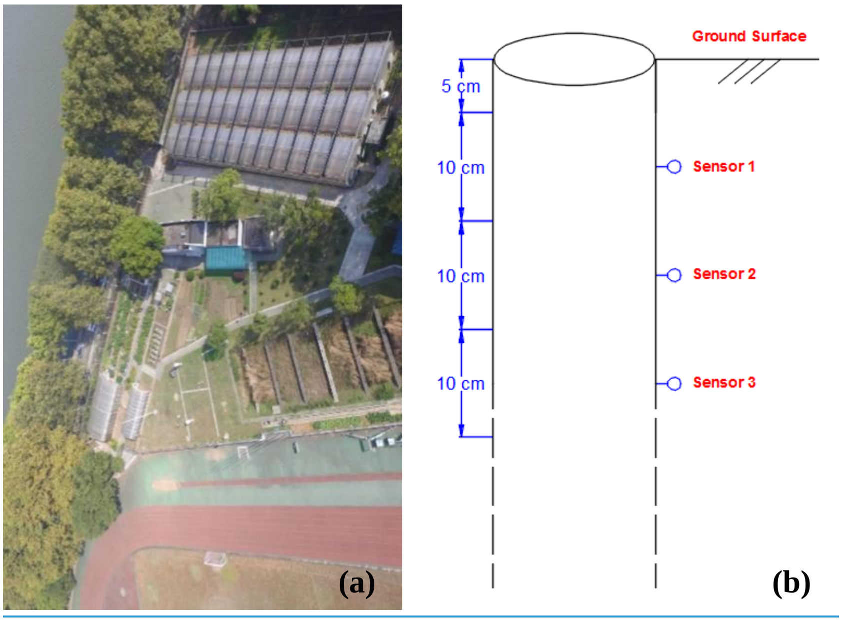


Fig. ~~9~~10. The experimental site: (a) plan view, and (b) the cross-sectional view of the FDR sensor.

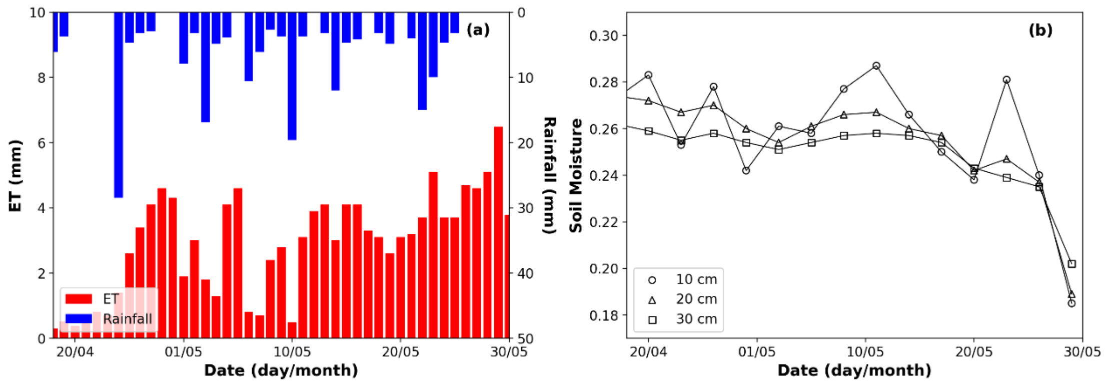


Fig. ~~10~~11. The meteorological information and observed soil moistures over the experimental time. (a) Observed rainfall and calculated potential evaporation. (b) Temporal change of soil moisture data at three different observed depths (10 cm, 20 cm and 30 cm).

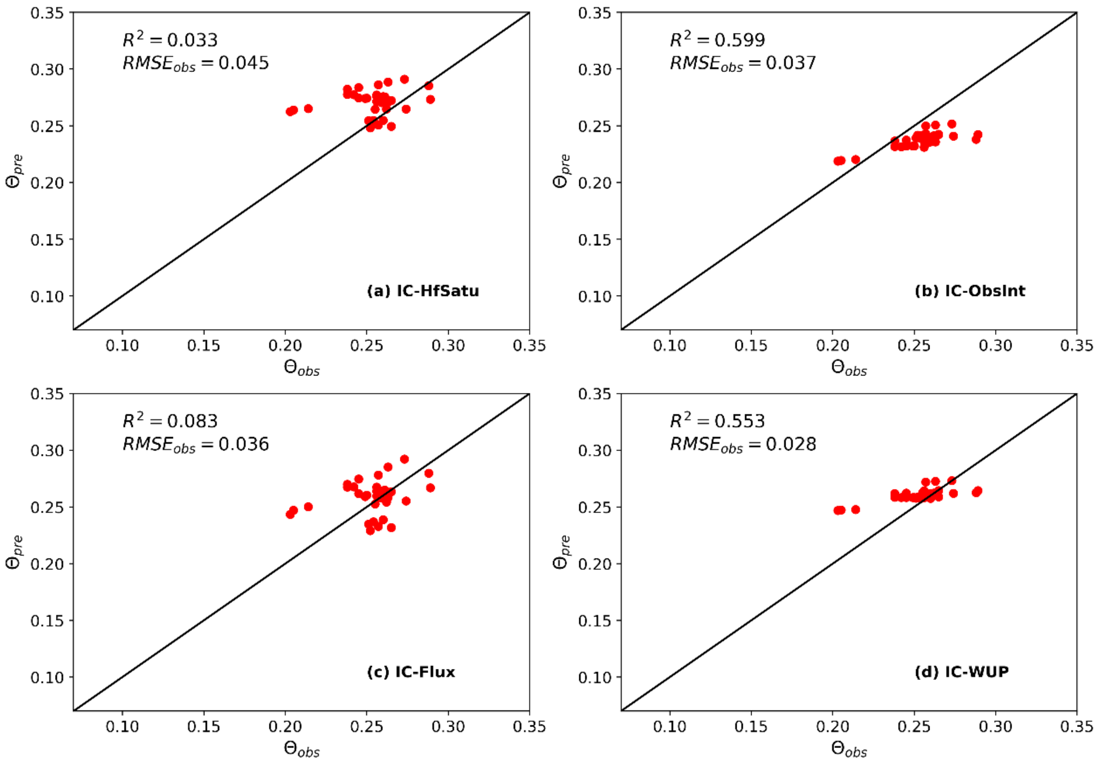


Fig. ~~11~~12. The comparisons between soil moisture observations and predictions ~~at all depths~~ (with ~~four~~estimated parameters from IES combined with different ~~initial conditions~~initialization methods) at all observation depths.