

The consistencies of the ensemble Kalman filter (EnKF) and iterative EnKF algorithms for two-dimensional nonlinear unsaturated flow in randomly heterogeneous soil are discussed. A modified Restart EnKF with low computational effort and sound consistency is proposed. The effects of various factors, e.g., observation type and damping factor, are investigated.

X. Song, L. Shi, and J. Yang, State Key Lab. of Water Resources and Hydro-power Engineering Science, Wuhan Univ., Wuhan 430072, China; M. Ye and I.M. Navon, Dep. of Scientific Computing, Florida State Univ., Tallahassee, FL 32306. *Corresponding author (jzyang@whu.edu.cn).

Vadose Zone J.
doi:10.2136/vzj2013.05.0083
Received 14 May 2013.

© Soil Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA.

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Numerical Comparison of Iterative Ensemble Kalman Filters for Unsaturated Flow Inverse Modeling

Xuehang Song, Liangsheng Shi, Ming Ye, Jinzhong Yang,* and I. Michael Navon

This study evaluated three algorithms of the iterative ensemble Kalman filter (EnKF). They are Confirming EnKF, Restart EnKF, and modified Restart EnKF developed to resolve the inconsistency problem (i.e., updated model parameters and state variables do not follow the Richards equation) in vadose zone data assimilation due to model nonlinearity. While Confirming and Restart EnKF were adapted from literature, modified Restart EnKF was developed in this study to reduce computational costs by calculating only the mean simulation, not all the ensemble realizations, from time $t = 0$. A total of 11 cases were designed to investigate the performance of EnKF, Confirming EnKF, Restart EnKF, and modified Restart EnKF with different types and spatial configurations of observations (pressure head and water content) and different values of observation error variance, initial guess of ensemble mean and variance, ensemble size, and damping factor. The numerical study showed that Confirming EnKF produced considerable inconsistency for the nonlinear unsaturated flow problem, which differs from the apparent consensus opinion that Confirming EnKF can resolve the inconsistency problem. In contrast, Restart EnKF and its modification can resolve the inconsistency problem. Restart EnKF and its modification outperformed EnKF and Confirming EnKF in the various cases considered in this study. It was also found that combining different types of observations can achieve better assimilation results, which is useful for monitoring network design.

Abbreviations: EnKF, ensemble Kalman filter; IC, inconsistency.

Inverse modeling is always required in vadose zone modeling to estimate optimum soil hydraulic parameters in soil water retention and hydraulic conductivity functions and to quantify parameter uncertainty (Vrugt et al., 2008). Among various inverse approaches, data assimilation methods have become popular because they can update not only model parameters (e.g., soil hydraulic parameters) but also system states (e.g., hydraulic head). The ensemble Kalman filter (EnKF) (Evensen, 1994, 2009; Burgers et al., 1998) is one of the most popular sequential data assimilation methods, and it has been recently used for soil moisture estimating (Reichle, 2008) and inverse modeling in vadose zone hydrology (e.g., Li and Ren, 2011; Wu and Margulis, 2011, 2013). As a Monte Carlo method, EnKF avoids evolving the covariance matrix of the probabilistic distribution function of the state vector (Johns and Mandel, 2007). It does not require the direct calculation of the objective function value or the evaluation of the tangent linear operator (as in the extended Kalman filter) or adjoint equations (as in variational data assimilation). Furthermore, EnKF is inherently compatible with parallel computation techniques because each ensemble member can be independently run on a different processor. Although EnKF does not require linearization of nonlinear systems, however, EnKF may not be suitable for parameter estimation for highly nonlinear systems (Sakov et al., 2012) because nonlinear systems always correspond to non-Gaussian distributions of model parameters and model errors. A vadose zone system is highly nonlinear due to a rapid change in surface moisture (the driving force) and, more importantly,

the nonlinear relation between soil moisture and pressure head. Therefore, research is needed to better understand the applicability of EnKF to vadose zone inverse modeling.

When applying EnKF to unsaturated flow inverse modeling, the nonlinearity may cause a possible inconsistency between updated parameters and updated system states, i.e., updated pressure heads and soil hydraulic parameters do not follow the Richards equation. It happens because EnKF updates model parameters and state variables synchronously using linear update formulae, which cannot guarantee that the updated parameters and state variables follow the nonlinear Richards equation. An extreme example is that updates of the state variables are physically unreasonable, e.g., water saturations being <0 or >1.0 . While the extreme cases can be easily identified and fixed, the inconsistency problem is always hidden and difficult to resolve. In addition, the inconsistency may accumulate in assimilation steps and will eventually lead to incorrect parameter estimates and model predictions. There is an urgent need to resolve the inconsistency problem.

Considerable attention has been paid to resolve the problems of nonlinearity and inconsistency in ensemble data assimilation, and various approaches have been developed. One of the most prominent methods is the particle filter (Moradkhani, 2005; Montzka et al., 2011; Rings et al., 2012), which uses a set of particles to represent the posterior density and thus makes no restrictive assumption about the probability density function of the state vector. There are also other approaches such as merging EnKF and particle filter (Hoteit et al., 2008) and modifying EnKF (e.g., the non-Gaussian EnKF of Anderson, 2010). This study focused on the iterative EnKF, which modifies EnKF to include an option of updating the state variables by rerunning the nonlinear model with the updated model parameters. In comparison with the other approaches for nonlinear systems, iterative EnKF is still within the framework of EnKF and thus theoretically straightforward. The concept of iterative updating can be traced back to Jazwinski (1970) and Navon (1998) for global and local iterations of the extended Kalman filter; the difference between global and local iterations is whether the update is conducted across many assimilation cycles (global) or across a single assimilation cycle (local). For EnKF, global and local iterations can be found in Vrugt et al. (2005) and Zupanski and Collins (2005), respectively. With respect to the starting time of each iteration, iterative EnKF algorithms can be classified into two categories: (i) a Confirming EnKF (Wen and Chen, 2006) that reruns the nonlinear model from the previous assimilating step (Moradkhani et al., 2005; Wen and Chen, 2006; Li and Reynolds, 2007; Lorentzen and Naevdal, 2011); and (ii) a Restart EnKF that reruns the nonlinear model from time zero (Reynolds et al., 2006; Gu and Oliver, 2007; Thulin et al., 2007; Wang et al., 2010; Chen, 2012; Elsheikh et al.,

2013). Hendricks Franssen and Kinzelbach (2008) compared EnKF, Confirming EnKF, and Restart EnKF in a numerical study of synthetic groundwater modeling with moderately and strongly heterogeneous transmissivity fields and found that the latter two methods gave only slightly better results than EnKF. The seemingly consensus opinion is that Confirming EnKF can resolve the inconsistency problem. However, Zafari and Reynolds (2007) showed in theoretical and numerical studies that Confirming EnKF cannot resolve the inconsistency problem. This study contributes to the debate in the context of vadose zone modeling through a numerical study.

The objective of this study was twofold: (i) to conduct a comprehensive numerical study for better understanding the strength and weakness of Confirming EnKF and Restart EnKF; and (ii) to modify the Restart EnKF algorithm to improve its computational efficiency because rerunning the nonlinear model from time zero is computationally expensive. The numerical study showed that the inconsistency problem occurs in EnKF results and cannot be resolved by using Confirming EnKF but by Restart EnKF. This finding is consistent with that of Zafari and Reynolds (2007) but different from that of Hendricks Franssen and Kinzelbach (2008) and thus challenges the seemingly consensus opinion that Confirming EnKF has been accepted as a common practice in many studies to alleviate inconsistency (Gu and Oliver, 2006; Krymskaya et al., 2008; Hendricks Franssen and Kinzelbach, 2008; Li et al., 2010; Zagayevskiy et al., 2012; Zhang et al., 2012).

The numerical experiments of this study have several unique features. First, different from the previous studies of data assimilation that have been focused on one-dimensional homogeneous or layered soil columns (e.g., Montzka et al., 2011; Lü et al., 2011; Li and Ren, 2011; Wu and Margulis, 2011), this study looked at two-dimensional, heterogeneous soils with more complicated flow patterns and a larger number of state variables. It revealed challenges of applying existing EnKF algorithms to more realistic vadose modeling in randomly heterogeneous soils. The numerical experiment also comprehensively investigated the factors (observation error variance, initial guess, and ensemble size) that may affect the possible inconsistency in data assimilation. The damping factor and two types of observations (pressure head and water content) were also considered. The damping factor was used in the manner of Wu and Margulis (2011) to reduce filter divergence. Evaluating data values of the two types of observations is helpful to design monitoring networks for site characterization and reduction of predictive uncertainty, as reported in previous studies (Abbaspour et al., 2000; Jacques et al., 2002; Abbasi et al., 2003a, 2003b; Zhang et al., 2003; Mirus et al., 2009; Verbist et al., 2009; Wöhling and Vrugt, 2011). The EnKF, Confirming EnKF, Restart EnKF, and modified Restart EnKF are described, followed by a discussion of the numerical results.

Unsaturated Flow and Ensemble Kalman Filter Methodologies

Governing Equation

This study considered a two-dimensional, transient unsaturated flow system in a randomly heterogeneous soil. The unsaturated flow is governed by the Richards equation:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial x} \left[K(h) \frac{\partial h}{\partial x} \right] + \frac{\partial}{\partial z} \left[K(h) \frac{\partial h}{\partial z} + K(h) \right] - S(t, z, h) \quad [1]$$

where θ is the soil volumetric water content [L^3/L^3], t is time [T], h is the soil water pressure head [L], $K(h)$ is the unsaturated hydraulic conductivity [L/T], x is the horizontal coordinate [L], z is the vertical coordinate [L], and $S(t, z, h)$ is the root water uptake or other source–sink term [T^{-1}]. Solution of Eq. [1] requires a relation describing the soil water retention characteristics. The van Genuchten–Mualem (van Genuchten, 1980) equation is one of the most widely used relations:

$$\theta(h) = \begin{cases} \theta_r + \frac{\theta_s - \theta_r}{(1 + |\alpha h|^n)^m} & h < 0 \\ \theta_s & h \geq 0 \end{cases} \quad [2]$$

$$K(h) = \begin{cases} K_s S_e^{1/2} \left[1 - (1 - S_e^{1/m})^m \right]^2 & h < 0 \\ K_s & h \geq 0 \end{cases} \quad [3]$$

where

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r} \quad [4]$$

is the effective water saturation (dimensionless), θ_r and θ_s are the residual and saturated volumetric water contents [L^3/L^3], respectively, K_s is the saturated hydraulic conductivity [L/T], α and n are parameters related to the soil pore-size distribution (dimensionless), and $m = 1 - 1/n$ (dimensionless). Hysteresis of water retention curves was not considered in this study. With the given boundary and initial conditions, Eq. [1] can be solved using standard Galerkin finite element methods. The variably saturated finite element code SWMS-2D of Šimůnek et al. (1994) was used in this study.

Ensemble Kalman Filter with Augmented Vector

The EnKF method is briefly described here. Detailed description of the method can be found in Evensen (1994) and Naevdal et al.

(2005). The augmented vector, \mathbf{y}_n , of model parameters and state variables is denoted as

$$\mathbf{y}_n = \begin{bmatrix} \mathbf{m}_n \\ \mathbf{u}_n \end{bmatrix} \quad [5]$$

where \mathbf{m}_n and \mathbf{u}_n are the vectors of model parameters (e.g., hydraulic conductivity) and state variables (e.g., pressure head), respectively, at time t_n . For a set of observations $\mathbf{d}_{\text{obs},n}$, available at time $t = t_n$, its relation with the true but unknown state variable $\mathbf{d}_{\text{true},n}$ and the true augmented state vector $\mathbf{y}_{\text{true},n}$ can be expressed as

$$\mathbf{d}_{\text{obs},n} = \mathbf{d}_{\text{true},n} + \boldsymbol{\varepsilon}_{\text{obs},n} = \mathbf{H}\mathbf{y}_{\text{true},n} + \boldsymbol{\varepsilon}_{\text{obs},n} \quad [6]$$

where $\boldsymbol{\varepsilon}_{\text{obs},n}$ is observation error, which is assumed to be Gaussian with mean $E(\boldsymbol{\varepsilon}_{\text{obs},n}) = 0$ and the covariance of observation error $\mathbf{C}_{\mathbf{D}_n} = E(\boldsymbol{\varepsilon}_{\text{obs},n} \boldsymbol{\varepsilon}_{\text{obs},n}^T)$. It is assumed that the observation error is uncorrelated at different observation times. The observation operator, \mathbf{H} , represents the relation between the augmented state vector and the observation vector.

Ensemble Kalman filter is a Monte Carlo method with N_e ensemble members of \mathbf{y}_n . For each member, the model parameters and state variables in vector \mathbf{y}_n are updated simultaneously by updating the model analysis, \mathbf{y}_n^a , from the model forecast, \mathbf{y}_n^f , via

$$\mathbf{y}_{n,j}^a = \mathbf{y}_{n,j}^f + \mathbf{K}_n (\mathbf{d}_{\text{uc},n,j} - \mathbf{H}\mathbf{y}_{n,j}^f) \quad [7]$$

where the subscript j represents the j th realization of the ensemble, $\mathbf{d}_{\text{uc},n,j}$ is a randomly perturbed observation according to Eq. [6], and \mathbf{K}_n is the Kalman gain, defined as

$$\mathbf{K}_n = \mathbf{C}_{\mathbf{y}_n^f \mathbf{y}_n^f} \mathbf{H}^T (\mathbf{H} \mathbf{C}_{\mathbf{y}_n^f \mathbf{y}_n^f} \mathbf{H}^T + \mathbf{C}_{\mathbf{D}_n})^{-1} \quad [8]$$

and the covariance matrix, $\mathbf{C}_{\mathbf{y}_n^f \mathbf{y}_n^f}$, is estimated by

$$\mathbf{C}_{\mathbf{y}_n^f \mathbf{y}_n^f} \approx \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left[(\mathbf{y}_{n,j}^f - \langle \mathbf{y}_n^f \rangle) (\mathbf{y}_{n,j}^f - \langle \mathbf{y}_n^f \rangle)^T \right] \quad [9]$$

where $\langle \mathbf{y}_n^f \rangle$ denotes the ensemble mean of \mathbf{y}_n^f .

In the study of Hendricks Franssen and Kinzelbach (2008), a damping factor was introduced to reduce the inbreeding problem that results in increasing underestimation of ensemble variance with time. This is achieved by decreasing the magnitude of the update by revising Eq. [7] as

$$\mathbf{y}_{n,j}^a = \mathbf{y}_{n,j}^f + \alpha \mathbf{K}_n (\mathbf{d}_{\text{uc},n,j} - \mathbf{H}\mathbf{y}_{n,j}^f) \quad [10]$$

where α is the damping factor, with values between 0 and 1. Wu and Margulis (2011, 2013) conducted a similar operation but limited the damping factor to the parameter update. Because model parameters and state variables were updated in two separate steps

in this study, the damping factor was used in the manner of Wu and Margulis (2011).

Confirming, Restart, and Modified Restart Ensemble Kalman Filters

Figure 1 illustrates the basic ideas of Confirming EnKF, Restart EnKF, modified Restart EnKF, as well as their major differences from the original EnKF. Confirming EnKF includes in EnKF an additional step, called “confirming”, to ensure that the updated state variables and model parameters are consistent, which, however, may be only “plausibly consistent” (Aanonsen et al., 2009) because Confirming EnKF cannot resolve the inconsistency problem, as shown below. At each assimilation step, Confirming EnKF updates the only model parameter vector \mathbf{m}_n and subsequently uses the updated parameters to compute the state variable vector \mathbf{u}_n at time t_n by running the simulator again from the previous time t_{n-1} via

$$\mathbf{u}_{n,j}^a = f_{n-1 \rightarrow n}(\mathbf{m}_{n,j}^a, \mathbf{u}_{n-1,j}^a) \quad [11]$$

where $f_{n-1 \rightarrow n}$ refers to the model simulator from time $n - 1$ to n .

The major difference between Confirming EnKF and Restart EnKF is the starting time of rerunning the forward model. Instead

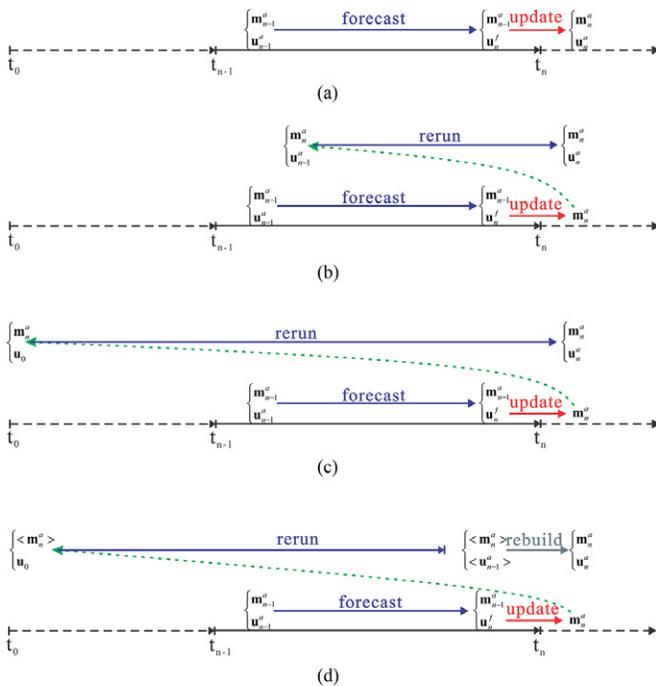


Fig. 1. Flowcharts of (a) original ensemble Kalman filter (EnKF), (b) Confirming EnKF, (c) Restart EnKF, and (d) modified Restart EnKF; t_0 is the initial time, \mathbf{m}_n and \mathbf{u}_n are the vectors of model parameters (e.g., hydraulic conductivity) and state variables (e.g., pressure head), respectively, at time t_n , $\langle \mathbf{m}_n \rangle$ and $\langle \mathbf{u}_n \rangle$ are the means of \mathbf{m}_n and \mathbf{u}_n , respectively, the superscript f indicates forecast and the superscript a indicates assimilated.

of rerunning from the previous time, Restart EnKF reruns from time zero via

$$\mathbf{u}_{n,j}^a = f_{0 \rightarrow n}(\mathbf{m}_{n,j}^a, \mathbf{u}_{0,j}^a) \quad [12]$$

where $f_{0 \rightarrow n}$ refers to the model simulator from time 0 to t_n . This update procedure is different from that of Gu and Oliver (2007) because local optimization for further parameter updating was not conducted in this study due to computational cost. Including the parameter update using local optimization at each assimilation step is expected to improve the performance of Restart EnKF. While Eq. [12] makes Restart EnKF intrinsically consistent, as shown below, the computational cost of Restart EnKF is unaffordable for computationally expensive models because it requires rerunning all the ensemble members from $t = 0$ at each assimilation step.

To improve the computational efficiency of Restart EnKF, a modified Restart EnKF was developed in this study to rerun only the mean of the ensemble. The modification is based on the observation that the ensemble spread is similar in EnKF, Confirming EnKF, and Restart EnKF at any assimilation step, regardless of the prediction values of the three algorithms. In each assimilation step of the modified Restart EnKF, only the model parameters are updated, and the mean of the updated model parameters is approximated by

$$E(\mathbf{m}_n^a) \approx \frac{1}{N_c} \sum_{j=1}^{N_c} \mathbf{m}_{n,j}^a \quad [13]$$

The ensemble mean is used to rerun the simulator from time zero to the current time to estimate an updated mean of the state variables via

$$E(\mathbf{u}_n^{rc}) \approx f_{0 \rightarrow n} E(\mathbf{m}_n^a) \quad [14]$$

Subsequently, a new ensemble of the state variables is constructed by replacing the mean of each realization with the new one in two steps. The first step consists in evaluating the fluctuation of the EnKF ensemble realizations around their mean via

$$\Delta \mathbf{u}_{n,j}^f = \mathbf{u}_{n,j}^f - \frac{1}{N_c} \sum_{j=1}^{N_c} \mathbf{u}_{n,j}^f \quad [15]$$

The second step is to construct the ensemble of the modified Restart EnKF by imposing the updated mean on the fluctuation of the forecast ensemble via

$$\mathbf{u}_{n,j}^a = E(\mathbf{u}_n^{rc}) + \Delta \mathbf{u}_{n,j}^f \quad [16]$$

It should be noted that the modified Restart EnKF is based on two assumptions. First, the state variables computed by the parameter ensemble mean are regarded as an approximation of the state

variable ensemble mean. The other assumption is that parameter variances change slightly in each assimilation step, which is the reason why the fluctuation of the state variable corresponding to the updated parameters can be replaced by that of the forecast. While the second assumption is not difficult to satisfy (because the observation information is gradually incorporated into the assimilation), the first one is strict and is satisfied only when the parameter ensemble mean is the same as the apparent mean obtained through perturbation theories (e.g., Ye et al., 2004) or the effective mean obtained using spectrum theories (e.g., Yeh et al., 1985).

Numerical Experiments

Experiment Design

A synthetic experiment was designed to evaluate the four EnKF algorithms mentioned above in 11 cases (Table 1), with different values of the observation error variance, initial mean and variance, ensemble size, and damping factor, as well as different types of observations and configurations. Figure 2 shows a sketch of the synthetic problem. The size of the two-dimensional domain is 200 by 200 [L²], and the domain is discretized into a grid with 50 by 50 uniform elements, each of which has the size of 4 by 4 [L²]. The bottom and two lateral sides are impervious boundaries, and the top side is composed of steady-state rainfall of 0.2 [L/T]. The initial pressure head is -200 [L] at the top and gradually changes to 0 at the bottom, representing a hydrostatic condition. A reference field of log hydraulic conductivity ($Y = \ln K_s$) was generated by a Karhunen–Loeve expansion (Zhang and Lu, 2004), with mean $\langle Y \rangle = \langle \ln K_s \rangle = 2$, variance $\sigma_Y^2 = 0.7$, and a separated exponential covariance function. The horizontal and vertical correlation lengths of the reference Y field and initial Y realizations were 50 [L] and 20 [L], respectively. The remaining parameters in Eq. [2] and [3] for the unsaturated flow problems were assumed to be deterministic

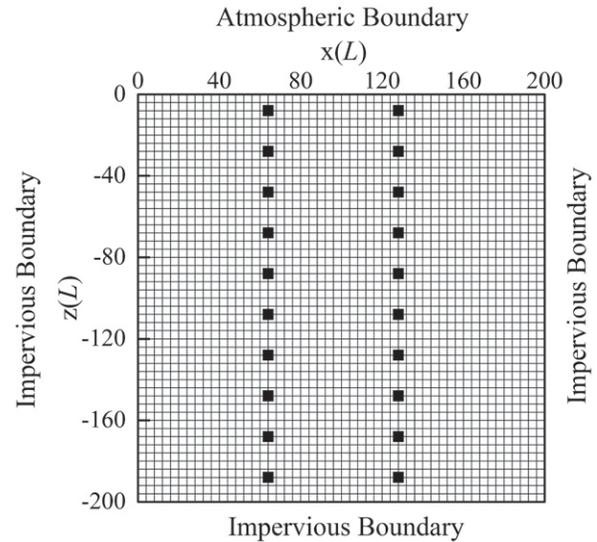


Fig. 2. Study domain, boundary conditions, and 20 locations (squares) for observations of pressure head and moisture content.

constants, with $\theta_r = 0.0001$, $\theta_s = 0.399$, $\alpha = 0.0174$, and $n = 1.3757$. The pressure head, h , and moisture content, θ , observations at 20 locations were drawn from the simulation with the true parameters. The total simulation time was 200 [T]; the initial time step was 0.001 [T], the minimum time step was 0.001 [T], and the maximum time step was 0.5 [T]. The observations every 1 [T] were assimilated.

Evaluation of the Algorithms for the Inconsistency Problem

The performances of the four EnKF algorithms were evaluated in this study using the root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N_g} \sum_{i=1}^{N_g} (\bar{x}_i^a - x_i^t)^2} \quad [17]$$

and Ensemble Spread

$$\text{Ensemble spread} = \sqrt{\frac{1}{N_g} \sum_{i=1}^{N_g} \text{VAR}(x_i^a)} \quad [18]$$

where x_i^t is the reference value of the logarithm of the hydraulic conductivity, Y , or the pressure head, h , \bar{x}_i^a and $\text{VAR}(x_i^a)$ are the ensemble mean and variance, respectively, at each element, and $N_g = 2500$ is the number of elements in the computational grid. The RMSE measures the deviation between the ensemble mean and reference field, and the ensemble spread represents the estimated uncertainty of the ensemble. Lower RMSE indicates better estimation. If the uncertainty of the state is estimated properly, the ensemble spread should be close to the RMSE. The inconsistency (IC)

Table 1. Parameter sets for 11 cases of numerical experiments with observations of pressure head h and water content θ .

Case	Observation error variance	Mean of initial realizations	Variance of initial realizations	No. of realizations	Damping factor	Observations
Reference	–	2.0	0.7	–	–	–
Case 1	0.64	2.0	0.7	1000	1	20 h
Case 2	64	2.0	0.7	1000	1	20 h
Case 3	0.64	2.0	0.4	1000	1	20 h
Case 4	0.64	2.0	1.0	1000	1	20 h
Case 5	0.64	3.0	0.4	1000	1	20 h
Case 6	0.64	3.0	1.0	1000	1	20 h
Case 7	0.64	2.0	0.7	100	1	20 h
Case 8	0.64	2.0	0.7	1000	0.1	20 h
Case 9	0.0025	3.0	1.0	1000	1	20 θ
Case 10	0.64 h , 0.0025 θ	3.0	1.0	1000	1	10 h , 10 θ
Case 11	0.64 h , 0.0025 θ	3.0	1.0	1000	1	10 h , 10 θ

between Y and b is indirectly measured by comparing the mean values before and after rerunning the simulator via

$$IC = \sqrt{\frac{1}{N_g} \sum_{i=1}^{N_g} (\bar{x}_i^a - \bar{x}_i^r)^2} \quad [19]$$

where \bar{x}_i^r is the mean of b computed by rerunning the simulator from time zero. Smaller IC values suggest smaller inconsistency.

The inconsistency problem was investigated in Case 1 (Table 1), in which the mean and variance of the initial realizations of the Y field were set to 2.0 and 0.7, respectively, the same as those of the reference Y field. A relatively large ensemble size of 1000 was used to constrain spurious correlations and the resulting filter divergence, which may happen with a small ensemble size. Figure 3 compares the RMSE and ensemble spread of Y and b for EnKF, Confirming EnKF, Restart EnKF, modified Restart EnKF, and an unconditional run without data assimilation. It shows that the results of Restart EnKF and modified Restart EnKF are similar and significantly better than those of EnKF and Confirming EnKF. The RMSE values of Y for EnKF and Confirming EnKF can even be larger than those of an unconditional model run without data assimilation. After $t = 25$, the RMSE of Y and b increased dramatically for EnKF and Confirming EnKF. This was attributed to inconsistency and its accumulation with time because Fig. 4 shows that the IC values of EnKF and Confirming EnKF increased rapidly after $t = 25$. On the contrary, the IC values are zero for Restart EnKF and remained low for the modified Restart EnKF. Figures 3 and 4 suggest that, while the inconsistency problem cannot be resolved by Confirming EnKF, it can be resolved by Restart EnKF and its modification. Figure 5 plots the spatial distribution of IC at different time steps for EnKF, Confirming EnKF, and modified Restart EnKF. It shows that the inconsistency started at the top

of the domain and gradually moved downward with the wetting front. The inconsistency was the largest for Confirming EnKF and smallest for the modified Restart EnKF.

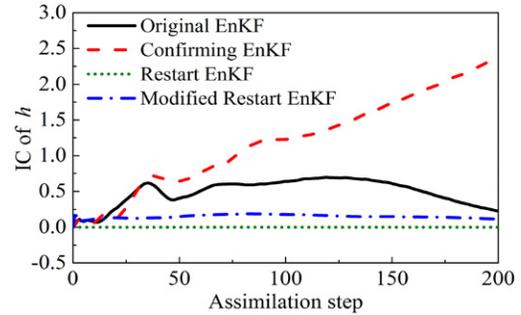


Fig. 4. Inconsistency (IC) values for pressure head b of the four ensemble Kalman filter (EnKF) algorithms in Case 1.

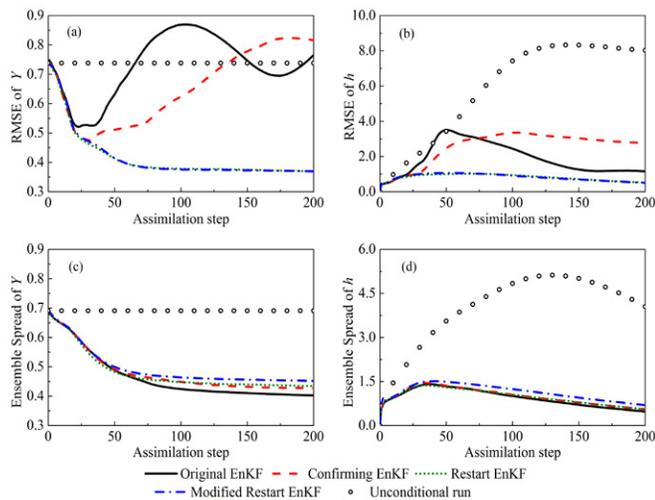


Fig. 3. Comparison of (a) RMSE of Y ($\ln K_G$, the saturated hydraulic conductivity), (b) RMSE of pressure head b , (c) ensemble spread of Y , and (d) ensemble spread of b of four ensemble Kalman filter (EnKF) algorithms and an unconditional run without assimilation in Case 1.

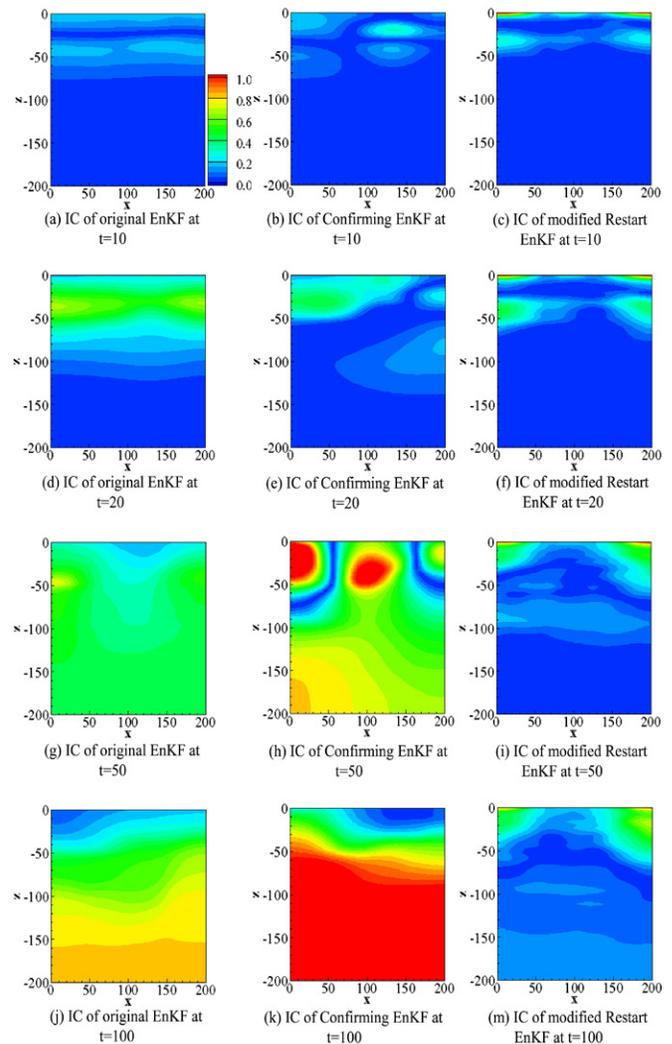


Fig. 5. Inconsistency (IC) values of the original ensemble Kalman filter (EnKF), Confirming EnKF, and modified Restart EnKF at (a–c) time $t = 10$, (d–f) $t = 20$, (g–i) $t = 50$, and (j–l) $t = 100$ in Case 1.

Spatial distributions of the mean and variance of the estimated Y fields of the four EnKF algorithms at $t = 200$ are shown in Fig. 6 and 7. Figure 6 also illustrates the reference and initial Y field. The estimated Y fields were similar for Restart EnKF and its modification; they captured the patterns of the reference field and were better than those of EnKF and Confirming EnKF, which contained a large number of aberrant points. The contours of σ_Y in Fig. 7 are similar for the four methods, which was also true for other time steps, as indicated by the ensemble spread shown in Fig. 3. This suggests that the four methods exhibit similar performance in terms of ensemble spreading, which is the basis of the modified Restart EnKF.

Effects of Data Assimilation Parameters

We investigated the performance of EnKF, Confirming EnKF, Restart EnKF, and the modified Restart EnKF in eight cases (Table 1) under different parameters, i.e., observation error variance, initial mean and variance, ensemble size, and damping factor. The inconsistency under the different parameters and its damage to the assimilation results were particularly analyzed. Case 1 is the base case to mimic the reference case, and the other seven cases were adjusted from Case 1. The rationales of designing the seven

cases are given below. The assimilation data are the pressure head values observed at the locations shown in Fig. 2.

Observation Error Variance

Case 2 was identical to Case 1, except that the observation error variance increased from 0.64 to 64, considering that pressure head measurement errors can be substantial. Equation [7] suggests that a larger observation error variance C_{D_e} will lead to a smaller correction to the forecast. Therefore, a larger observation error may lead to less numerical inconsistency of EnKF in nonlinear problems. Figure 8 shows the values of RMSE and ensemble spread in Cases 1 and 2. The RMSE curves suggest that Restart EnKF and modified Restart EnKF produced poorer Y estimates when the observation error variance was larger, while the results of the original EnKF and Confirming EnKF were somewhat improved due to the alleviated inconsistency. However, the inconsistency problem still existed for EnKF and Confirming EnKF, as shown in Fig. 9, which plots the IC values of Case 2. The ensemble spread curves are higher when measurements with lower precision are used due to the smaller magnitude of correction during the assimilation.

Initial Guess of Ensemble Mean and Variance

In EnKF, the initial realizations of Y are generated based on the prior information (e.g., mean and variance) and may deviate from the true field. The effects of the prior variance on the performance of the four EnKF algorithms were investigated by decreasing the value of the prior variance from 0.7 in Case 1 to 0.4 in Case 3 and increasing it to 1.0 in Case 4; the value of the prior mean remained

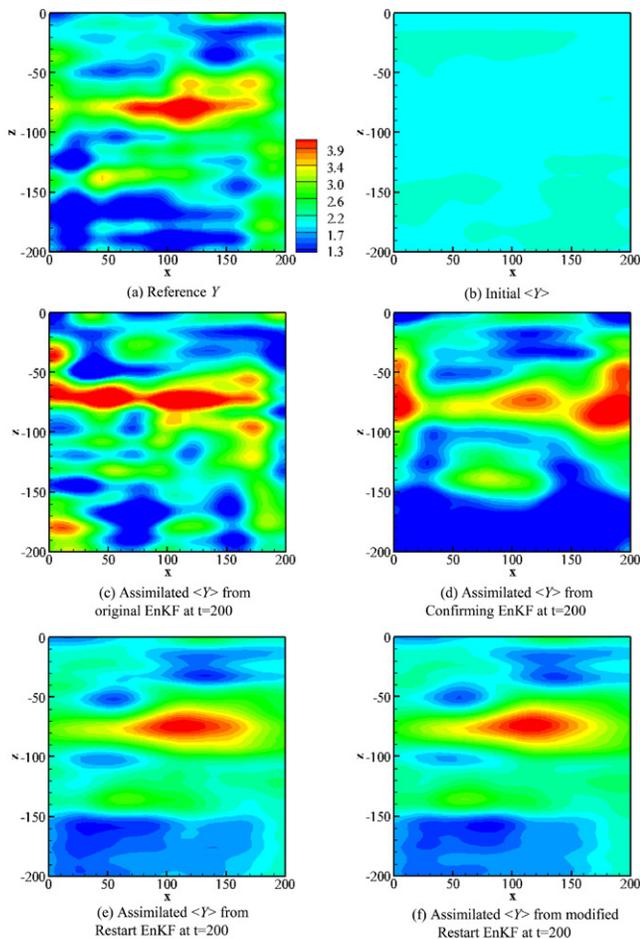


Fig. 6. Spatial distribution of (a) reference Y ($\ln K_s$, the saturated hydraulic conductivity), (b) initial Y , and assimilated Y of (c) original ensemble Kalman filter (EnKF), (d) Confirming EnKF, (e) Restart EnKF, and (f) modified Restart EnKF at time $t = 200$ in Case 1.

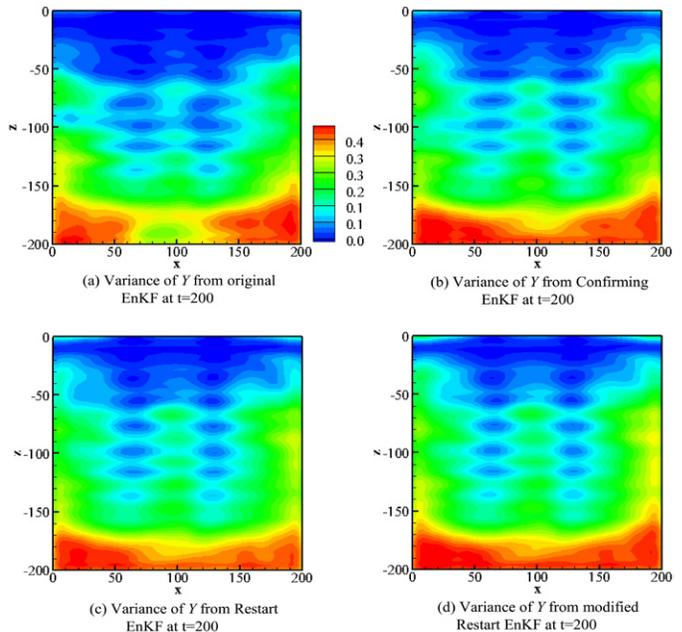


Fig. 7. Spatial distribution of variance of Y ($\ln K_s$, the saturated hydraulic conductivity) for (a) original ensemble Kalman filter (EnKF), (b) Confirming EnKF, (c) Restart EnKF, and (d) modified Restart EnKF at time $t = 200$ in Case 1.

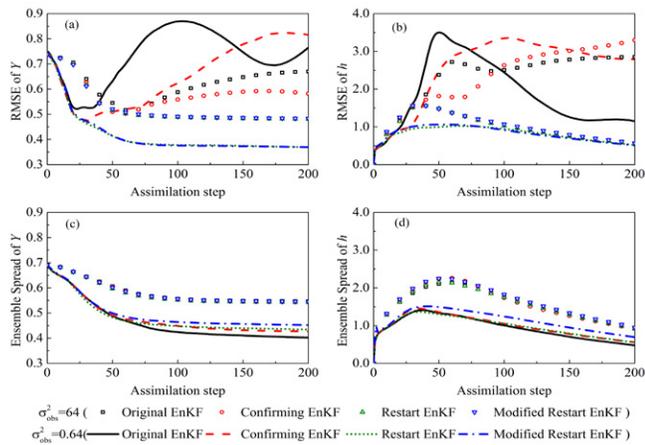


Fig. 8. Comparison of (a) RMSE of Y ($\ln K_S$, the saturated hydraulic conductivity), (b) RMSE of pressure head h , (c) ensemble spread of Y , and (d) ensemble spread of h of four ensemble Kalman filter (EnKF) algorithms in Cases 1 and 2. Observation error variance σ_{obs}^2 increased from 0.64 in Case 1 to 64 in Case 2.

at 2.0. Cases 5 and 6 are identical to Cases 3 and 4, respectively, except that the value of the prior mean increased from 2.0 to 3.0.

Figure 10 compares the RMSE values of Y in Cases 1, 3, and 4. The figure shows that, when there is no rerunning in the original EnKF, the RMSE values are smaller when the initial variance value is larger (Fig. 10a). When the rerunning was for the assimilation step from t_{n-1} to t_n in Confirming EnKF, the RMSE values were similar for the three prior variance values in early assimilation steps but became larger for larger initial values in later steps (Fig. 10b). In this case study, the value of the prior variance barely affected the RMSE curves of Restart EnKF and its modification (Fig. 10c and 10d). There are two reasons for the lower dependence of the prior variance for Restart EnKF and the modified Restart EnKF. The first is that the prior means for Cases 1, 3, and 4 were identical. If the algorithms ran correctly (without divergence or inconsistency), the RMSE curves for Cases 1, 3, and 4 should be close. The second reason is that the given prior variances for Cases 1, 3, and 4 were large enough to capture the possible real Y field. For example, if a very small variance of 0.0001 was given to Case 3, a constant

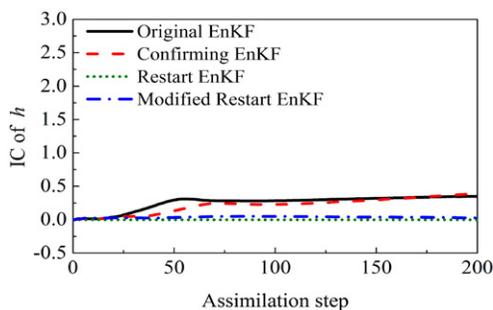


Fig. 9. Inconsistency (IC) values for pressure head h of the four ensemble Kalman filter (EnKF) algorithms in Case 2.

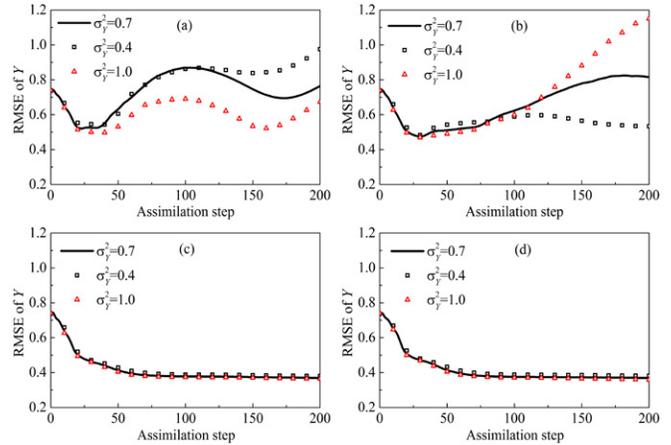


Fig. 10. Comparison of RMSE values of Y ($\ln K_S$, the saturated hydraulic conductivity) for (a) original ensemble Kalman filter (EnKF), (b) Confirming EnKF, (c) Restart EnKF, and (d) modified Restart EnKF with prior variance σ_Y^2 values of 0.7 (Case 1), 0.4 (Case 3), and 1.0 (Case 4).

RMSE would be produced and a large difference between Case 1 and Case 3 would be observed.

Figure 11 compares the RMSE values of Y in Cases 3 to 6, in which the initial mean increased from $\langle Y \rangle = 2.0$ (the true value) in Cases 3 and 4 to $\langle Y \rangle = 3.0$ in Cases 5 and 6. For all four EnKF algorithms, the RMSE values of Cases 5 and 6 are larger than those of Cases 3 and 4, indicating the effects of the initial guess of $\langle Y \rangle$ on the data assimilation results. Because the modified Restart EnKF depends on $\langle Y \rangle$ more than the other three algorithms, caution should be used when the prior mean value is uncertain. In Cases 5 and 6, Restart EnKF and its modification outperformed EnKF and Confirming EnKF because the former two methods have smaller RMSE values. In addition, for Restart EnKF and its modification, because the initial guess of $\langle Y \rangle$ deviates from its true value,

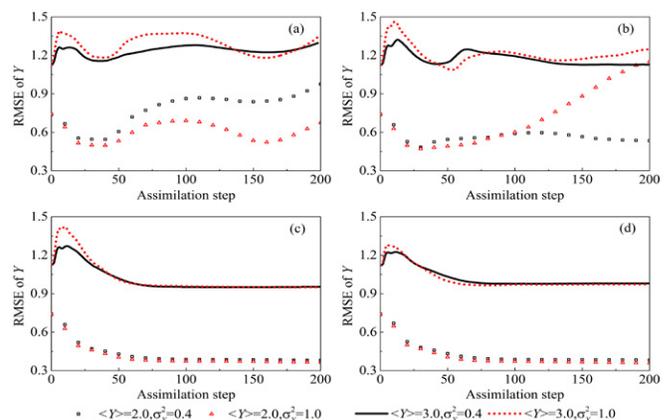


Fig. 11. Comparison of RMSE values of Y ($\ln K_S$, the saturated hydraulic conductivity) for (a) original ensemble Kalman filter (EnKF), (b) Confirming EnKF, (c) Restart EnKF, and (d) modified Restart EnKF with four different sets of prior variance σ_Y^2 values in Cases 3 to 6.

the RMSE curves rise sharply at early time; the RMSE values drop gradually as more measurements are incorporated into the data assimilation. The RMSE values of EnKF and Confirming EnKF are less desirable because the RMSE values of EnKF are larger at later time than at early time and the RMSE values of Confirming EnKF at late time are similar to those at early time.

Ensemble Size

To investigate the impact of ensemble size, Case 1 was resimulated using a reduced ensemble size of 100 (Case 7). To further evaluate the uncertainty of ensemble selection, five different ensembles were used in Case 7. Figure 12 compares the RMSE of the estimated Y with 1000 and 100 realizations. When the ensemble size was reduced to 100, all four EnKF algorithms produced larger RMSE values, indicating worse assimilation performance. In particular, Fig. 12a shows that the original EnKF collapsed for all five smaller ensembles. The performance of Confirming EnKF with 1000 realizations was surprisingly worse than that with 100 realizations. Thus, an improved performance of Confirming EnKF cannot be expected even when a large ensemble size is used due to the impacts of inconsistency. The performance of the RMSE of b (Fig. 13) is better than that of the RMSE of Y in that the RMSE of b does not diverge for EnKF (Fig. 13a). In comparison with the RMSE of Y , the RMSE of b with 100 realizations is close to that with 1000 realizations. In addition, the RMSE curves of b among the five ensembles are more similar than those of Y , especially for Restart EnKF and its modification, which suggests that the ensemble size has a smaller impact on the estimation of b than of Y . This is the strength of the Restart EnKF and its modification because a large ensemble size is computationally impractical in real-world applications of data assimilation.

Effects of Damping Factor

Based on Case 1, Case 8 was designed to consider the impact of the damping factor ($\alpha = 0.1$) on parameter updates according to

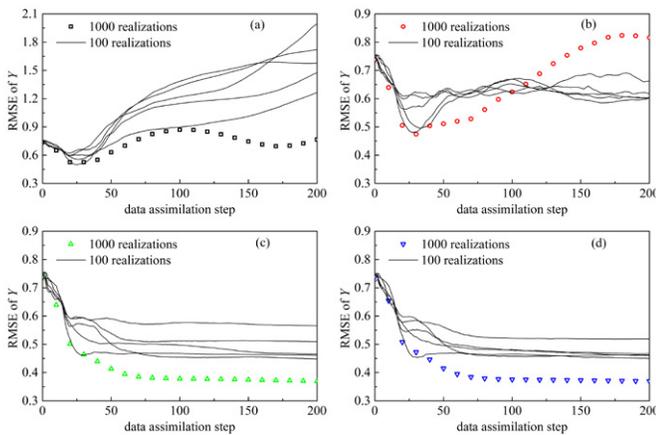


Fig. 12. Comparison of RMSE values of Y ($\ln K_s$, the saturated hydraulic conductivity) for (a) original ensemble Kalman filter (EnKF), (b) Confirming EnKF, (c) Restart EnKF, and (d) modified Restart EnKF with ensemble sizes of 1000 (Case 1) and 100 (Case 7). Five ensembles of 100 realizations were used in Case 7.

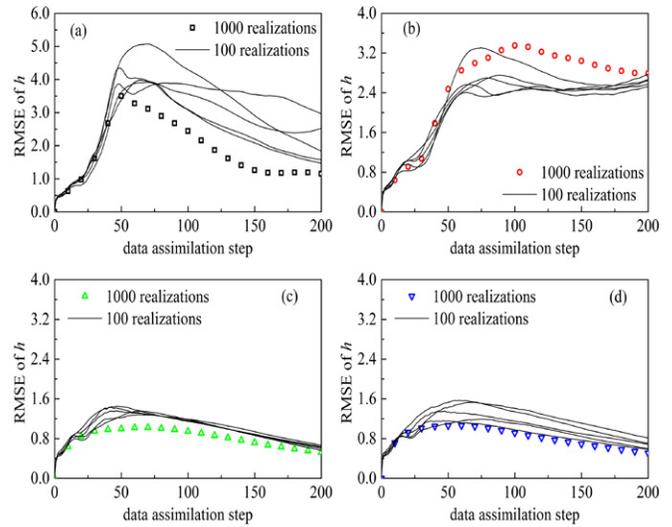


Fig. 13. Comparison of RMSE values of pressure head b for (a) original ensemble Kalman filter (EnKF), (b) Confirming EnKF, (c) Restart EnKF, (d) modified Restart EnKF with ensemble sizes of 1000 (Case 1) and 100 (Case 7). Five ensembles of 100 realizations were used in Case 7.

Eq. [10]. The comparisons of Cases 1 and 8 for all four EnKF algorithms are shown in Fig. 14. Including the damping factor resulted in a significant reduction of the RMSE for EnKF, and the reduced RMSE is close to that of Restart EnKF and its modification. However, the RMSE value of Confirming EnKF with the damping factor increased substantially compared with the one without the damping factor; the RMSE values of Restart EnKF and its modification with the damping factor are also higher than the original ones. Including the damping factor led to a moderate increase in the ensemble spread for the four algorithms. We thus conclude that the damping factor can improve the stability of EnKF but may help little for Confirming EnKF. Including the damping factor also worsened the performance of Restart EnKF and its modification because the observation data value was reduced.

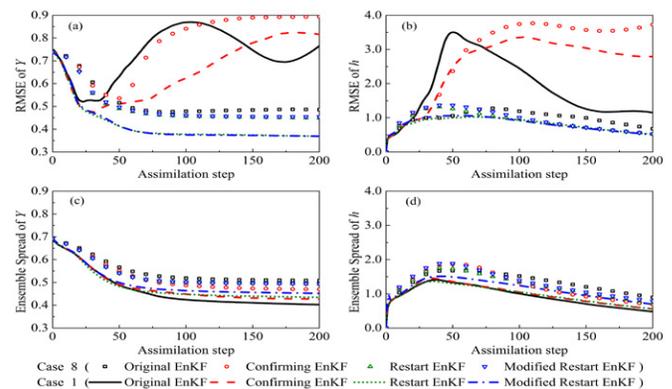


Fig. 14. Comparison of (a) RMSE of Y ($\ln K_s$, the saturated hydraulic conductivity), (b) RMSE of pressure head b , (c) ensemble spread of Y , and (d) ensemble spread of b of the four ensemble Kalman filter (EnKF) algorithms without (Case 1) and with (Case 8) damping factor $\alpha = 0.1$.

Effects of Observation Type

In Cases 1 to 8, the assimilation data were 20 observations of pressure head obtained in the two vertical lines shown in Fig. 2. The effects of different observation types (pressure head and water content) and the configuration of the observations were investigated in Cases 9 to 11. Case 9 was identical to Case 6 but with 20 observations of water content. Because water content can be accurately measured in practice, the observation error variance was set to 0.0025. Cases 10 and 11 were modified from Case 9 by replacing the 20 observations of water contents by 10 observations of pressure head and 10 observations of water content. In Case 10, the observations of water content were above the observations of pressure head; the locations of water content and pressure head were opposite in Case 11.

The performance of the different algorithms for Case 6 and Case 9 are compared in Fig. 15. In Case 6 with pressure head used for assimilation, the original EnKF and Confirming EnKF performed worse than the Restart EnKF and its modification. In Case 8 with water content used for assimilation, however, the RMSE values of Y are visually the same for all the algorithms and the RMSE values of h are similar at early time. This is also the case for the ensemble spread of Y and h . This suggests that using water content for assimilation may be helpful in alleviating the inconsistency problem for the original EnKF and Confirming EnKF because water content is less sensitive than pressure head to soil water movement. For example, a slight change in water content may correspond to a dramatic change in pressure head, especially in dry soil. However, it should be noted that valuable information on water movement may be lost for parameter estimation when using moisture content for assimilation.

Figure 16 for Case 10 is similar to Fig. 15 for Case 9, except for a surprisingly improved RMSE of Y and h observed at the first

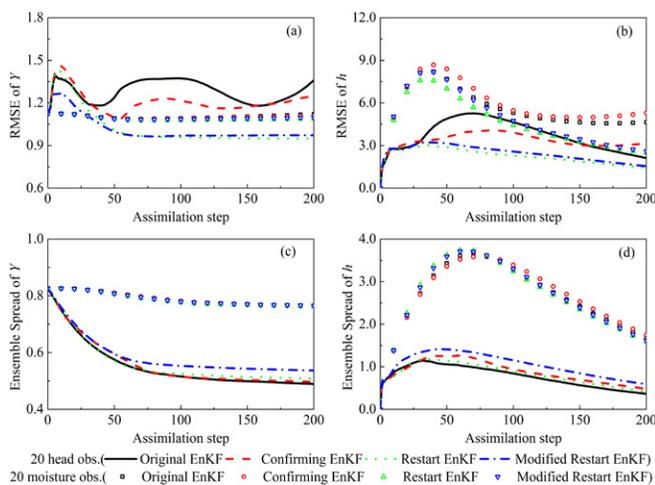


Fig. 15. Comparison of (a) RMSE of Y ($\ln K_s$, the saturated hydraulic conductivity), (b) RMSE of pressure head h , (c) ensemble spread of Y , and (d) ensemble spread of h of the four ensemble Kalman filter (EnKF) algorithms with 20 assimilation data (obs.) for pressure head (Case 6) and water content (Case 9).

100 assimilation steps, even with original EnKF and Confirming EnKF. The RMSE curves of Y for EnKF and Confirming EnKF increase at later times because the inconsistency occurred as the water infiltrated into the deeper soil. Figure 17 for Case 11 does not show any improvement for EnKF and Confirming EnKF; instead, the two algorithms diverge as in Case 6. Comparing all the monitoring strategies, Case 10 (10 water content measurements at the top and 10 head measurements at the subsoil) obtained the best estimation of Y in terms of the smallest RMSE for Restart EnKF and the modified Restart EnKF. Generally, more correction will be given to the parameter estimation when more sensitive data are available. For our investigation, it seems that sensitive data do not always bring benefits to the assimilation because its associated correction may be too large (for example, the early rising in Fig. 16 for Case 6), and incorporating some insensitive data is helpful to improve the performance, especially when the flow zone is undergoing a relatively drastic change in pressure head (for example, the early time of infiltration in our cases).

Comparing Fig. 15 to 17 shows that the RMSE and ensemble spreading of h are significantly improved in Case 11. However, none of alternative monitoring strategies in Cases 9 to 11 yielded better results than the original one in Case 6.

Conclusions

This numerical study aimed at evaluating three algorithms of iterative EnKF (i.e., Confirming EnKF, Restart EnKF, and a modified Restart EnKF) for resolving the inconsistency problem in estimating soil hydraulic parameters of unsaturated flow problems. While Confirming and Restart EnKF were adapted from literature, the modified Restart EnKF was developed in this study to reduce the computational cost by recalculating only the mean simulation, rather than all the ensemble realizations, from time $t = 0$. A

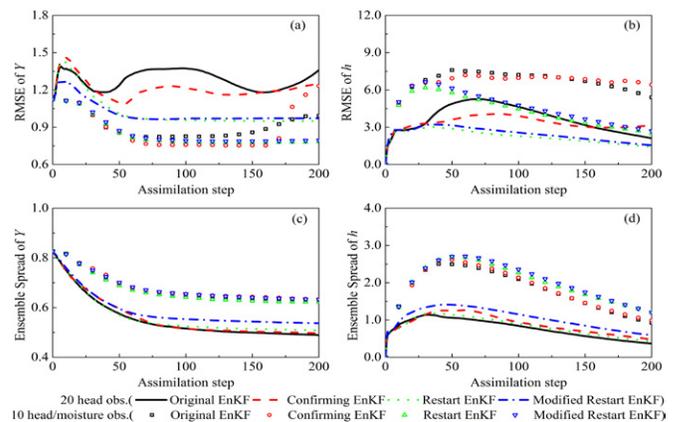


Fig. 16. Comparison of (a) RMSE of Y ($\ln K_s$, the saturated hydraulic conductivity), (b) RMSE of pressure head h , (c) ensemble spread of Y , and (d) ensemble spread of h of the four ensemble Kalman filter (EnKF) algorithms with 20 assimilation data (obs.) for pressure head (Case 6) and 10 data for water content located above 10 data for pressure head (Case 10).

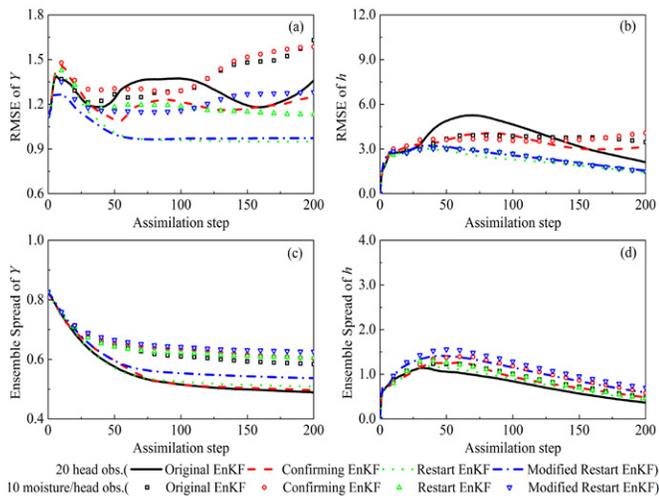


Fig. 17. Comparison of (a) RMSE of Y ($\ln K_s$, the saturated hydraulic conductivity), (b) RMSE of pressure head h , (c) ensemble spread of Y , and (d) ensemble spread of h of the four ensemble Kalman filter (EnKF) algorithms with 20 assimilation data (obs.) for pressure head (Case 6) and 10 data for water content located below 10 data for pressure head (Case 11).

total of 11 cases were designed to investigate the performance of EnKF, Confirming EnKF, Restart EnKF and the modified Restart EnKF algorithms with different values of the observation error variance, the initial guess of ensemble mean and variance, ensemble size, and the damping factor as well as different observation types and configurations of observation data. The major findings of the numerical study are summarized as follows:

1. Although there is a seeming consensus opinion that Confirming EnKF can resolve or alleviate the inconsistency problem, it was found in this study that Confirming EnKF produced considerable inconsistency for the nonlinear unsaturated flow problem in a randomly heterogeneous soil. For the numerical example, the inconsistency problem started at the top of the domain and gradually moved downward with the wetting front.
2. In contrast, Restart EnKF can fully resolve the inconsistency problem and produced the best results of data assimilation among the four EnKF algorithms tested in this study. While the local optimization of Gu and Oliver (2007) was not conducted in this study, it is not expected to alter this conclusion. Modified Restart EnKF can substantially reduce computational costs and produce assimilation results as accurate as those of Restart EnKF.
3. Using data of higher observation variance may alleviate the inconsistency problem to a certain extent for EnKF and Confirming EnKF due to the small amount of correction during the assimilation.
4. The original EnKF and Confirming EnKF are more likely to produce erratic Y RMSE values under different initial prior variances. However, Restart EnKF and its modification work more stably regardless of the initial variance.
5. When the initial guess of the ensemble mean of Y deviates from the true value, its impacts on Restart EnKF and its modification are smaller than on EnKF and Confirming EnKF, suggesting robustness of Restart EnKF.

6. When the ensemble size decreases, performance deterioration is less for Restart EnKF and its modification than for EnKF and Confirming EnKF.
7. Including the damping factor can improve the stability of EnKF but not Confirming EnKF. However, it may deteriorate the performance of Restart EnKF and its modification due to the loss of data value.
8. Due to the rapid soil water movement driven by atmospheric boundary conditions, using moisture content (instead of pressure head) observations in the topsoil is helpful to maintain stability of the assimilation process. On the contrary, using more sensitive data like pressure head in the subsoil can better capture the details of water movement. Combining different types of observations can achieve better assimilation results.

Although we preliminarily discussed the application of iterative EnKF algorithms for parameter estimation in the vadose zone and fully investigated the consistency introduced by nonlinear unsaturated flow, it should be noted that the present work was based on several important assumptions, such as greatly simplified boundary conditions (e.g., the constant rain rate of the upper boundary and the impervious lower boundary), a relatively small variance of the saturated hydraulic conductivity and a known variogram, limited anisotropy, and other predetermined parameters in the van Genuchten–Mualem equation. A more comprehensive study considering effects like boundary conditions and other parameters is still required in the future.

Acknowledgments

This study is supported by Natural Science Foundation of China through Grants 51079101, 51179132, and 51279141 and by the Fundamental Research Funds for the Central Universities through Grant 2012206020216. Ming Ye is supported by the DOE Early Career Award DE-SC0008272. We also thank Jasper Vrugt and the anonymous reviewers for their thoughtful comments.

References

- Aanonsen, S., G. Nævdal, D. Oliver, A. Reynolds, and B. Vallès. 2009. The ensemble Kalman filter in reservoir engineering: A review. *SPE J.* 14:393–412. doi:10.2118/117274-PA
- Abbasi, F., D. Jacques, J. Šimůnek, J. Feyen, and M.Th. van Genuchten. 2003a. Inverse estimation of soil hydraulic and solute transport parameters from transient field experiments: Heterogeneous soil. *Trans. ASAE* 46:1097–1111.
- Abbasi, F., J. Šimůnek, J. Feyen, M.Th. van Genuchten, and P.J. Shouse. 2003b. Simultaneous inverse estimation of soil hydraulic and solute transport parameters from transient field experiments: Homogeneous soil. *Trans. ASAE* 46:1085–1095.
- Abbaspour, K., R. Kasteel, and R. Schulin. 2000. Inverse parameter estimation in a layered unsaturated field soil. *Soil Sci.* 165:109–123. doi:10.1097/00010694-200002000-00002
- Anderson, J.L. 2010. A non-Gaussian ensemble filter update for data assimilation. *Mon. Weather Rev.* 138:4186–4198. doi:10.1175/2010MWR3253.1
- Burgers, G., P.J. van Leeuwen, G. Evensen, and P. van Leeuwen. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* 126:1719–1724. doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2
- Chen, Y. 2012. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Math. Geosci.* 44:1–26. doi:10.1007/s11004-011-9376-z
- Elsheikh, A.H., M.F. Wheeler, and I. Hoteit. 2013. An iterative stochastic ensemble method for parameter estimation of subsurface flow models. *J. Comput. Phys.* 242:696–714. doi:10.1016/j.jcp.2013.01.047
- Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* 99(C5):10143–10162. doi:10.1029/94JC00572

- Evensen, G. 2009. The ensemble Kalman filter for combined state and parameter estimation: Monte Carlo techniques for data assimilation in large systems. *IEEE Contr. Syst. Mag.* 29:83–104. doi:10.1109/MCS.2009.932223
- Gu, Y., and D.S. Oliver. 2006. The ensemble Kalman filter for continuous updating of reservoir simulation models. *J. Energy Resour. Technol.* 128:79–87. doi:10.1115/1.2134735
- Gu, Y., and D. Oliver. 2007. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE J.* 12:438–446. doi:10.2118/108438-PA
- Hendricks Franssen, H.J., and W. Kinzelbach. 2008. 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):W09408. doi:10.1029/2007WR006505
- Hoteit, I., D. Pham, G. Triantafyllou, and G. Korres. 2008. Particle Kalman filtering for data assimilation in meteorology and oceanography. In: *Proceedings of the 3rd WCRP International Conference on Reanalysis*, Tokyo, 28 Jan.–1 Feb. 2008. Univ. of Tokyo, Tokyo, p. 6.
- Jacques, D., J. Šimůnek, A. Timmerman, and J. Feyen. 2002. Calibration of Richards' and convection–dispersion equations to field-scale water flow and solute transport under rainfall conditions. *J. Hydrol.* 259:15–31. doi:10.1016/S0022-1694(01)00591-1
- Jazwinski, A.H.H.A.H. 1970. *Stochastic processes and filtering theory*. Academic Press, San Diego.
- Johns, C.J., and J. Mandel. 2007. A two-stage ensemble Kalman filter for smooth data assimilation. *Environ. Ecol. Stat.* 15:101–110. doi:10.1007/s10651-007-0033-0
- Krymskaya, M.V., R.G. Hanea, and M. Verlaan. 2008. An iterative ensemble Kalman filter for reservoir engineering applications. *Comput. Geosci.* 13:235–244. doi:10.1007/s10596-008-9087-9
- Li, C., and L. Ren. 2011. Estimation of unsaturated soil hydraulic parameters using the ensemble Kalman filter. *Vadose Zone J.* 10:1205–1227. doi:10.2136/vzj2010.0159
- Li, G., and A. Reynolds. 2007. An iterative ensemble Kalman filter for data assimilation. In: *Proceedings of the SPE Annual Technical Conference and Exhibition*, Anaheim, CA, 11–14 Nov. 2007. Soc. Pet. Eng., Richardson, TX. SPE Paper 109808. doi:10.2118/109808-MS
- Li, H., S. Chen, D. Yang, and P. Tontiwachwuthikul. 2010. Ensemble-based relative permeability estimation using B-spline model. *Transp. Porous Media* 85:703–721. doi:10.1007/s11242-010-9587-7
- Lorentzen, R.J., and G. Naevdal. 2011. An iterative ensemble Kalman filter. *IEEE Trans. Automat. Contr.* 56:1990–1995. doi:10.1109/TAC.2011.2154430
- Lü, H., Z. Yu, Y. Zhu, S. Drake, Z. Hao, and E.A. Sudicky. 2011. Dual state-parameter estimation of root zone soil moisture by optimal parameter estimation and extended Kalman filter data assimilation. *Adv. Water Resour.* 34:395–406. doi:10.1016/j.advwatres.2010.12.005
- Mirus, B.B., K.S. Perkins, J.R. Nimmo, and K. Singha. 2009. Hydrologic characterization of desert soils with varying degrees of pedogenesis: 2. Inverse modeling for effective properties. *Vadose Zone J.* 8:496–509. doi:10.2136/vzj2008.0051
- Montzka, C., H. Moradkhani, L. Weihermüller, H.-J.H. Franssen, M. Canty, and H. Vereecken. 2011. Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. *J. Hydrol.* 399:410–421. doi:10.1016/j.jhydrol.2011.01.020
- Moradkhani, H. 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resour. Res.* 41:1–17.
- Moradkhani, H., S. Sorooshian, H.V. Gupta, and P.R. Houser. 2005. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Adv. Water Resour.* 28:135–147. doi:10.1016/j.advwatres.2004.09.002
- Naevdal, G., L. Johnsen, S. Aanonsen, and E. Vefring. 2005. Reservoir monitoring and continuous model updating using ensemble Kalman filter. *SPE J.* 10:66–74. doi:10.2118/84372-PA
- Navon, I.M. 1998. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dyn. Atmos. Oceans* 27:55–79. doi:10.1016/S0377-0265(97)00032-8
- Reichle, R.H. 2008. Data assimilation methods in the Earth sciences. *Adv. Water Resour.* 31:1411–1418. doi:10.1016/j.advwatres.2008.01.001
- Reynolds, A.C., M. Zafari, G. Li, and R.A. Zafari. 2006. Iterative forms of the ensemble Kalman filter. In: *Proceedings of the 10th European Conference on the Mathematics of Oil Recovery*, Amsterdam, 4–7 Sept. 2006. Eur. Assoc. Geosci. Eng., Houten, the Netherlands, p. A030.
- Rings, J., J.A. Vrugt, G. Schoups, J.A. Huisman, and H. Vereecken. 2012. Bayesian model averaging using particle filtering and Gaussian mixture modeling: Theory, concepts, and simulation experiments. *Water Resour. Res.* 48:W05520. doi:10.1029/2011WR011607
- Sakov, P., D.S. Oliver, and L. Bertino. 2012. An iterative EnKF for strongly nonlinear systems. *Mon. Weather Rev.* 140:1988–2004. doi:10.1175/MWR-D-11-00176.1
- Šimůnek, J., T. Vogel, and M.Th. van Genuchten. 1994. The SWMS_2D code for simulating water flow and solute transport in two-dimensional variably saturated media (Version 1.21). U.S. Salinity Lab., Riverside, CA.
- Thulin, K., G. Li, S. Aanonsen, and A. Reynolds. 2007. Estimation of initial fluid contacts by assimilation of production data with EnKF. In: *Proceedings of the SPE Annual Technical Conference and Exhibition*, Anaheim, CA, 11–14 Nov. 2007. Soc. Pet. Eng., Richardson, TX. SPE 109975. doi:10.2118/109975-MS
- van Genuchten, M.Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* 44:892–898. doi:10.2136/sssaj1980.03615995004400050002x
- Verbist, K., W.M. Cornelis, D. Gabriels, K. Alaerts, and G. Soto. 2009. Using an inverse modeling approach to evaluate the water retention in a simple water harvesting technique. *Hydrol. Earth Syst. Sci. Discuss.* 6:4265–4306. doi:10.5194/hessd-6-4265-2009
- Vrugt, J.A., C.G.H. Diks, H.V. Gupta, W. Bouten, and J.M. Verstraten. 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resour. Res.* 41:W01017. doi:10.1029/2004WR003059
- Vrugt, J.A., P.H. Stauffer, Th. Wöhling, B.A. Robinson, and V.V. Vesselinov. 2008. Inverse modeling of subsurface flow and transport properties: A review with new developments. *Vadose Zone J.* 7:843–864. doi:10.2136/vzj2007.0078
- Wang, Y., G. Li, and A. Reynolds. 2010. Estimation of depths of fluid contacts by history matching using iterative ensemble-Kalman smoothers. *SPE J.* 15:509–525. doi:10.2118/119056-PA
- Wen, X.-H., and W. Chen. 2006. Real-time reservoir model updating using ensemble Kalman filter with confirming option. *SPE J.* 11:431–442.
- Wöhling, T., and J.A. Vrugt. 2011. Multiresponse multilayer vadose zone model calibration using Markov chain Monte Carlo simulation and field water retention data. *Water Resour. Res.* 47:W04510. doi:10.1029/2010WR009265
- Wu, C.-C., and S.A. Margulis. 2011. Feasibility of real-time soil state and flux characterization for wastewater reuse using an embedded sensor network data assimilation approach. *J. Hydrol.* 399:313–325. doi:10.1016/j.jhydrol.2011.01.011
- Wu, C.-C., and S.A. Margulis. 2013. Real-time soil moisture and salinity profile estimation using assimilation of embedded sensor datastreams. *Vadose Zone J.* 12(1). doi:10.2136/vzj2011.0176
- Ye, M., S.P. Neuman, A. Guadagnini, and D.M. Tartakovsky. 2004. Nonlocal and localized analyses of conditional mean transient flow in bounded, randomly heterogeneous porous media. *Water Resour. Res.* 40:W05104. doi:10.1029/2003WR002099
- Yeh, T.-C.J., L.W. Gelhar, and A.L. Gutjahr. 1985. Stochastic analysis of unsaturated flow in heterogeneous soils: 1. Statistically isotropic media. *Water Resour. Res.* 21:447–456. doi:10.1029/WR021i004p00447
- Zafari, M., and A. Reynolds. 2007. Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter. *SPE J.* 12:382–391. doi:10.2118/95750-PA
- Zagayevskiy, Y., A.H. Hosseini, and C.V. Deutsch. 2012. Constraining a heavy oil reservoir to temperature and time lapse seismic data using the EnKF. In: P. Abrahamson et al., editors, *Geostatistics Oslo 2012*. Quant. Geol. Geostat. 17. Springer, Dordrecht, the Netherlands, p. 145–158.
- Zhang, D., and Z. Lu. 2004. An efficient, high-order perturbation approach for flow in random porous media via Karhunen–Loève and polynomial expansions. *J. Comput. Phys.* 194:773–794. doi:10.1016/j.jcp.2003.09.015
- Zhang, Y., H. Li, and D. Yang. 2012. Simultaneous estimation of relative permeability and capillary pressure using ensemble-based history matching techniques. *Transp. Porous Media* 94:259–276. doi:10.1007/s11242-012-0003-3
- Zhang, Z.F., A.L. Ward, and G.W. Gee. 2003. Estimating soil hydraulic parameters of a field drainage experiment using inverse techniques. *Vadose Zone J.* 2:201–211.
- Zupanski, M., and F. Collins. 2005. Maximum likelihood ensemble filter: Theoretical aspects. *Mon. Weather Rev.* 133:1710–1726. doi:10.1175/MWR2946.1