Stoch Environ Res Risk Assess (2009) 23:1155–1167
DOI 10.1007/s00477-008-0289-4

ORIGINAL PAPER

Using data assimilation method to calibrate a heterogeneous conductivity field and improve solute transport prediction with an unknown contamination source

Chunlin Huang · Bill X. Hu · Xin Li · Ming Ye

Published online: 20 November 2008

Abstract Hydraulic conductivity distribution and plume initial source condition are two important factors affecting solute transport in heterogeneous media. Since hydraulic conductivity can only be measured at limited locations in a field, its spatial distribution in a complex heterogeneous medium is generally uncertain. In many groundwater contamination sites, transport initial conditions are generally unknown, as plume distributions are available only after the contaminations occurred. In this study, a data assimilation method is developed for calibrating a hydraulic conductivity field and improving solute transport prediction with unknown initial solute source condition. Ensemble Kalman filter (EnKF) is used to update the model parameter (i.e., hydraulic conductivity) and state variables (hydraulic head and solute concentration), when data are available. Two-dimensional numerical experiments are designed to assess the performance of the EnKF method on data assimilation for solute transport prediction. The study results indicate that the EnKF method can significantly improve the estimation of the hydraulic conductivity distribution and solute transport prediction by assimilating hydraulic head measurements with a known solute initial condition. When solute source is unknown, solute prediction by assimilating continuous measurements of solute concentration at a few points in the plume well captures the plume evolution downstream of the measurement points.

Keywords Data assimilation · Ensemble Kalman filter · Solute transport · Hydraulic conductivity · Steady-state flow

C. Huang · X. Li
Cold and Arid Regions Environmental and Engineering Research Institute, CAS, 730000 Lanzhou, China

B. X. Hu (✉) · M. Ye
Department of Geological Sciences, Florida State University, Tallahassee, FL 32306, USA
e-mail: hu@gly.fsu.edu

M. Ye
Department of Scientific Computing, Florida State University, Tallahassee, FL 32306, USA

1 Introduction

The effective management of subsurface contamination requires development and application of groundwater flow and solute transport models to accurately predict contaminant distribution and migration in complex subsurface environments. Natural media are generally heterogeneous and hydraulic parameters vary significantly in space. Characterization of field heterogeneity is always difficult due to limited project budget, time, and/or available measurement techniques. We always face a problem that parameter measurements are too sparse to describe spatial distribution of hydraulic parameters. To resolve this problem, inverse methods are commonly used to calibrate parameter distributions and provide more reliable model predictions, as described in several review articles (McLaughlin and Townley 1996; Zimmerman et al. 1998; Carrera et al. 2005). In particular, it is vital to utilize various kinds of observations, which contain different levels of signature of site heterogeneity and can help resolve the non-uniqueness problem of inverse modeling (Poeter and Hill 1997). Many time-invariant inverse algorithms have been developed for this purpose using cokriging methods (Hoeksema and Kitanidis 1984; Sun 1994; Yeh and Zhang 1996; Yeh and Liu 2000; Zhu and Yeh 2005). In the conventional inverse methods, the calibration data (e.g.,

measurements of state variables) are commonly used in a batch, not sequential, mode. This diminishes flexibility of model calibration and increases computational cost (e.g., in calculation of the sensitivity matrix). More importantly, using calibration data in the batch model significantly reduces potential of eliminating influence of incorrect initial conditions on model predictions. With improvement of measurement technology, more and more continuous groundwater observations become available. Suitable approaches should be invoked to dynamically reconcile these temporal observations of different kinds. The data assimilation method is one of these kind approaches.

Originating from meteorology and oceanography (Daley 1991), data assimilation method has been developed for improvement of operational weather forecasts and ocean dynamics prediction (Bennett 1992). In hydrology, data assimilation methods have been used to incorporate remote sensing data dynamically to improve estimation of land surface variables. One primary application is to estimate soil moisture or soil temperature profile in vertical direction by assimilating remote sensing data or observation in situ (Houser et al. 1998; Reichle et al. 2002a, b; Li et al. 2004; Huang et al. 2008a, b). Data assimilation methods have been applied to assimilate geophysical data to characterize a medium heterogeneity (Christakos 2002, 2005). The method has also used in hydrodynamic modeling to deal with nonlinearity and bias (Sorensen et al. 2004). Liu and Gupta (2007) recently provided a review on applications of data assimilation in hydrology with a focus on uncertainty analysis.

Kalman filter (KF) (Kalman 1960) is a method of sequential data assimilation that recursively assimilates observations when they become available with time. While KF was developed for linear systems, Ensemble Kalman filter (EnKF) (Evensen 2003, 2006) was developed for nonlinear systems without model linearization as in the Extended Kalman Filter (EKF) (Jazwinski 1970). The KF and EKF have been applied to improve the distribution of soil moisture and other quantities (McLaughlin 1995). The EnKF was applied to estimate soil moisture profile (Reichle et al. 2002a) and its advantage over EKF was also assessed (Reichle et al. 2002b). Margulis et al. (2002) utilized the EnKF to assimilate land data obtained from field experiments at the Southern Great Plains to estimate soil moisture. Andreadis and Lettenmaier (2006) used the EnKF to assimilate remotely sensed snow observation into a macroscale hydrology model to improve the prediction for snow cover prediction. Clark et al. (2006) applied the EnKF to assimilate snow covered area information into hydrologic and land surface models to improve streamflow simulation. Huang et al. (2008a, b) adopted the EnKF to assimilate remote sensing data into land surface model to improve soil moisture and soil temperature profile estimation. The KF, EKF, and EnKF have received more and more attention in groundwater, surface hydrology, and petroleum reservoir modeling (Van Geer et al. 1991; Yangxiao et al. 1991; Vrugt et al. 2005a, b; Gu and Oliver 2005, 2006; Drecourt et al. 2006; Wen and Chen 2005; Chen and Zhang 2006; Zhang et al. 2007). Recent development of the EnKF is not limited to updating system state variables, as in conventional data assimilation, but simultaneously updates state variables and model parameters to yield more accurate model predictions (Moradkhani et al. 2005).

Current applications of the EnKF in groundwater modeling are limited to flow models. Utilizing the EnKF method to assimilate concentration data for prediction of solute transport in groundwater is a new research area. Vrugt et al. (2005a, b) applied a SODA method to couple parameter optimization and sequential data assimilation for estimation of sorption parameters by using Li concentration at a production well. The SODA method estimates the parameters in an outer loop using a global optimization method in a batch mode and then conducts data assimilation in an inner loop using the EnKF. In this study, by using a synthetic case for demonstration purpose, we apply an EnKF method jointly, rather than separately, to estimate model parameters and conduct data assimilation. The study results demonstrate the feasibility of using the EnKF to gradually reduce influence of unknown source term on model prediction. In our best knowledge, it is the first time to address this issue in groundwater modeling.

Predicting evolution of a solute plume requires knowing not only the spatial distributions of hydraulic and transport properties but also solute source such as its location and solute release function. In many field cases, when groundwater contamination is found in one or several locations (e.g., monitoring wells), solute plume has already formed. The locations of the contaminate sources may be unknown, or if the source locations are identified, the solute release functions are generally unknown. Under the condition of unknown solute source, it is still an open question that how to use concentration measurements at limited monitoring wells to improve the prediction of solute plume distribution in future or solute concentration at downstream of the monitoring wells. In this study we address this issue by using the EnKF in a manner similar to its applications in weather prediction. By assimilating concentration observations, solute plume distribution will be adjusted and long-term behavior of the plume can be accurately simulated. This is a unique property of the data assimilation method, not shared by conventional inverse methods that use calibration data in the batch mode. It is worth mentioning that in this study, we focus on gradually reduce influence of unknown source term on transport predictions using the EnKF, do not intend to characterize the source term as in the backward probability model of

Neupauer and Wilson (2001) and Neupauer and Lin (2006).

2 Data assimilation method

A data assimilation system is composed of a model operator, an observation operator, and a data assimilation algorithm. For a groundwater transport system, the model operator is usually a flow-transport model used to simulate the spatial-temporal evolution of hydraulic head and solute concentration. In this study, the model operator is a steady-state flow and a transient transport model. The observation operator is used to build the relationship between simulated state variables and observations. In this study, the state variables and observations are the same, i.e., hydraulic head and concentration at observation wells. The data assimilation algorithm used in this study is the EnKF, which utilizes observations of head and concentration to update the state variables produced by the flow and transport model.

2.1 Flow and transport models

The classical convection-dispersion equation is adopted as the governing equation for the nonreactive chemical transport (Zheng and Wang 1999; Zheng and Bennett 2002), which is expressed as

$$\frac{\partial(\theta C^k)}{\partial t} = \frac{\partial C}{\partial x_i}\left(\theta D_{ij}\frac{\partial C^k}{\partial x_j}\right) - \frac{\partial}{\partial x_i}\left(\theta v_i C^k\right) + q_s C_s^k + \sum R_n \tag{1}$$

where C^k (M L^{-3}) is dissolved concentration, D_{ij} (L^2 T^{-1}) is hydrodynamic dispersion coefficient tensor, $v_i = \frac{q_i}{\theta}$ (L T^{-1}) is seepage or average linear pore water velocity, θ (–) and q_i (L T^{-1}) being porosity and Darcy velocity, q_s (T^{-1}) is volumetric flow rate per unit volume of aquifer due to fluid sources and sinks, and C_s^k (M L^{-3}) is concentration of the source or sink flux. $\sum R_n$ (M L^{-3} T^{-1}) is chemical reaction term. Here, L, T and M denote any consistent unit of length, time and mass. The Darcy velocity, q_i, is estimated via the Darcy's law

$$q_i = -K_{ij}\frac{\partial h}{\partial x_j} \tag{2}$$

where K_{ij} (L T^{-1}) is the hydraulic conductivity tensor and for an isotropic medium, $K_{ij} = 0$ $(i \neq j)$, and h (L) is hydraulic head. The groundwater seepage velocity, v_i, is related to the specific discharge as

$$v_i = \frac{q_i}{\theta} = -\frac{K_{ij}}{\theta}\frac{\partial h}{\partial x_j} \tag{3}$$

where θ (L^3/L^3) is the medium's porosity. Hydraulic head is obtained by solving the groundwater flow equation

$$\frac{\partial}{\partial x_i}\left(K_{ij}\frac{\partial h}{\partial x_j}\right) = S_s\frac{\partial h}{\partial t} \tag{4}$$

where S_s (L^{-1}) is specific storage of the aquifer. If the flow is in steady state, the term on right-hand side will be zero. This study focus on groundwater flow and non-reactive transport modeling, chemical reactions are not considered, but can be integrated in future study.

2.2 Ensemble Kalman filter

To make this paper self-contained, we briefly introduce the EnKF method with perturbed observation proposed by Burgers et al. (1998). In this algorithm, predictions of state variables are given by their ensembles. Assuming normal distribution of model predictions, the ensemble mean is supposed to be the best estimate of the true state, and prediction error around the mean is measured by covariance of the ensemble (Evensen 2003). The covariance, P, of forecast and analysis error of a random variable x are defined as

$$P^{\mathrm{f}} \cong P_{\mathrm{e}}^{\mathrm{f}} = \overline{(x^{\mathrm{f}} - \bar{x}^{\mathrm{f}})(x^{\mathrm{f}} - \bar{x}^{\mathrm{f}})^{\mathrm{T}}} \tag{5}$$

$$P^{\mathrm{a}} \cong P_{\mathrm{e}}^{\mathrm{a}} = \overline{(x^{\mathrm{a}} - \bar{x}^{\mathrm{a}})(x^{\mathrm{a}} - \bar{x}^{\mathrm{a}})^{\mathrm{T}}} \tag{6}$$

where the overbar denotes the ensemble mean, and the superscripts a and f refer to analysis and forecast, respectively, the two necessary steps of the EnKF. In the forecast step, forecasted state variables of each ensemble member is updated according to

$$X_{i,t+1}^{\mathrm{f}} = M\left(X_{i,t}^{\mathrm{a}}\right) + u_i \quad u_i \sim N(0, Q) \tag{7}$$

where $X_{i,t+1}^{\mathrm{f}}$ is the forecasted state variable of the ith ensemble member at time $t+1$; $X_{i,t}^{\mathrm{a}}$ is analyzed state variable of the ith ensemble member at time t; $M(\cdot)$ is model operator, which is the flow or/and transport model in groundwater modeling; u_i is model error vector, which is assumed to satisfy Gaussian distribution with zero mean and covariance matrix Q.

In the analysis step, the observation data are perturbed by adding random observation errors. The forecast of each ensemble member is updated as follows (Burgers et al. 1998):

$$X_{i,t+1}^{\mathrm{a}} = X_{i,t+1}^{\mathrm{f}} + P_{t+1}^{\mathrm{f}}H^{\mathrm{T}}\left[HP_{t+1}^{\mathrm{f}}H^{\mathrm{T}} + R\right]^{-1} \left[(Y_{t+1} + \varepsilon_i) - H\left(X_{i,t+1}^{\mathrm{f}}\right)\right] \quad \varepsilon_i \sim N(0, R) \tag{8}$$

where $H(\cdot)$ is observation operator used to convert the model state variables to observations; Y_{t+1} is observation data at time $t+1$; ε_i is random error vector of observation with zero mean and covariance matrix R; P_{t+1}^{f} is forecast background covariance matrix at time $t+1$. Based on Eq. (5), replacing the ensemble covariance by sample

covariance of N ensemble members, $P^{f}_{t+1}H^{T}$ and $HP^{f}_{t+1}H^{T}$ in Eq. (8) can be calculated via (Burgers et al. 1998)

$$P^{f}_{t+1}H^{T} = \frac{1}{N-1}\sum_{i=1}^{N}\left[X^{f}_{i,t+1} - \overline{X^{f}_{t+1}}\right]\left[H\left(X^{f}_{i,t+1}\right) - H\left(\overline{X^{f}_{t+1}}\right)\right]^{T} \quad (9)$$

$$HP^{f}_{t+1}H^{T} = \frac{1}{N-1}\sum_{i=1}^{N}\left[H\left(X^{f}_{i,t+1}\right) - H\left(\overline{X^{f}_{t+1}}\right)\right]\left[H\left(X^{f}_{i,t+1}\right) - H\left(\overline{X^{f}_{t+1}}\right)\right]^{T} \quad (10)$$

The analysis state estimate at time $t + 1$ is given by the mean of the ensemble members. The analyzed ensemble is then integrated forward until the next observation is available and the process is repeated. In comparison with commonly used inverse methods (e.g., general least square and maximum likelihood methods), the EnKF can dynamically adjust system estimates, without reprocessing existing data when new observations become available.

2.3 Experimental design

In this study, the EnKF algorithm is implemented in a heterogeneous saturated medium with two-dimensional steady-state flow and transient-state solute transport. As shown in Fig. 1, the flow domain is a square and discretized into 20 × 20 square cells. Each cell is of one unit area. For simplicity and generality, the units of length, mass, and time are not specified in this study. The flow field is surrounded by Dirichlet boundaries on the left and right sides and Neumann boundaries on the top and bottom borders. A point source is located at point (5, 10) and a nonreactive solute is continuously injected into the medium at an injection rate of 50 M L^{-3} T^{-1}. The parameters required by flow-transport model are listed in Table 1. The statistical distribution of the log hydraulic conductivity field (ln$K$) is assumed to be statistically stationary with the exponential covariance function, $C(h)$,

$$C(\mathbf{h}) = \sigma^2 \exp\left(-\left[\frac{h_x^2}{\lambda_x^2} + \frac{h_y^2}{\lambda_y^2}\right]^{\frac{1}{2}}\right) \quad (11)$$

where h_x and h_y are the lags in the horizontal and vertical directions and $\mathbf{h} = (h_x, h_y)$; λ_x and λ_y are correlation lengths (integral scale) along horizontal and vertical directions, respectively. In our study, we choose $\lambda_x = 4$, $\lambda_y = 2$. Furthermore, we assume that the ln$K$ values are known at the locations of 16 hydraulic head observation wells marked by crosses in Fig. 1. Then a certain amount of realizations of ln$K$ fields are generated using the sequential Gaussian simulation (SGSIM) modular in GSLIB software (Deutsch and Journel 1998) and are used in our data assimilation experiment.

Two case studies are used to demonstrate EnKF's capacity of improving solute concentration prediction through assimilating hydraulic head, and concentration data. The first one (referred to as Case 1 hereinafter) is to assimilate hydraulic head measurements to identify conductivity field and update flow and transport predictions. As shown in Fig. 1a, there are 16 hydraulic head observations wells that are distributed uniformly in study domain. Additionally, the wells, A_1, A_2, A_3, and A_4, located at downstream are used to validate the study results. In this case, we assumed that the concentration source is known and is released continuously. The hydraulic conductivity and hydraulic head fields are considered as the state variables (i.e., variable X in Sect. 2.2). Procedure of the data assimilation using the EnKF is shown in the flowchart of Fig. 2a. First, ensemble of initial

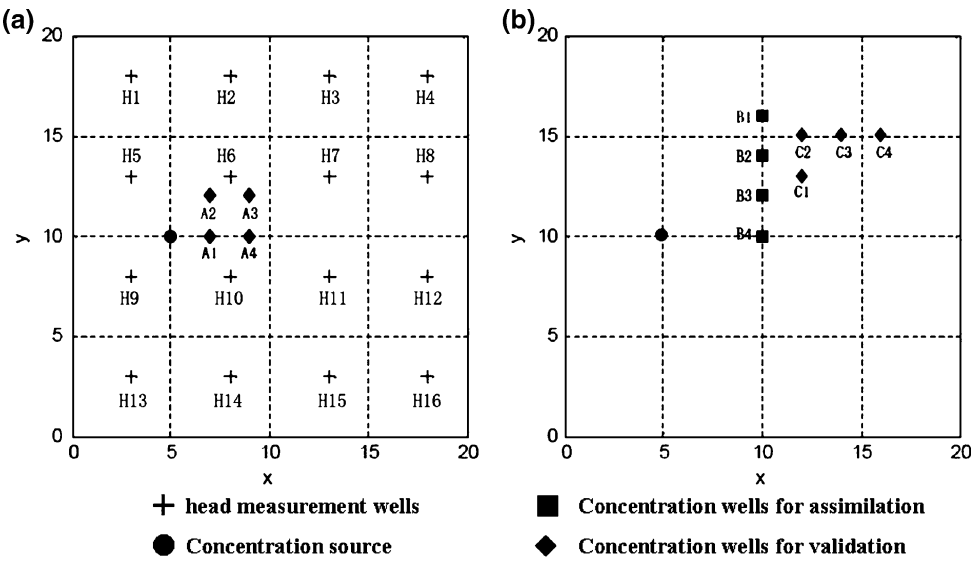

Fig. 1 The domain of the case study field and distribution of wells. **a** The case of assimilating hydraulic head data, **b** the case of assimilating concentration data

Table 1 Summary of parameter required by flow-transport model

Parameter	Value	Unit
Cell width along rows	20	L
Cell width along columns	20	L
Hydraulic conductivity	$[e^{-4}, e^{4}]$	$L\ T^{-1}$
Saturated thickness	1	L
Longitudinal dispersivity	20	L
Transverse dispersivity	4	L
Porosity	0.3	–
Concentration injection rate	50	$M\ L^{-3}\ T^{-1}$
Simulation time (t)	200	T

hydraulic conductivity fields are generated using the SGSIM, and the generated hydraulic conductivity fields are used in flow model to calculate hydraulic head in the study area. These hydraulic conductivity fields and heads are served as the first-step forecast results. When there are the hydraulic head measurements at the observation wells in current time step, which are used to update the hydraulic conductivity and hydraulic head distributions through EnKF algorithm. The updated conductivity and head values are applied to calculate groundwater flow velocity, which is used, in turn, to calculate solute concentration based on transport model. Finally, the updated hydraulic conductivity, hydraulic head, and concentration are then used to reinitialize the flow and transport models at next time step.

The second experiment (referred to as Case 2 hereinafter) is to assimilate concentration data to improve solute transport prediction with unknown contamination source. The distribution of observation wells is shown in Fig. 1b. It is assumed that the conductivity field is fixed, but the concentration source is unknown. Time-series concentration data are available at four observation wells, B_1, B_2, B_3, and B_4. The concentration is considered as the state variable and concentration data at the four wells are assimilated into the transport model to improve the concentration prediction at downstream of the observation wells. The concentration observations in wells, C_1, C_2, C_3, and C_4, are used to validate the assimilation results. Procedure of the data assimilation is shown in Fig. 2b. As shown in the figure, first, based on the given hydraulic head boundary condition and hydraulic conductivity field, we use the steady-state flow model to calculate hydraulic head and flow velocity in the study domain. Then, when the concentration data are larger than 0.1, about 3% of the peak concentration value, the data are assimilated into calculation to update the concentration field, which is then used to reinitialize the transport model at next time step. We have tested that assimilation of the large amount of low

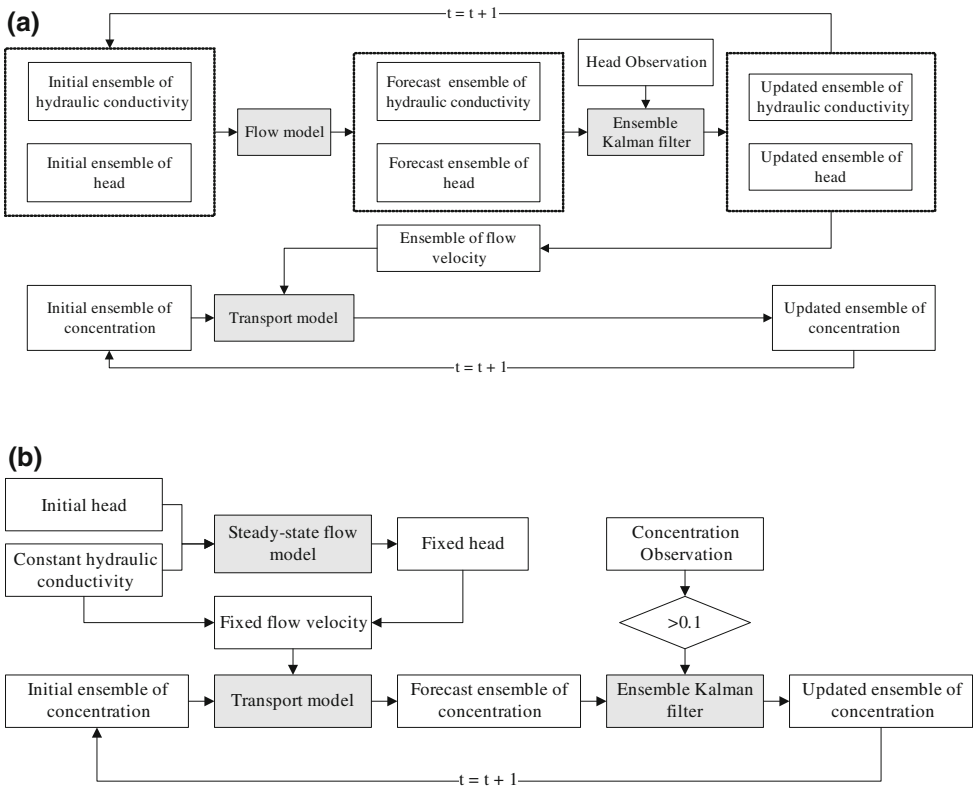

Fig. 2 The flowchart of data assimilation scheme for groundwater. **a** Flowchart of hydraulic head assimilation. **b** Flowchart of assimilating concentration measurements

concentration data, less than 0.1, into concentration calculation will hardly affect the updating results, but significantly increase computation.

For the proposed two cases, we conduct two numerical experiments to demonstrate the method of assimilating hydraulic head and concentration data, respectively. The observations and validation data used in this study are obtained from the synthetic "true" data. The availability of the "true" data allows us to determine the feasibility of the assimilation approach under known conditions. We adopt one of the multiple realizations of ln$K$ field as the reference field to run flow and transport model under given boundary and initial conditions to obtain hydraulic head distribution and temporal–spatial propagation of solute plume. At measurement locations, the hydraulic head and concentration observations are extracted then are artificially added stochastic noise with zero mean and variance 0.01 to generate a set of new observations series which were used to perform our assimilation experiments. Additionally, we assume the state model is perfect, thus the system is free of model error. As commonly adopted in groundwater modeling, we also assume that the observation errors are unbiased (the mean is aero) and independent (the observation error covariance is diagonal). The observation error standard deviation is set to be 0.1 for both the hydraulic head and concentration measurements.

3 Results and analysis

3.1 Influence of ensemble size

In the EnKF algorithm, the underlying statistics of a random field are estimated from a limited set of ensemble

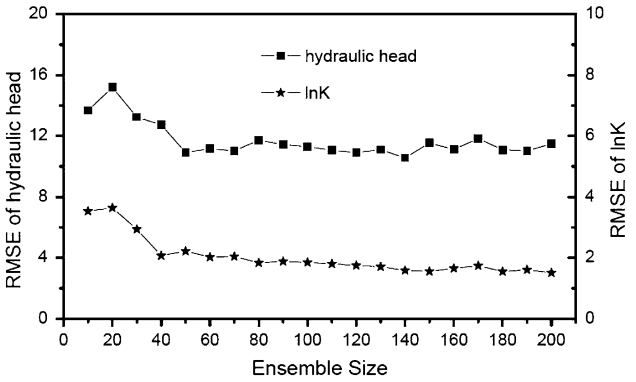

Fig. 3 The influence of ensemble size on the results of hydraulic head and conductivity fields

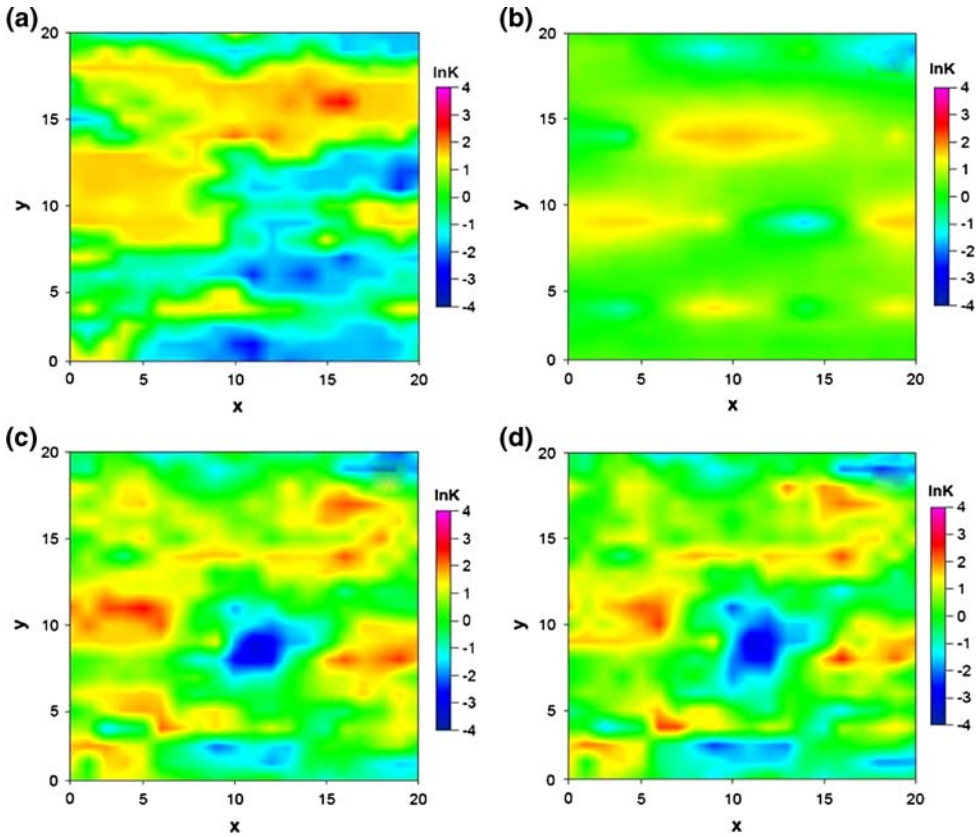

Fig. 4 Comparison of ln$K$ fields in Case 1: **a** reference field of ln$K$; **b** ensemble mean of initial ln$K$ field; **c** ensemble mean of ln$K$ field at the 5th assimilation step; **d** ensemble mean of ln$K$ field at the 50th assimilation step

members. It is expected that the estimation will improve as the ensemble size (or number of realizations) increases, so better results may be achieved by enlarging the ensemble size. On the other hand, the increase of the ensemble size will increase the computation. Therefore, it is required to find an optimal way to balance the calculation accuracy and efficiency. For the specific case we propose in this study, we also need to address the issue of ensemble size or the realization number before we do other studies. For case 1, based on 200 realizations of the ln$K$ field, we use various realization numbers, from 10 to 200, to conduct flow calculation, then compare the root mean square error (RMSE) results of ln$K$ and hydraulic head from different realization sizes to seek the influence of realization size on calculation results. The RMSE is used as a criterion to indicate the goodness of the results, which is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i^{\mathrm{t}} - \bar{S}_i^{\mathrm{f}})^2} \tag{12}$$

Here, N is cells number in experimental domain, S_i^{t} and $\bar{S}_i^{\mathrm{f}}$ stands for the true and estimated ln$K$ fields or hydraulic head fields, respectively. Since this is a synthetic case the

Table 2 The statistics of mean and variance of ln$K$ fields

	Reference field	Initial ensemble mean field	Assimilated ensemble mean field	
			5th step	50th step
Mean	−0.03402	0.44596	0.31168	0.22018
Variance	1.8546	0.37937	1.18776	1.20037

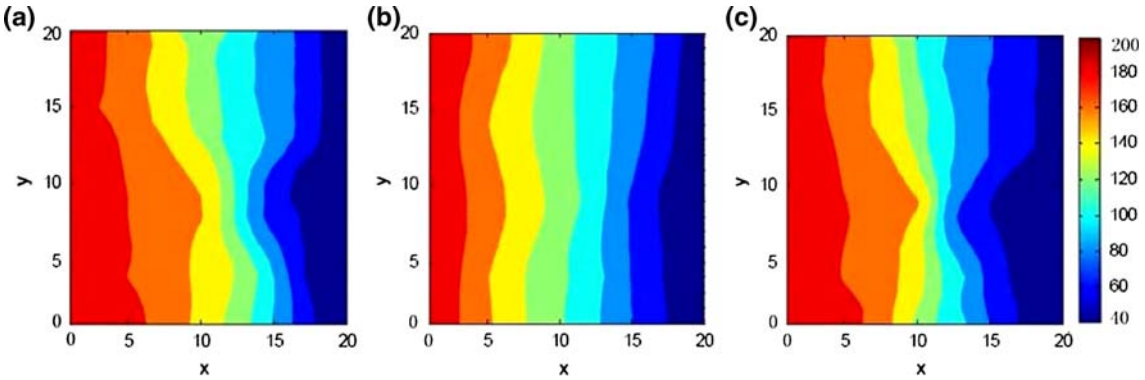

Fig. 5 Comparison of hydraulic head field for Case 1: **a** reference field; **b** ensemble mean field of simulation; **c** ensemble mean field of assimilation

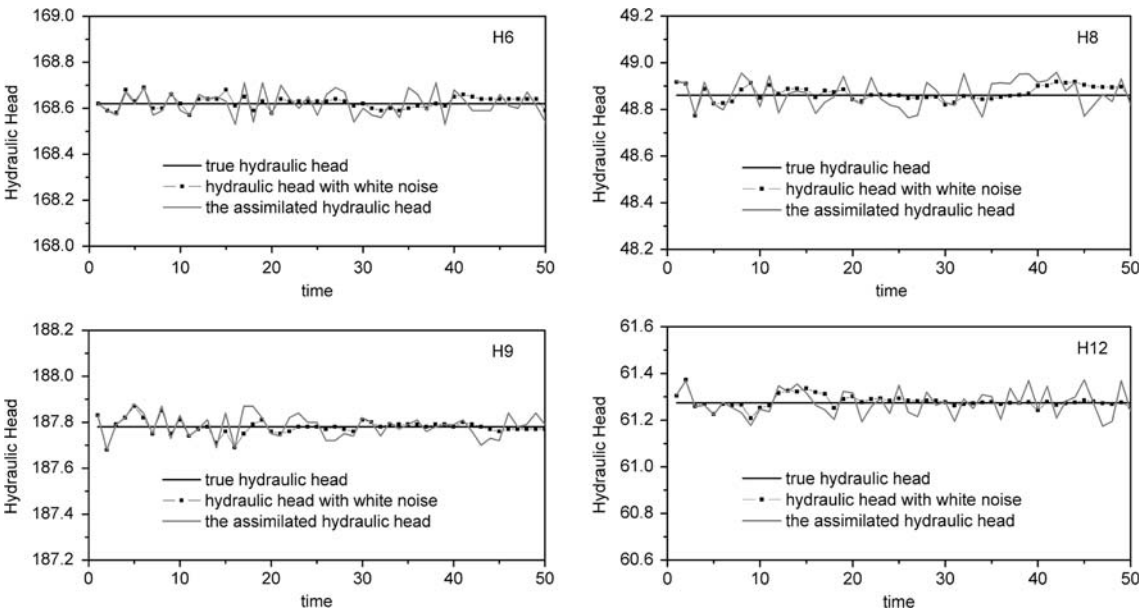

Fig. 6 Comparison of hydraulic head variation at observation wells for Case 1

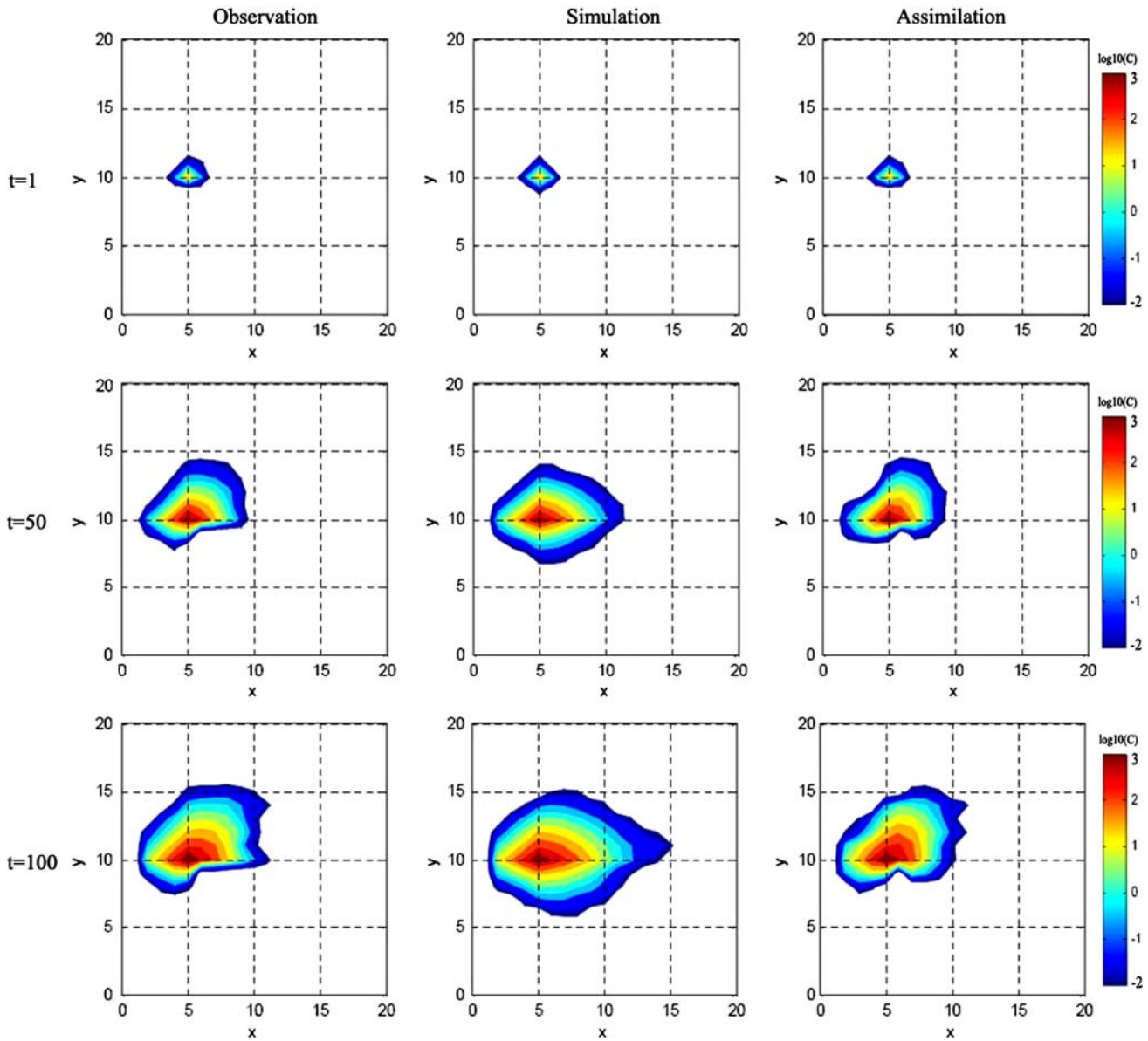

Fig. 7 The concentration field propagation by observed, simulated, and assimilated at the 1st, 50th, and 100th assimilation steps in Case 1

true field is known. The calculated RMSE results of ln$K$ and hydraulic head with different realization numbers are shown in Fig. 3. One can see from the figure that when the ensemble size is small, the RMSE values of ln$K$ and hydraulic head generally decrease with the increase of the realizations. However, when the realization number reaches 60–80, the RMSE values almost keep unchanged with the increase of the realizations. Therefore, the ensemble size is chosen to be 100 realizations.

3.2 Assimilation of hydraulic head

The reference ln$K$ field, the sample mean of the initial ln$K$ realizations and the updated ln$K$ fields at the 5th and 50th assimilation step are plotted in Fig. 4. The figures depict significant difference between the reference and the sample mean of the initial ln$K$ realizations. Although only 16 hydraulic head observations are assimilated at every time step in Case 1, the updated mean ln$K$ field is closer to the reference field after 50 assimilation steps than the initial field. The statistics of mean and variance of ln$K$ field are summarized in Table 2. From the table one can see that updated mean ln$K$ value decreases from the initial value of 0.45–0.22 while the "true" value is −0.03, and the variance increases from 0.38 to 1.20, while the "true" value is 1.85. Therefore, the improvement is significant, but there still exists a gap between the "true" values and the updated values, which indicates that there is an efficiency limit to use steady-state hydraulic head data to calibrate conductivity field. Further, in comparison of the 5th and 50th step

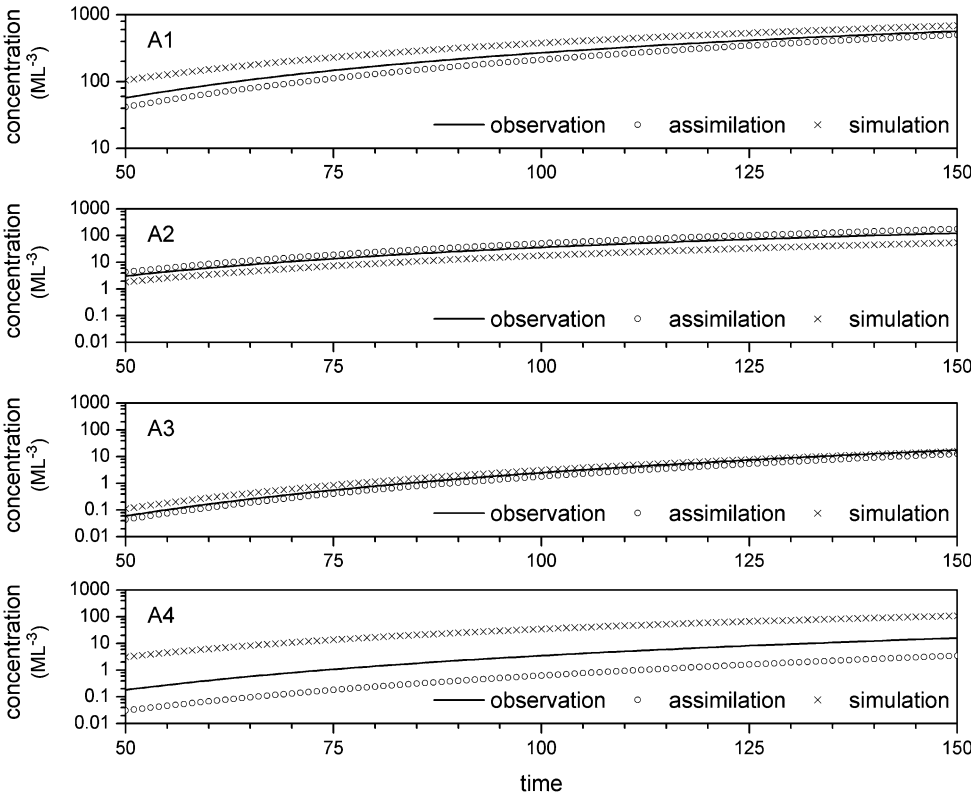

Fig. 8 The comparison of the observed, assimilated, and ensemble simulated concentration at A1, A2, A3 and A4 wells in Case 1

results, one can tell that improvement of the updating results decreases with the assimilation step increases.

The improvement of hydraulic conductivity field is also reflected by hydraulic head variation, which is shown in Fig. 5. After 50 assimilation steps, the spatial pattern of the ensemble mean head of assimilation results becomes closer to that of the reference field than the initial simulation results. The time-series of hydraulic head at four measurement wells are also shown in Fig. 6. The assimilated hydraulic head is close to the "true" measurements, although white noise is added in the hydraulic head measurements. The reason is that the hydraulic head and conductivity are adjusted at every assimilation step. It should be pointed out that the hydraulic head measurement is much more accurate than conductivity measurement in field. The measurement error can be limited to 5 cm. However, since the conductivity calibration is very sensitive to the head value, especially locally, the measurement error can lead to significant change of local conductivity field.

Based on the assimilated hydraulic conductivity and head results, we also calculate the solute transport process, which is called assimilated transport results. The variation of concentration field by observed, simulated, and assimilated is shown in Fig. 7. The simulated solute plume propagates more quickly than "true" field and the plume distribution is significantly different from the "true" one. After 50 simulation time steps, the assimilated concentration plume is much closer to the "true" one: the plume mean movement is roughly same and plume shape is also similar. We also calculate the concentration breakthrough curves at four observation wells, and the results are shown in Fig. 8. These results also indicate the data assimilation will improve the calculation results, despite that the improvement is not very significant for this case.

3.3 Assimilation of concentration data

In this case study, we use the concentration data collected at monitoring wells to update the prediction for solute plume evolution. As shown in Fig. 1b, there are four observation wells, B_1–B_4, located at the downstream of the plume. The breakthrough curves obtained from the wells are used in the data assimilation to update the solute calculation results. The concentration breakthrough curves collected from wells, C_1–C_4, are used to check the assimilation results. In this case, the hydraulic conductivity field is supposed to be known, but the plume source is unknown. The data assimilation is used to improve forecast of concentration by assimilating limited concentration observations in wells B_1–B_4 under the condition of unknown contamination source. The concentration propagation by observed and assimilated fields at the 150th, 200th, and 300th time steps are shown in Fig. 9. It is clearly shown from the figure that

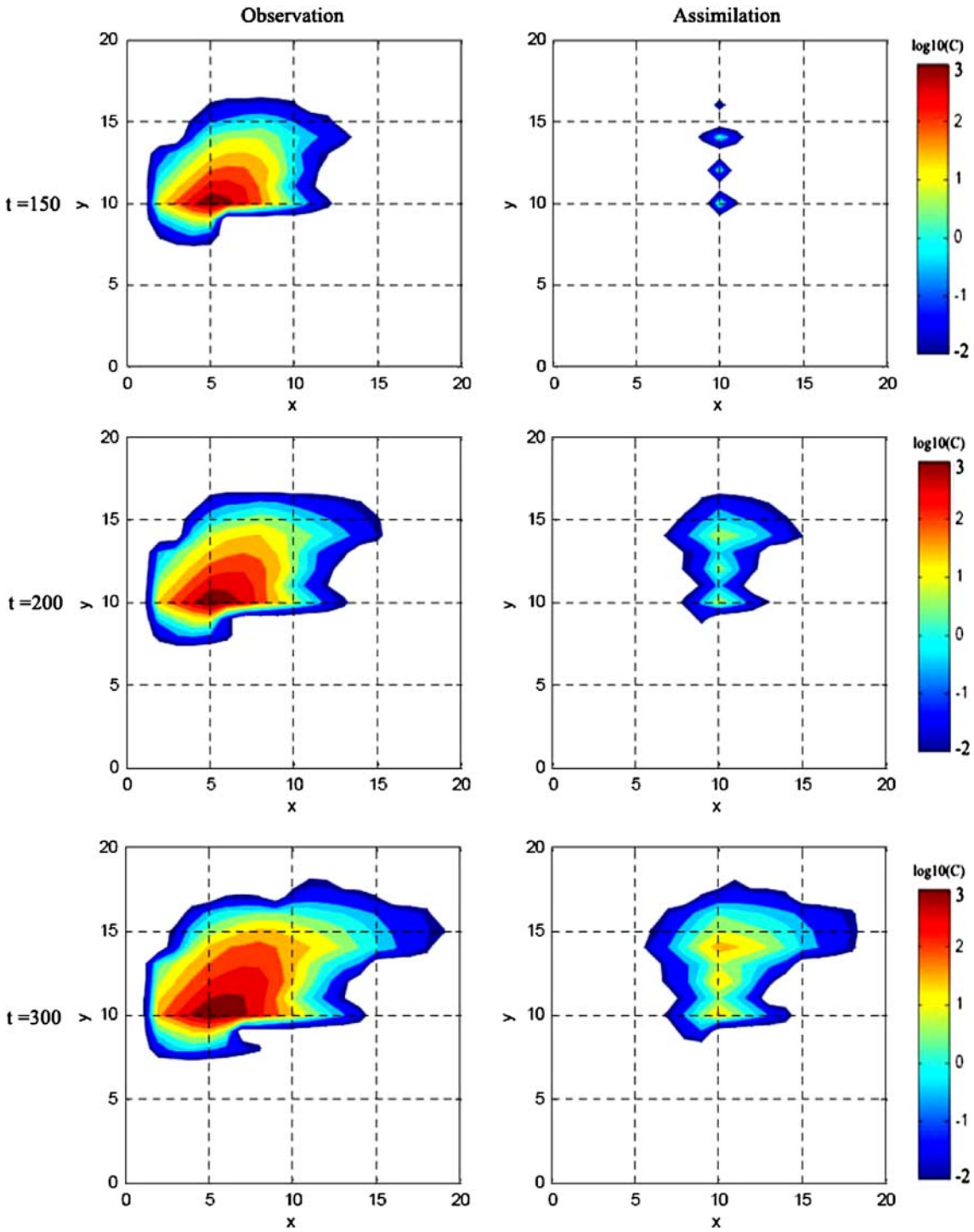

Fig. 9 Evolution of contaminant plume in the "true" and assimilated fields at the 150th, 200th, and 300th assimilation steps in Case 2

although only concentration observations in four wells are assimilated into our scheme in every time step, the improvement for plume prediction is significant. After the assimilation, the spatial pattern of concentration at the down gradient area becomes significantly closer to the "true" field. Additionally, the comparison of the observed and assimilated concentrations at C_1, C_2, C_3 and C_4 wells are plotted and shown in Fig. 10. When more and more concentration data collected at B_1, B_2, B_3 and B_4 wells are assimilated into calculation, the assimilation results at C_1, C_2, C_3 and C_4 wells become more and more similar to the "true" values. The study results indicate that when the

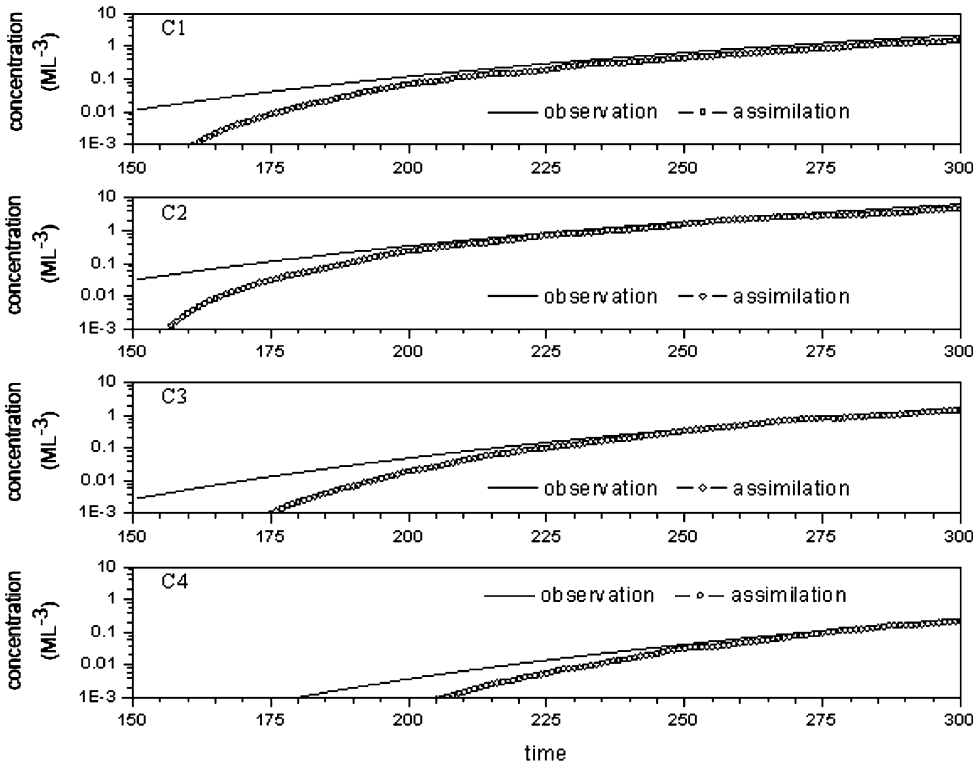

Fig. 10 The comparison of the observed and assimilated concentration at C1, C2, C3 and C4 wells in Case 2

conductivity field is known, but the solute source is unknown, the concentration data assimilation can make the concentration plume distribution and breakthrough curves in the observation wells in the downstream of the observation wells almost match the "true" ones. It should be pointed out that when the conductivity is unknown, but the solute source is known, the concentration data can also be applied to update the conductivity field. However, if both the solute source and conductivity field are unknown, we cannot use concentration data to update the two quantities. In this condition, we have to use other data, such as hydraulic head data, to do the calibration, hydraulic data for conductivity and concentration for unknown source.

4 Summary and conclusions

In this study, the EnKF is used to estimate heterogeneous hydraulic conductivity field by assimilating hydraulic head measurements, and to improve prediction of solute plume with unknown solute source condition by assimilating concentration measurements at monitoring wells. Two synthetic cases are designed to demonstrate applicability and effectiveness of the data assimilation method proposed in this study for parameter estimation and improvement of solute transport predictions. The developed method is general and applicable to transient flow and transport in a three-dimensional domain. In the first case study, we consider to use head measurements to identify a heterogeneous hydraulic conductivity field. The $\ln K$ and hydraulic head fields are supposed to be state variables and the hydraulic head measurements in limited wells are assimilated to calibrate the conductivity field. In the second case study, the data assimilation improves solute transport prediction in a known conductivity field, but unknown solute source location and solute release function. By using the EnKF method developed in this study, we assimilate the concentration data obtained at monitoring wells to examine how effective we can improve the concentration prediction at the downstream of the monitoring wells. Based on the study results, the following conclusions are obtained.

1. The EnKF can be used to effectively calibrate a heterogeneous conductivity field by assimilation of hydraulic head measurements. After the assimilation, the spatial distributions of hydraulic head, hydraulic conductivity, and concentration fields become significantly closer to reference fields than those without data assimilation.
2. The EnKF can be also used to assimilate concentration measurements obtained at limited monitoring wells to improve concentration prediction. The second case study results indicate that by locating the concentration monitoring wells in the downstream of the

contamination source, the solute transport process in the downstream of the monitoring wells can be accurately predicted by continuously assimilating the concentration data obtained in the monitoring wells even though the solute source location and solute release function are unknown. The EnKF method developed here provides an approach for many environmental projects to design concentration monitoring wells for prediction of solute transport.

3. In this study, the forecast model is assumed known and effect of measurement and model error is not studied. Therefore, identifying spatial distribution of model error and quantifying model error is still an issue needed to study.
4. The model predictability depends on ensemble size and the complexity of a medium heterogeneity characteristics and hydraulic conditions. In this study, we analyze the influence of ensemble size and conclude that 100 realizations will be accurate enough for flow calculation based on the specific conductivity field. However, the influence of the observation locations, and the measurement timings and types, and develop methods of jointly estimating flow and transport parameters need to be studied in future.

Acknowledgments This work is supported by the Chinese Academy of Science (CAS) International Partnership Project "The basic research for water issues of Inland River Basin in Arid Region" (CXTD-Z2005-2), the CAS West Development Action Plan Project (grant number: KZCX2-XB2-09) and the National Science Foundation of China (NSFC) project (grant number: 40528003, 40771036, and 40801126). The fourth author is supported in part by the FYAP program of the Florida State University.

References

Andreadis KM, Lettenmaier DP (2006) Assimilating remotely sensed snow observations into a macroscale hydrology model. Adv Water Resour 29(6):872–886

Bennett AF (1992) Inverse methods in physical oceanography. Cambridge University Press, London, 346 pp

Burgers G, van Leeuwen PJ, Evensen G (1998) Analysis scheme in the ensemble Kalman filter. Mon Weather Rev 126:1719–1724

Carrera J, Alcolea A, Medina A, Hidalgo J, Slooten LJ (2005) Inverse problem in hydrogeology. Hydrogeol J 13:206–222

Chen Y, Zhang D (2006) Data assimilation for transient flow in geologic formations via ensemble Kalman filter. Adv Water Resour 29:1107–1122

Christakos G (2002) On the assimilation of uncertain physical knowledge bases: Bayesian and non-Bayesian techniques. Adv Water Resour 25:1257–1274

Christakos G (2005) Methodological developments in geophysical assimilation modeling. Rev Geophys 43(2):RG2001. doi:10.1029/2004RG000163

Clark MP, Slater AG, Barrett AP (2006) Assimilation of snow covered area information into hydrologic and land-surface models. Adv Water Resour 29(8):1209–1221

Daley R (1991) Atmospheric data analysis. Cambridge University Press, New York, p 457

Deutsch CV, Journel AG (1998) GSLIB: geostatistical software library and user's guide, 2nd edn. Oxford University Press, New York, 369 pp

Drecourt JP, Madsen H, Rosbjerg D (2006) Calibration framework for a Kalman filter applied to a groundwater model. Adv Water Resour 29(5):719–734

Evensen G (2003) The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn 253:343–367

Evensen G (2006) Data assimilation: the ensemble Kalman filter. Springer, New York

Gu YQ, Oliver DS (2005) History matching of the PUNQ-S3 reservoir model using the ensemble Kalman filter. SPE J 10(2):217–224

Gu YQ, Oliver DS (2006) The ensemble Kalman filter for continuous updating of reservoir simulation models. J Energy Resour Technol Trans ASME 128(1):79–87

Hoeksema RJ, Kitanidis PK (1984) An application of the statistical approach to the inverse problem in two-dimensional groundwater modeling. Water Resour Res 20(7):1003–1020

Houser PR, Shuttleworth WJ, Gupta HV (1998) Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. Water Resour Res 34(12):3405–3420

Huang CL, Li X, Lu L, Gu J (2008a) Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter. Remote Sens Environ 112(3):888–900

Huang CL, Li X, Lu L (2008b) Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter. Remote Sens Environ 112(4):1320–1336

Jazwinski AH (1970) Stochastic processes and filtering theory. Elsevier, New York

Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82(D):35–45

Li X, Koike T, Pathmathevan M (2004) A very fast simulated reannealing (VFSA) approach for land data assimilation. Comput Geosci 30:239–248

Liu Y, Gupta HV (2007) Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. Water Resour Res 43:W07401. doi:10.1029/2006WR005756

Margulis SA, McLaughlin D, Entekhabi D, Dunne S (2002) Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. Water Resour Res 38(12):1299. doi:10.1029/2001WR001114

McLaughlin D (1995) Recent development in hydrologic data assimilation. Rev Geophys 33(suppl):977–984

McLaughlin D, Townley LR (1996) A reassessment of the groundwater inverse problem. Water Resour Res 32(5):1131–1161

Moradkhani HS, Sorooshian H, Gupta V, Houser PR (2005) Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Adv Water Resour 28:135–147

Neupauer RM, Lin R (2006) Identifying sources of a conservative groundwater contaminant using backward probabilities conditioned on measured concentrations. Water Resour Res 42:W03424. doi:10.1029/2005WR004115

Neupauer RM, Wilson JL (2001) Adjoint-derived location and travel time probabilities for a multidimensional groundwater system. Water Resour Res 37(6):1657–1668

Poeter EP, Hill MC (1997) Inverse models: a necessary next step in ground-water modeling. Ground Water 35(2):250–260

Reichle RH, McLaughlin DB, Entekhabi D (2002a) Hydrologic data assimilation with the ensemble Kalman filter. Mon Weather Rev 130:103–114

Reichle RH, Walker JP, Koster RD (2002b) Extended versus ensemble filtering for land data assimilation. J Hydrometeorol 3:728–740

Sorensen JVT, Madsen H, Madsen H (2004) Data assimilation in hydrodynamic modelling: on the treatment of non-linearity and bias. Stoch Environ Res Risk Assess 18(7):228–244

Sun NZ (1994) Inverse problems in groundwater modeling. Kluwer, Dordrecht

Van Geer FC, Te Stroet CBM, Zhou X (1991) Using Kalman filtering to improve and quantifying the uncertainty of numerical groundwater simulations: 1. The role of system noise and its calibration. Water Resour Res 27(8):1987–1994

Vrugt JA, Diks CGH, Gupta HV, Bouten W, Verstraten JM (2005a) 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 JA, Robinson BA, Vesselinov VV (2005b) Improved inverse modeling for flow and transport in subsurface media: combined parameter and state estimation. Geophys Res Lett 32:L18408. doi:10.1029/2005GL023940

Wen X-H, Chen W-H (2005) Real-time reservoir model updating using ensemble Kalman filter. SPE J 11(4):431–442

Yangxiao Z, Te Stroet CBM, van Geer FC (1991) Using Kalman filtering to improve and quantifying the uncertainty of numerical groundwater simulations: 2. Application to monitoring network design. Water Resour Res 27(8):1995–2006

Yeh T-CJ, Liu S (2000) Hydraulic tomography: development of a new aquifer test method. Water Resour Res 36(8):2095–2105

Yeh T-CJ, Zhang J (1996) A geostatistical inverse method for variably saturated flow in the vadose zone. Water Resour Res 32(9):2757–2766

Zhang D, Lu Z, Chen Y (2007) Dynamic reservoir data assimilation with an efficient, dimension-reduced Kalman filter. SPE J 12(1):108–117

Zheng C, Bennett GD (2002) Applied contaminant transport modeling, 2nd edn. Wiley, New York

Zheng C, Wang P (1999) MT3DMS: A modular three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems; documentation and user's guide, Contract Report SERDP-99-1. U.S. Army Engineer Research and Development Center, Vicksburg, MS

Zhu J, Yeh T-CJ (2005) Characterization of aquifer heterogeneity using transient hydraulic tomography. Water Resour Res 41:W07028. doi:10.1029/2004WR003790

Zimmerman DA et al (1998) A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Water Resour Res 34(6):1373–1413