

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Data Assimilation Application to the Subsurface Flow and Solute Transport

Bill X. Hu and Juxiu Tong

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/56705>

1. Introduction

Natural media are generally heterogeneous. The medium properties (hydraulic parameters such as hydraulic conductivity and porosity) are often treated as random space functions. Without an adequate description of the spatial distributions of the hydraulic parameters, such as hydraulic conductivity and porosity, the ability of the numerical model to predict groundwater flow and solute transport is limited and tends to deteriorate over time. However, characterization of field heterogeneity is always difficult due to economical and technical limitations, which make the predictions, deteriorate over time.

Many inverse algorithms have been developed in the field of hydrogeology (Constantinescu et al, 2007; Hoeksema and Kitanidis, 1984; Poeter and Hill, 1997; Sun, 1994; Sun and Yeh, 1992; Vrugt et al., 2005a; Vermeulen et al., 2006; Yeh and Liu, 2000; Yeh and Zhang 1996, Zhu and Yeh, 2005). With newly developed techniques, more observations become available. But the observation data from these techniques are usually indirectly related to the model parameters, which are of the most interest in practice. Moreover, there may be different kinds of measurements. Therefore, it is necessary to develop approaches to dynamically reconcile the observation information of different kinds. The data assimilation method is established in this situation (Thomsen and Zlatev, 2008).

Data assimilation is different from the traditional parameter estimation methods. Originating from meteorology and oceanography (Daley, 1991; LeDimet and Talagrand, 1986), the data assimilation method has been developed for improvement of operational weather forecasts and ocean dynamics prediction (Evensen, 2003, 2004; Fang et al., 2006; Houtekamer and Mitchell, 1998, 2001; Thomsen and Zlatev, 2008). Data assimilation methods have been applied to many study areas (Andreadis and Lettenmaier, 2006; Aubert et al., 2003; Clark et al., 2006; McLaughlin and Townley, 1996; Mclaughlin, 2002; Natvik and Evensen, 2003; Reichle, 2008).

In geophysics, data assimilation methods have been used to assimilate geophysical data to characterize medium heterogeneity (Christakos, 2005). Data assimilation methods have been used in hydrology for uncertainty and sensitivity analysis (Liu and Gupta, 2007; Van Geer et al., 1991; Vrugt et al., 2005b) and in the petroleum industry (Oliver et al., 2008).

The traditional Kalman filter (KF) is an efficient sequential data assimilation method for linear dynamics and measurement processes with Gaussian error statistics (Kalman, 1960; Kalman and Bucy, 1961; Drecourt, 2003; Drecourt et al., 2006; Gelb, 1974; Maybeck, 1979; Tipireddy et al., 2008; Zou et al., 1991; Zhang et al., 2007). To assimilate data for nonlinear dynamics and measurement processes, the extended Kalman filter (EKF) was developed (Jazwinski, 1970). However, the EKF is very unstable if the nonlinearities are strong. Furthermore, this method is not computationally feasible for large-scale environmental systems.

To overcome these limitations, the ensemble Kalman filter (EnKF) is proposed (Evensen, 1994). The Ensemble Kalman filter (EnKF) method is a Monte-Carlo variant of the KF, consisting of a set (or ensemble) of parallel short-term forecasts and data assimilation cycles. Statistics derived from the short-term forecasts are used to estimate background error covariances during the subsequent data assimilation step. Further, ensemble-based techniques have two potential advantages over the traditional EKF (Cohn, 1997): (1) the computational cost of ensemble techniques are significantly less than the EKF because covariances are estimated using a limited-size random sample, while the error covariances of EKF for each model component are propagated using linear tangent and adjoints of the fully nonlinear model, and exorbitant computational expense for a high-dimensional model [however, it may be possible to reduce the computations by computing in a reduced-dimensional subspace, e.g., Fisher (1998)]; (2) in terms of accuracy, ensemble filters may be more accurate than the EKF since covariances are estimated by propagating model states with a fully nonlinear model rather than under assumptions of linearity. Moreover, it is also able to account for the possible model noise/error.

Recent development of the EnKF is not limited to updating system state variables (such as hydraulic head in transient groundwater flow modeling and solute concentration in solute transport modeling) as in conventional data assimilation, but also allow updating state variables and model parameters (such as hydraulic conductivity in transient groundwater flow modeling or in transient solute transport on the basis of the steady state groundwater flow) to yield more accurate model predictions. The application of the EnKF in hydraulic modeling (including groundwater flow and solute transport) has also received more and more attention. Chen and Zhang (2006) have successfully applied data assimilation for transient groundwater flow in geologic formations via EnKF, but in their study, they did not consider the influence of different boundaries and the locations of observation wells on the data assimilation results. Huang et al. (2009) used the EnKF data assimilation method to calibrate a heterogeneous conductivity field using steady state groundwater flow data and calibrate solute plume initial condition using tracer test data.

All of the above works are done by applying the EnKF to calibrate the hydraulic conductivity by assimilating the hydraulic head. For the solute transport modeling, though Liu et al. (2008) applied the EnKF to calibrate the conductivity fields by assimilating tritium concentra-

tion at the MADE site, they first assimilated hydraulic head observations to update the hydraulic conductivity and obtained the flow field for the solute transport process. They also incorporated the updated variable longitudinal dispersivity during the solute transport modeling, which will influence the solute prediction and the corresponding inversed results including the updated hydraulic conductivity field.

All uncertainty sources, including parameters and output observations, need to be considered for improvement of forecast capability (Wang et al., 2009). In most physical systems, states and/or parameters are bounded, which introduces constraints on their estimates. While in sequential data assimilation techniques like the EnKF, the state variables are updated by minimizing the mean square error rather than following physical principles, which may result in the violation of physical constraints and many methods have been proposed to solve these problems (Prakash et al., 2008; Wang et al., 2009). In the conservative solute transport modeling, solute concentrations have their own physical meaning. So it is better to pose some constraints to them in case they violate the physical meaning after the application of the EnKF method (Hemant and Olive, 2009, Sridhar et al., 2007; Thacker, 2007). We call this method as constrained EnKF method.

According to the above introduction about the EnKF, the EnKF application in groundwater is classified into two types: using groundwater flow data or conservative solute transport data to calibrate the hydraulic parameters, e.g. hydraulic conductivity and corresponding state variables, e.g., hydraulic head and conservative solute concentrations. The rest is arranged as follows:

In section 2, the EnKF data assimilation method is introduced based on the numerical forward groundwater flow and conservative solute transport models.

In section 3, the research design for the groundwater flow and conservative solute transport is described. The synthetic or illustrative examples or cases for the groundwater flow and conservative solute transport will be depicted in detail. In our research, we will apply EnKF method to update a hydraulic conductivity field through assimilating transient head data in the groundwater flow modeling part. The study focuses on investigating the capability of EnKF to update a hydraulic conductivity field through assimilating transient head data, and explores the influences of different factors on assimilation results. In the conservative solute part, we will assimilate solute transport data to calibrate conductivity fields via the constrained EnKF method, where the conservative solute concentration is constrained to be non-negative. The constrained EnKF method is used to identify a two-dimensional (2D) heterogeneous hydraulic conductivity field with mixed Neumann/ Dirichlet boundary conditions through assimilating conservative solute concentration observations. The study focuses on investigating the capability of constrained EnKF to update a hydraulic conductivity field, and explores the influences of different factors on assimilation results.

In section 4, the updated results through assimilating groundwater flow and conservative solute transport data will be provided, and then discussion and analysis of the influences of various factors on the results will be provided in detail. Conclusions for the inverse modeling of the groundwater flow and conservative solute transport will be made.

2. Data assimilation method

Since data assimilation is a process in which estimates of model parameters and variables are updated by requiring consistency with observations and governing equations, a data assimilation system is composed of a model operator, an observation operator and a data assimilation algorithm.

2.1. Transient stochastic flow model

The governing equation for transient groundwater flow in unsaturated-saturated medium is (Bear, 1972; Arlen, 2005):

$$\nabla \cdot \mathbf{q}(x, t) + g(x, t) = S_s \frac{\partial h(x, t)}{\partial t} \quad (1)$$

$$\mathbf{q}(x, t) = -K(x) \nabla h(x, t) \quad (2)$$

where \mathbf{q} [LT^{-1}] is the Darcy's flux; $h(x, t)$ [L] is the hydraulic head; \mathbf{x} [L] is the spatial parameters, *i.e.*, in the three-dimensional domain, \mathbf{x} means x, y, z ; $g(x, t)$ [T^{-1}] is the external sink or source term; $K(x)$ [LT^{-1}] is the hydraulic conductivity; ∇ stands for $(\partial/\partial x, \partial/\partial y, \partial/\partial z)$; S_s is the specific storage. In this study, the hydraulic conductivity $K(x)$ is considered as a log-normally distributed random space function while specific storage S_s is treated as a deterministic constant. Since $K(x)$ is a random function, the flow equations become stochastic nonlinear partial differential equations.

2.2. Transient conservative solute transport model

When the right side in Equation (1) is 0, then it becomes a governing equation for the steady state flow. On the basis of the mass balance and Fick's law, the governing equation for the transport of conservative solute in groundwater flow can be written as (Zheng, 1990):

$$\frac{\partial \theta c(x, t)}{\partial t} = \nabla \cdot (\theta D \cdot \nabla c(x, t)) - \nabla \cdot (\mathbf{q}(x, t) \cdot c(x, t)) - g(x, t) c_s \quad (3)$$

Where θ [L^3L^{-3}] is the porosity of the medium; $c(x, t)$ [ML^{-3}] is the solute concentration; D [L^2T^{-1}] is the hydrodynamic dispersion tensor; c_s [ML^{-3}] is the solute concentration of water entering from sources or sinks.

2.3. Ensemble Kalman filter

EnKF is a Monte Carlo method. The EnKF method is briefly introduced with perturbed observation proposed by Burgers et al. (1998). In this algorithm, predictions of model variables

S , including both the variables (hydraulic head in the groundwater flow modeling or solute concentration in the solute transport modeling) and parameters (hydraulic conductivity), 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 S are defined as

$$P^f \cong P_e^f = \overline{(S^f - \bar{S}^f)(S^f - \bar{S}^f)^T} = S^{f'}(S^{f'})^T / N - 1 \quad (4)$$

$$P^a \cong P_e^a = \overline{(S^a - \bar{S}^a)(S^a - \bar{S}^a)^T} = S^{a'}(S^{a'})^T / N - 1 \quad (5)$$

In the forecast step, forecasted model variables of each ensemble member are updated according to

$$S_{i,t+1}^f = M(S_{i,t}^a) + e_{1i} \quad e_{1i} \sim N(0, Q_s) \quad (6)$$

Where $S_{i,t+1}^f$ is the forecasted model variable of the i th ensemble member at time $t+1$; $S_{i,t}^a$ the analyzed model variable of the i th ensemble member at time t ; $M(\cdot)$ is model operator, which is the flow model in groundwater modeling and transport model in solute modeling; e_{1i} is the model error vector, which is assumed to satisfy a Gaussian distribution with zero mean and covariance matrix Q_s .

In the analysis step, the observation data are perturbed by adding random observation errors. The observation vector at the time step $t+1$ for each ensemble member is given by:

$$d_{i,t+1} = HS_{i,t+1}^{\text{true}} + e_{2i} \quad (7)$$

Where e_{2i} is random error vector of observation with zero mean and covariance matrix R_e .

The forecast of each ensemble member is updated as follows (Burgers et al., 1998):

$$S_{i,t+1}^a = S_{i,t+1}^f + P_e^f H^T (HP_e^f H^T + R_e)^{-1} (d_{i,t+1} - S_{i,t+1}^f) \quad (8)$$

Where H is the observation operator used to convert the model state variables to observations; $\bar{S}_{i,t+1}^f$ is the mean of forecast state vector of ensemble members at time $t+1$.

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 (hydraulic head or solute concentration) is available and the process is repeated. In comparison with commonly used inverse methods (e.g., general least squares and maximum likelihood methods), the EnKF can dynamically adjust the system estimate, without reprocessing existing data when new observations become available. The data assimilation procedure using the EnKF in this paper will be introduced later combined with the description of the study cases below.

3. Illustrative examples for the groundwater flow and conservative solute

3.1. Transient state groundwater flow

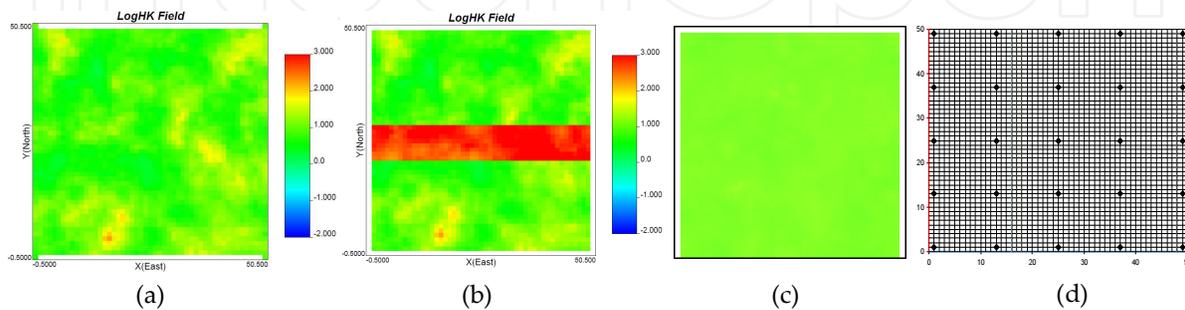
In this section a two-dimensional, transient flow in a heterogeneous saturated medium is used to demonstrate the capability of EnKF to estimate the hydraulic conductivity field by assimilating hydraulic head measurements and to explore the sensitivity of EnKF to various factors. The flow domain is a square with $L_x = L_y = 50.0 [L]$ (L is any consistent length unit) and uniformly discretized into 50×50 square grid cells. To investigate the effect of boundary condition on parameter calibration, two different boundary conditions for the study domain are proposed in this study. In the first case (referred to as Case 1 hereinafter), all four sides of the flow domain are prescribed as no-flow boundaries. There is a pumping well with continuous and constant pumping rate of $10 [L^3/T]$ (T is any consistent time unit) in the middle point of the flow domain (25, 25). In the second case (referred to as Case 2 hereinafter), the no-flow condition is still assumed for the top and bottom boundaries, but constant head boundary conditions are used on the left and right sides. Similarly, there is a pumping well in the middle point of the flow field (25, 25), but the initial pumping rate is $500 [L^3/T]$ and the pumping rate increases $5 [L^3/T]$ every $5 [T]$ to speed up the head decrease in the domain.

The conductivity field generator scheme developed by Hu et al. (2009) was used in this study to generate the reference (or real) conductivity field and the initial guess conductivity field in this study:

$$\log K = -0.457 - 0.0300M + 7.76n + 2.03 \log d_{10} \quad (9)$$

Where M is the matrix volume fraction of the grain size distribution in the medium; n is the porosity of the medium; d_{10} is the 10th percentile of grain size distribution. The distributions of M , n and d_{10} were obtained from the field experiment. In this study, a special case is considered in which d_{10} is a spatial random variable, its distribution is represented by a fractal function according to the field measurement, while the other two parameters, M and n are assumed to be constant and their mean values are used in Eq. (9). Using Hu et al.'s (2009) method, multiple, equally-probable realizations of the hydraulic conductivity field is generated based on Eq. (9). One of the generated realizations is shown in Figure 1(a), where it is apparent that the field is in the log form. One can see that the realized field is heterogeneous,

but statistically stationary. An ensemble of a sufficient number of realizations should reach a constant field. To generate a “real” or reference conductivity field, one realization is randomly picked and modified the field by inserting a high conductivity layer in the middle of the domain by setting the conductivity value at every point in the layer 100 times the original value. The constructed figure with log form is shown in Figure 1(b) and the red band part is the modified part. This field will be used to generate hydraulic head data at selected measurement points, while the other realizations, without modification, will be used as the initial “guess” conductivity field by averaging them. For example, In Figure 1(c), the 500 realizations mean logK field is taken as the representative initial “guess” conductivity field.



(Note: all the Figures of log hydraulic conductivity fields for the transient groundwater flow here and next have the same color scale)

Figure 1. Distribution of log conductivity field: (a) One realization of the initial guess field; (b) Real conductivity field; (c) Ensemble mean of 500 realizations; (d) Uniform distribution of 25 observation wells in the study domain

Based on the hydraulic boundary conditions and the conductivity fields described above, the flow equations are solved with the Galerkin finite-element method (FEM).

The simulated results based on the reference conductivity field at the chosen “measurement” locations will be used as the real values of the hydraulic heads at the locations. Considering possible measurement error in reality, the real head values are perturbed by white noises and the perturbed head results serve as the observed head values. In this study, the noise has a mean equal to 0 (indicating unbiased observations) and the standard deviation is 1% of the hydraulic head measurements. It should be pointed out that the EnKF method is not limited to this simplification of measurement error. In fact, the errors, assumed or inferred from the measurements, may vary with measurement types, times, and locations. In this study, the simulation time is evenly subdivided into many time steps of size 1 [T].

In case 1, in the vertical (z) direction, the bottom level of the study domain is -100 [L], and the upper level is 2.5 [L]. The initial hydraulic head throughout the domain is assumed to be 3.0 [L], so the saturated groundwater is confined at first, and the specific storage for this confined study medium is assumed to be 0.0001. After some time of the pumping the groundwater table goes below the upper level of the medium in some areas, as the groundwater becomes unconfined. The specific storage for the unconfined study medium is set to 0.1.

In case 2, bottom level of the study domain is -100 [L], and the upper level is 2.0 [L]. The constant hydraulic head on the left border is 3.0 [L], and the constant hydraulic head on the right border is 2 [L]. All the initial hydraulic head except the constant head boundaries are 0 [L]. The specific

storage and specific yield for the confined or unconfined study medium are the same as in case 1.

In the case study, the observation wells are distributed evenly in the study domain (e.g. 25 wells shown in Figure 1(d)), and continuous data can be obtained from the observation wells.

The procedure of the data assimilation using the EnKF is shown in a flowchart in Figure 2. First, ensembles of initial hydraulic conductivity fields are generated, and the generated hydraulic conductivity fields are used in transient flow model to calculate hydraulic head in the study area. These hydraulic conductivity fields and heads serve as the first-step forecast results. Second, if there are the hydraulic head measurements at the observation wells at the current time step, the measurements will be used to update the hydraulic conductivity and hydraulic head distributions through the EnKF algorithm. Finally, the updated hydraulic conductivity field and hydraulic heads are used to reinitialize the transient flow model at the next time step.

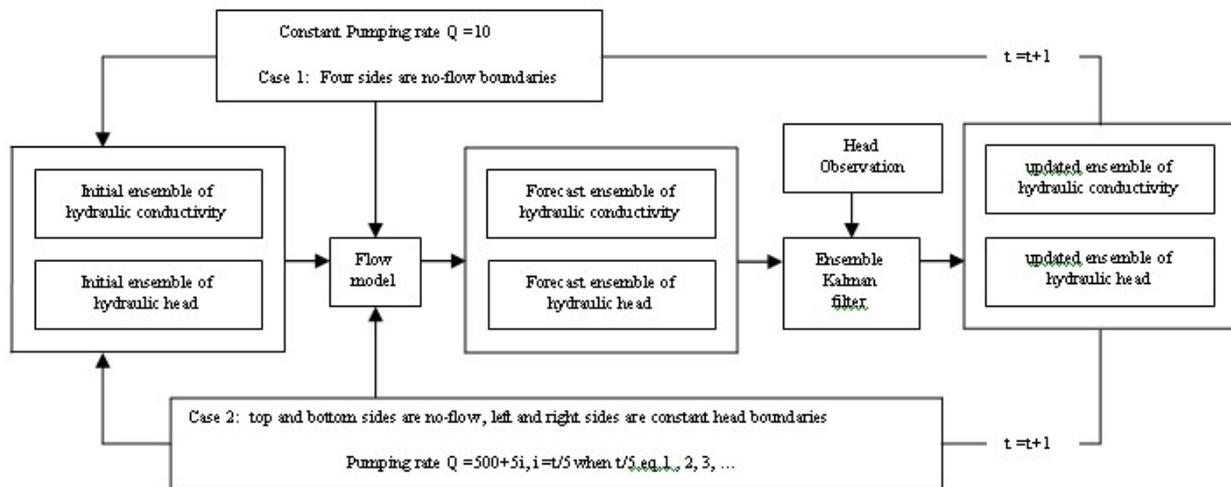


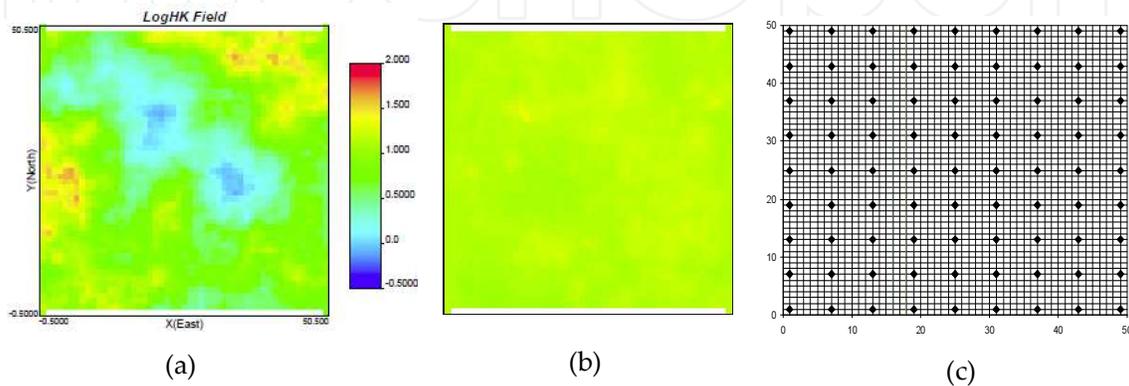
Figure 2. Flowchart of hydraulic head assimilation scheme for transient groundwater flow

3.2. Transient state conservative solute transport

The conservative solute transport modeling is based on the steady state groundwater flow. In our study, the same study case is chosen as Case 2 ($L_x = L_y = 50.0$ [L], 50×50 square grid cells, the top and bottom boundaries are the second Neumann condition with no-flow, but constant head boundary conditions on the left and right sides.) in part 3.1 except that there is no pumping well at all. In this way, the steady state flow can be obtained. With the same idea as in part 3.1, the ensemble realizations of the initial guessed hydraulic conductivity fields are produced by Hu et al.'s (2009) method. One of the generated realizations of log form is directly picked out and shown in Figure 3(a), which is also used as the true or reference hydraulic conductivity field in the solute transport modeling. The ensemble number is selected as 300, and these realizations mean logK field is taken as the representative initial "guess" conductivity field, which is shown in Figure 3(b). Uniform distributions of 25 and 81 observation wells

are studied here, and the former distribution have been presented in Figure 1(c), while the later distribution is presented in Figure 3(c). Continuous solute concentration data can be obtained from the observation wells where the solute reaches.

Based on the hydraulic boundary conditions and the conductivity fields described above, the forward groundwater flow equation is solved with the Galerkin finite-element method (FEM). Then the Darcy's velocity can be calculated and the transport equation is solved with the Galerkin FEM based on the initial instant injection of solute concentration somewhere.



(Note: all the Figures of log hydraulic conductivity fields for the conservative solute transport here and next have the same color scale)

Figure 3. Distribution of log conductivity field: (a) One realization or real conductivity field; (b) Ensemble mean of 300 realizations of the initial guess field; (c) Uniform distribution of 81 observation wells in the study domain

The observed concentration values are obtained by adding white noises to the real solute concentration values, which is the simulated results based on the reference conductivity field. The noise is unbiased and the standard deviation is 1% of the solute concentration measurements. The simulation time for the solute transport modeling is evenly subdivided into many time steps of size 0.1 [T] after the steady water velocity field is obtained by solving the governing flow equation.

To purely verify the ability of the EnKF method to update the hydraulic conductivity field through assimilating the solute concentration, all the other related parameters are assumed to be known except the hydraulic conductivity in the study domain. The solute longitudinal and transverse dispersivity is 0.1 [L] and 0.01 [L], respectively. The corresponding random porosity field is also one realization produced in Hu et al. (2009). In order to explore the EnKF ability to update the hydraulic conductivity field through assimilating the solute concentration, the random porosity field is the same for every realization.

Models usually have relevant physical laws or settings, but the EnKF formulation appears to be a promising approach for dealing with a wide class of nonlinear state estimation problems, it cannot handle bounds on the state variables (the solute concentrations) and/or the parameters (hydraulic conductivity here) that are being estimated (Hemant and Olive, 2009; Prakash et al., 2008; Thacker, 2007; Wang et al., 2009), which will result in the violation of physical

meaning. So it is necessary to account for bounds on state and parameters being estimated. The procedure of the data assimilation method using the constrained EnKF is shown in a flowchart as Figure 4. First, ensembles of initial parameters fields are generated, which are used in steady-state flow model to calculate solute concentration. If there are the solute concentration measurements at the observation wells at the current time step, the measurements will be used to update the hydraulic conductivity field and solute concentration distributions through the EnKF algorithm. If it is desired to apply the ensemble Kalman filter for state estimation when states are bounded, then it become necessary to modify the EnKF. Therefore, constraints are treated in a post processing after applying the EnKF (Wang et al., 2009), and bounds have to be deal with in EnKF framework (Prakash et al., 2008). For example, the solute concentrations should not be less than 0, if the updated concentrations become less than 0, then we set them as 1% of the concentration value for the last time step. In this way, the negative value of the solute concentration can be avoided. Finally, the updated reactive related parameters and solute concentration are used to reinitialize the conservative solute transport model at the next time step.

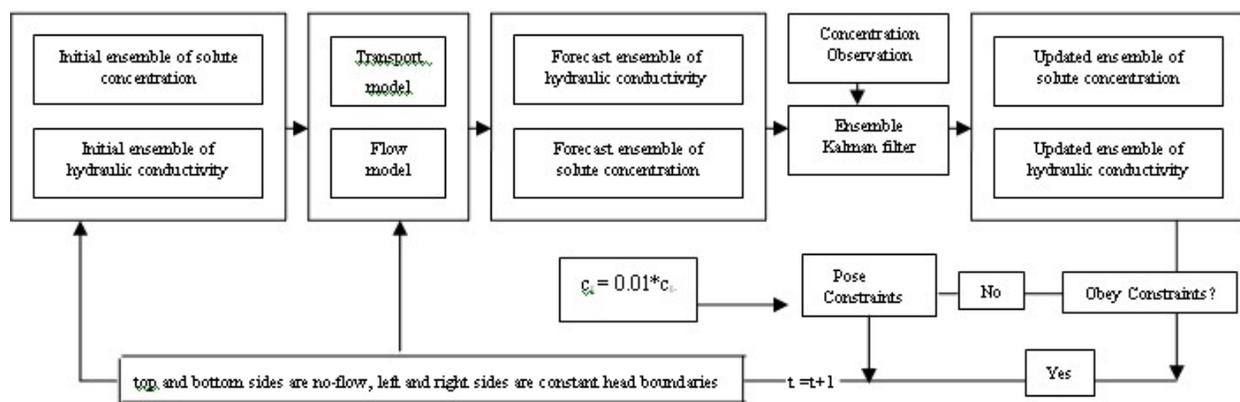


Figure 4. Flowchart of the constrained EnKF for solute concentration assimilation scheme with steady state groundwater flow

4. Application of data assimilation method to the groundwater flow and conservative solute transport

To determine the accuracy of the updated state vectors, the reference (also called true) simulation is compared to the EnKF simulation using the root mean squared error (RMSE) as the following:

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M \left(\bar{S}_i^f - S_i^t \right)^2} \tag{10}$$

where $RMSE$ is the criteria for the degree of the data assimilation capability, M is the total number of numerical grid elements in the study domain, which is 2601 in this hypothetical study case of section 3 or 4; \bar{S}_i^f and S_i^f stands for the estimated and reference hydraulic conductivities at grid i , respectively.

4.1. Transient state groundwater flow

In this part, the observations are assumed to be available for us for every simulation time, so the observed data can be assimilated for every current simulated time via the EnKF method.

4.1.1. Influence of ensemble size

To determine the ensemble size that should be used in our study, it is needed to study the influence of ensemble size on the data assimilation results. Here, it is assumed that there are 25 (5×5) observation wells, and they are distributed uniformly within four no-flow boundaries (Case 1).

The mean and variance values of the hydraulic conductivity field with and without data assimilation for different ensemble size is shown in Figures 5 and 6. The simulated hydraulic conductivity is compared with the real hydraulic conductivity fields for different ensemble size. The mean and variance values of the updated hydraulic conductivity fields for the ensemble size of 70, 90, 100, 200 are almost the same with the $RMSE K$ more than 10 [L/T] at some assimilation steps, I only plot the ensemble size of 90 as a representative for simplicity in Figures 5 and 6. Similarly, I only plotted the ensemble size of 300 and 500 as representatives. The initial ensemble mean hydraulic conductivity for different ensemble size (the value is an approximate constant as 13.10 [L/T]) is apparent at the first time step in Figure 5. It is shown that the first 2 or 3 time steps of data assimilation are the most efficient for every ensemble size simulation, with decrease of $RMSE K$ value during these time steps. Also, I found that after the first 5 time step, the effect of assimilating observation data became worse (Figure 6) for different ensemble sizes 300 and 500. So, it is better for us to assimilate only a few time step observation data. However, in Huang et al. (2009), the more data assimilation time steps, the more improved the updating hydraulic conductivity fields become. The criterion to end the process of data assimilation is introduced as:

$$DifK = \sqrt{\frac{1}{NM} \sum_{i=1}^M \sum_{j=1}^N (K_i^j - \bar{K}_i)^2} \leq 10^{-2} \quad (11)$$

Where $DifK$ is the mean difference of the updated value and the corresponding ensemble mean value over the domain; K_i^j is the model result of the j th ensemble for the i th grid; \bar{K}_i is the ensemble mean value at the i th grid. After the end of the data assimilation process, I consider the latest update as providing the most acceptable results.

It is very clear that the mean and variance values of hydraulic conductivity fields with data assimilation for the ensemble size of 300 and 500 are much closer to the real hydraulic conductivity field. This demonstrates that the data assimilation method for the transient flow via EnKF in this study can update the model parameters (hydraulic conductivities here) well. Moreover, the larger the ensemble size, the better the updated hydraulic conductivity field, which agrees with the results from Chen and Zhang (2006), Evensen (2003, 2004) and Huang et al. (2009). It is well known that with the increase of ensemble size, more effort will be needed for computation. Since the ensembles consisting of 300 and 500 members (realizations) yield similar results (the mean and variance values are very close to the real hydraulic conductivity field at a sufficient number of time steps), an ensemble consisting of 300 members is considered to be an appropriate number of realizations for both the results and computational efficiency. To compare the spatial distribution of the updated hydraulic conductivity fields under different assimilation steps with the real field, the spatial distributions of the log hydraulic conductivity field under the first and fourth time steps are plotted in Figures 7(a) and 7(b), respectively, with ensemble members of 300 and 25 observation wells. According to the plot, it can be seen that after four time step data assimilation, the updated mean hydraulic conductivity field is much better than the initial one shown in the Figure 1(c). Furthermore, it is very like the real field (Figure 1(b)), which suggest that the 2D saturated transient flow simulation by assimilating easily observed hydraulic head data can not only update the hydraulic head in time, but is also an efficient method to update other hydraulic parameters such as the hydraulic conductivity.

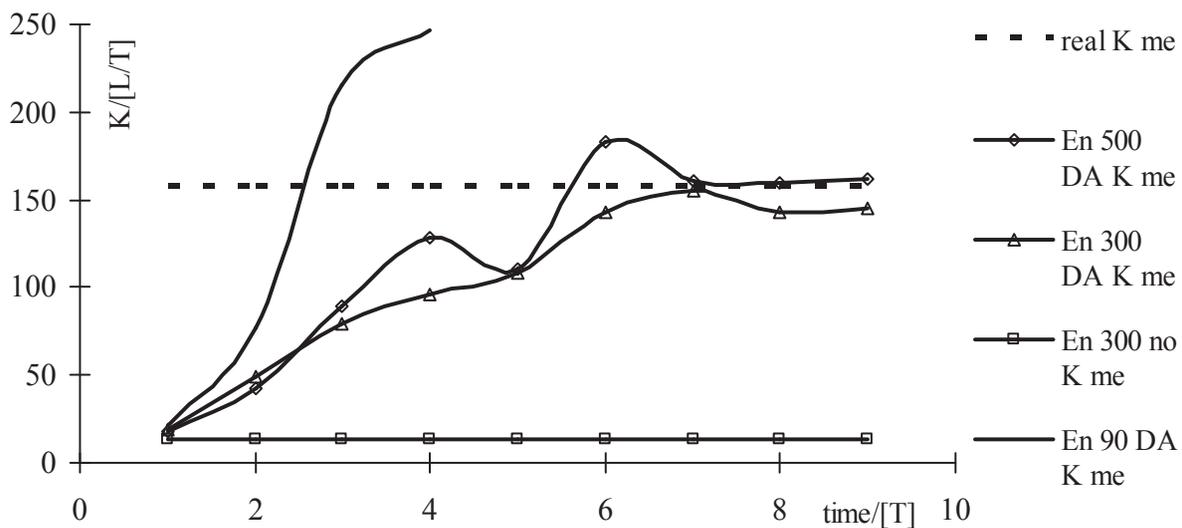


Figure 5. Hydraulic conductivity mean values of the real field, initial guessed field and calibrated fields through data assimilations with various ensemble sizes and under various time steps

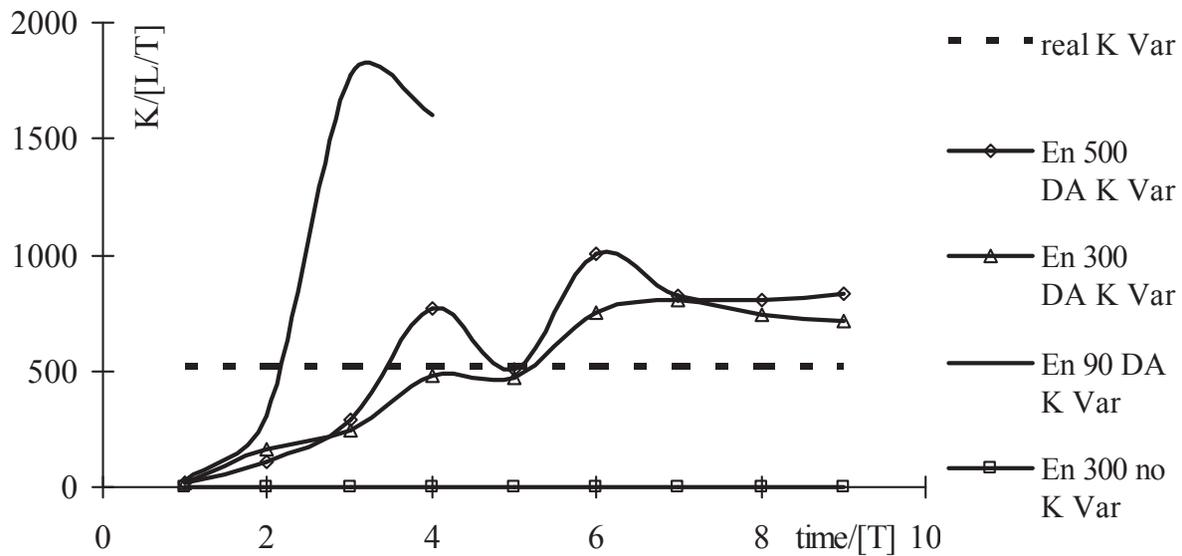


Figure 6. Hydraulic conductivity variance values of the real field, initial guessed field and calibrated fields through data assimilations with various ensemble sizes and under various time steps

4.1.2. Influence of boundary condition

In this section, the ensemble size of 300 is chosen as discussed above, 25 observation wells is selected to update the hydraulic conductivity by assimilating the observed hydraulic head data for different boundary conditions. The first condition is the one used in Case 1, four no-flow boundaries, the assimilation results have obtained and shown in Figure 7(a), (b).

Since the no-flow boundaries are not common in reality, the effects of non-zero flow boundary condition on data assimilation results are explored. In the second condition, we assume that there is a constant flow flux at each Neumann boundary: 2 [L/T] at the top boundary, 1 [L/T] at the bottom boundary, 1.5 [L/T] at the right boundary, -1.5 [L/T] at the left boundary, where the positive flux value means flow out of the study domain, while the negative value means flow into the domain. Under this boundary condition and with the same pumping rate as the first condition, the updated conductivity fields after the first and the fourth time step assimilations are shown in Figures 7(c) and 7(d). In comparison of Figure 7(b) and 7(d) with the real conductivity field in Figure 1(b), one can conclude that the updated conductivity field is similar to the real field only after four data assimilation steps when the boundaries are Neumann boundary condition, whether it is no flow or non-zero flow condition.

The third boundary condition is the one used in the Case 2, two no-flow boundaries and two constant hydraulic head boundaries. The updated hydraulic conductivity fields after 3 different time steps of data assimilation are shown in the Figure 8. It is shown from the figure that the updated hydraulic conductivity field is not similar to the real field even after the 50th assimilation step. The results shown in Figures 7 and 8 indicate that this data assimilation method for transient flow works better for the study domain with Neumann boundaries than that with mixed no-flow and constant head boundaries. In other words, data assimilation

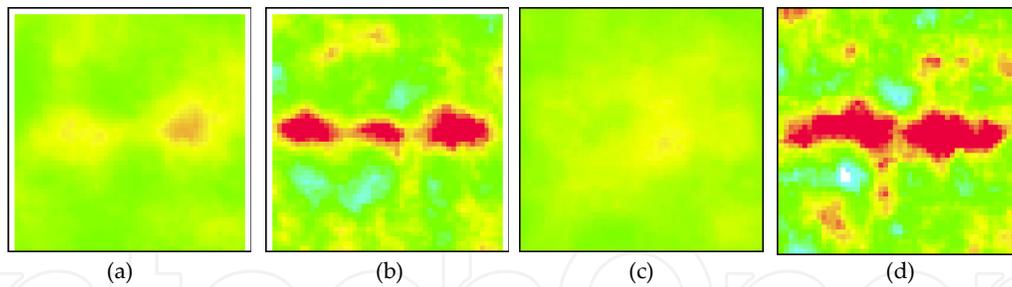


Figure 7. Assimilated logK field at 1th (a) and 4th step (b) in Case 1, and at 1th (c), 2th (d) and 4th (e) step with non-zero flow at the boundaries for ensemble size of 300 with 25 observation wells

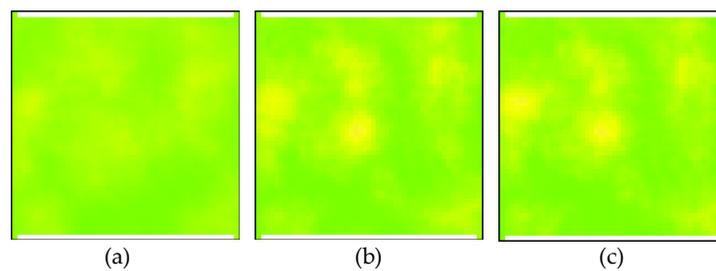


Figure 8. Identified logK in Case 2 for ensemble size of 300 with 25 observation wells at different assimilation steps: (a) the first step; (b) the 5th step; (c) the 50th step

method for transient flow does not work well for the study domain with mixed no-flow and constant head boundaries. These results can be explained by the boundary hydraulic characteristics. For the Neumann boundary condition, hydraulic head variation is directly related to the hydraulic conductivity distribution, without being constrained by constant head boundaries. While, for transient flow with a constant-head boundary, the hydraulic head variation will not be fully determined by the hydraulic conductivity field due to the head constraints imposed at the boundaries. Using extreme cases as examples, when the hydraulic conductivity is extremely large or small in the study domain with a constant-head boundary, the hydraulic head in the boundary is still constant. Therefore, the hydraulic conductivity field is not the only factor that determines the hydraulic head distribution in the transient flow domain, and the constant-head boundary is also a strong influence on the hydraulic head change in the study domain. Therefore, the constant head boundary condition will decrease the sensitivity of the transient hydraulic head change with hydraulic conductivity.

4.1.3. Impact of number of observation wells

To investigate the effect of the number of observation wells on the data assimilation, in addition to the conductivity field in Figure 7 calibrated through 25 observation wells, the data assimilation method is applied to Case 1 with 9, 16, 36, 49, 64 and 81 observation wells, where the observation wells are uniformly distributed in the study domain. The study results are shown in Figure 9.

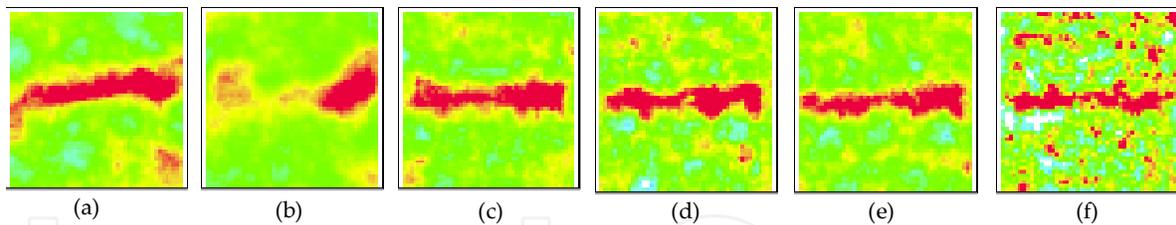


Figure 9. Calibrated hydraulic conductivity field with various observation wells: (a) 9 wells; (b) 16 wells; (c) 36 wells; (d) 49 wells; (e) 64 wells; and (f) 81 wells.

Comparing the real hydraulic conductivity field, the initial mean hydraulic conductivity field with the calibrated hydraulic conductivity fields with various number of observation wells shown in Figure 7 and 9, one can tell that with the increase of the observation wells, the updated hydraulic conductivity distribution will become closer to the real hydraulic conductivity field until the number of observation wells reaches 64. As shown in Figure 9 that when the observation wells are limited, such as 9 and 16 wells, the updated hydraulic conductivity is not close to the real hydraulic conductivity field, and the high conductivity channel is not well captured. However, when the observation wells become too many, such as 81 wells, the true information is too much, which will disturb the assimilation results and lead to “bad” calibration results. This phenomenon can be explained as one observed data from one observation well will influence a certain space around the well via EnKF by data assimilation. If the observation wells are too close to each other, some area will be calibrated by data from two wells at the same time. However, the observation wells and their influencing regions are considered to be independent in the data assimilation method, so, too many observation wells lead to over calibration of the hydraulic conductivity field. Moreover, the suitable uniform distance of two observation wells is still a problem and how to define the appropriate data assimilation frequency and space is an issue that requires further study.

From the assimilation results it is told that 25, 36, 49 and 64 observation wells are suitable for the calibration of the conductivity field with an ensemble size of 300, which is composed of 2601 numerical grids cells in the study area. In general, these results demonstrate that the data assimilation method using an appropriate number of observation wells for transient flow to calibrate a heterogeneous conductivity field for an ensemble size of 300 via EnKF is effective and convenient. Furthermore, it displays similarities with the real field.

4.1.4. Influence of boundary condition

To further investigate the influence of the observation wells on identification of a conductivity field from transient flow data via EnKF, three different observation well locations are set in case 1 with the ensemble size of 300. As shown in Figure 10(a)-(c), there are the same number of 16 hydraulic head observation wells in the study field, but their locations are distributed with three different patterns, denoted 16, 016, 0016, and their corresponding updated hydraulic conductivity field distributions through assimilating the observed hydraulic head data are displayed in the Figures 9(b), 11(a) and 11(b).

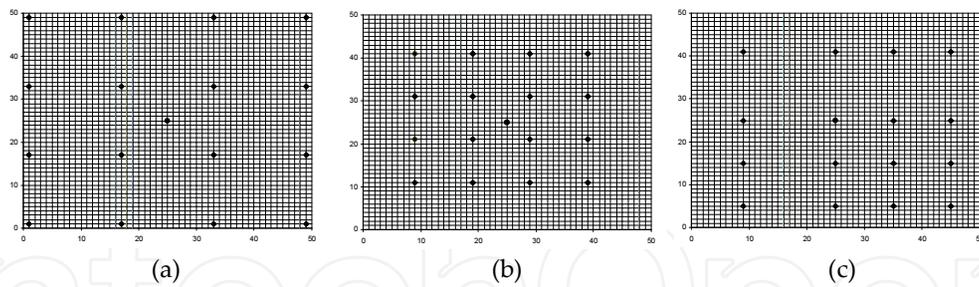


Figure 10. Different distribution patterns for 16 observation wells: (a) 16; (b) 016; and (c) 0016 (Diamonds are observation wells while the circle is the pumping well.)

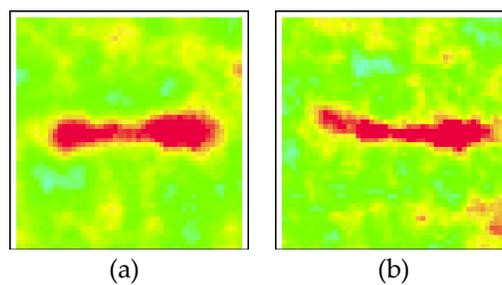


Figure 11. Calibrated hydraulic conductivity field with different distributions of 16 observation wells: (a) 016; and (b) 0016

In the first pattern shown in Figure 10(a), the 16 hydraulic head observation are uniformly distributed in the whole study domain, but there is not a single observation well located within the high hydraulic conductivity zone. As discussed above, the updated hydraulic conductivity field shown in Figure 9(b) does not represent the existence of the high conductivity zone well. The observed data will influence the updated data around the observation locations (Ott et al., 2004), so the updated hydraulic conductivity distribution cannot capture the zone of high hydraulic conductivity.

For the observation wells shown in Figure 10(b), they are still uniformly distributed. Although the observation well locations are still not in the high hydraulic conductivity zone, they are around that zone. Therefore, the updated hydraulic conductivity field shown Figure 11(a) is similar to the real field. The third distribution pattern of the observation wells is shown in Figure 10(c), and the wells are not distributed uniformly, and are located closer to each other on the lower right side while some are in the high hydraulic conductivity zone. Similar to the Figure 11(a), the updated hydraulic conductivity field shown in Figure 11(b), is very similar to the real conductivity field.

The study results indicate that the observation well locations will influence the assimilation results of the hydraulic conductivity field. Better results are obtained if the observation wells sample zones of distinct conductivity. Optimal design of observation well locations to maximize data efficiency is a research issue requiring further research.

4.1.5. Conclusions

In this study, the ensemble Kalman Filter (EnKF) is used to estimate heterogeneous hydraulic conductivity distributions by assimilating hydraulic head measurements in a transient flow pumping field. The measurement error is considered, but the model error is not. The forecast model is assumed to be known. Two synthetic cases with different boundaries are designed to investigate the capacity and effectiveness of the proposed data assimilation method to identify a conductivity distribution. The developed method is suitable for 2 or 3-D flow field, but only demonstrated on 2-D transient flow. In case 1, the four sides are Neumann boundaries with a constant pumping rate. While in case 2, the left and right sides are constant head boundaries and the top and bottom sides are still no-flow boundaries and the pumping rate changes every 5 [T]. The head measurements are used to identify a heterogeneous hydraulic conductivity field. Based on the study results, the following conclusions are obtained.

1. The EnKF can be used to effectively calibrate a heterogeneous conductivity distribution by assimilation of hydraulic head measurements in a transient flow field with no-flow boundaries. Only after a few (no more than 10) assimilation steps, the spatial distributions of hydraulic head and hydraulic conductivity become significantly closer to the real field distributions than those without data assimilation.
2. The model predictability depends on ensemble size. For our study cases, 300 realizations were found to be sufficient for transient flow calculation.
3. Using the EnKF to calibrate a heterogeneous conductivity field through assimilating transient head data, the calibration is very efficient for the study domain with all Neumann boundaries, while it is not efficient with any constant boundaries in study domain. The boundary condition will significantly affect the assimilation results.
4. The appropriate number of observation wells for our case study is 25, 36, 49 and 64 respectively. Compared with the total 2601 grids in the study area, the observation well number is still small. When well number is too small, the data cannot capture the variation, but when the number is too big, the data will be redundant and cause over calibration.
5. The location of the observation wells will also significantly influence the calibration results. Better results will be obtained if the observation wells adequately sample the heterogeneities. For the study case, if the observation wells are located in both high and low conductivity regions, the calibration results will be better than if they are only located in one region.
6. A criterion is introduced to determine the end of the data assimilation process, but how to develop a more suitable criterion needs further study. That criterion will allow the identification of the point where continued data assimilation will deteriorate the quality of the results.
7. The simulation time of the assimilation steps is very short, and this is a problem requires our deeper thinking. The model error is not considered here, which will pose too much confidence to the model estimate. Maybe the latter one is the cause of the former problem.

However, the model error is unknown here, so how to choose a model and determine the model error is another issue needs further thinking.

4.2. Transient state conservative solute transport

Similar to part 4.1, it is assumed that the solute concentration samples can be obtained in the observation wells with every simulation time step, so in this part the observed data will be assimilated through the constrained EnKF method (described in section 3.2) at every simulated time. Since the ensemble size of 300 is enough in part 4.1, 300 realizations will be used in this part too.

4.2.1. Small initial distribution of solute concentration

The solute concentration can be predicted after the water velocity field is obtained, and the initial instant injection is given. The solute concentration distribution of the initial instant injection is shown in Figure 12. The initial injection area is very small in Figure 12(a), and the larger on is in Figure 12(b). There are 81 observation wells as shown in Figure 3(c). I plot the mean and variance values of the hydraulic conductivity field with and without data assimilation under 300 ensemble sizes with small area of initial solute concentration in Figures 13 and 14, respectively. I compare the real conductivity field with simulated fields with and without the constrained EnKF method. The initial ensemble mean hydraulic conductivity under 300 ensemble size is 13.35 [L/T], as shown in Figure 13, while the real field mean is 7.92 [L/T]. It is shown in Figure 13 that in comparison with the initial conductivity field, the updated mean conductivity through the data assimilation method is much closer to the real mean value. However, some updated results could change from “good” at the first several steps to “bad” after many assimilation steps. The results at the first five or seven assimilated time steps are the most efficient. In Figure 14, the variance value of the updated result approaches the real value after several assimilation steps, however, the variance will continue increasing and deviates from that in the real field as time goes. The results are similar to the results in Figure 13. The *RMSE K* value also decreases in the first several time steps and then increases. The results suggest that it is only meaningful to use the data assimilation method to update parameters for the first several assimilation time steps. If we just take the results from the first several assimilation steps, the updated hydraulic conductivity field is significantly improved in comparison with the initial guessed field.

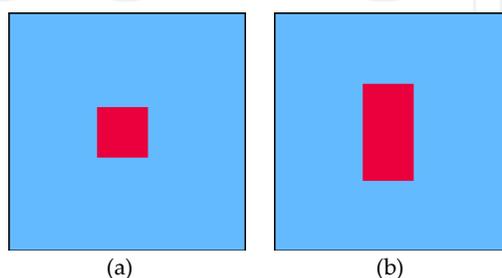


Figure 12. Solute concentration distribution of the initial instant injection (red zone): (a) small distribution; and (b) large distribution

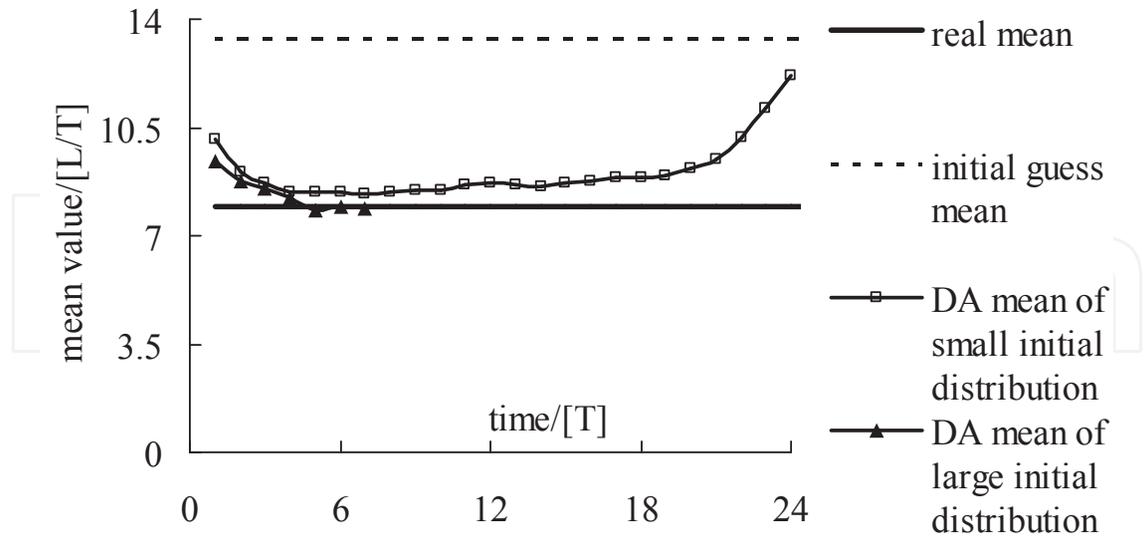


Figure 13. Mean values of hydraulic conductivity field: the real field, without and with data assimilation with small and large area of initial solute concentration under 81 observation wells

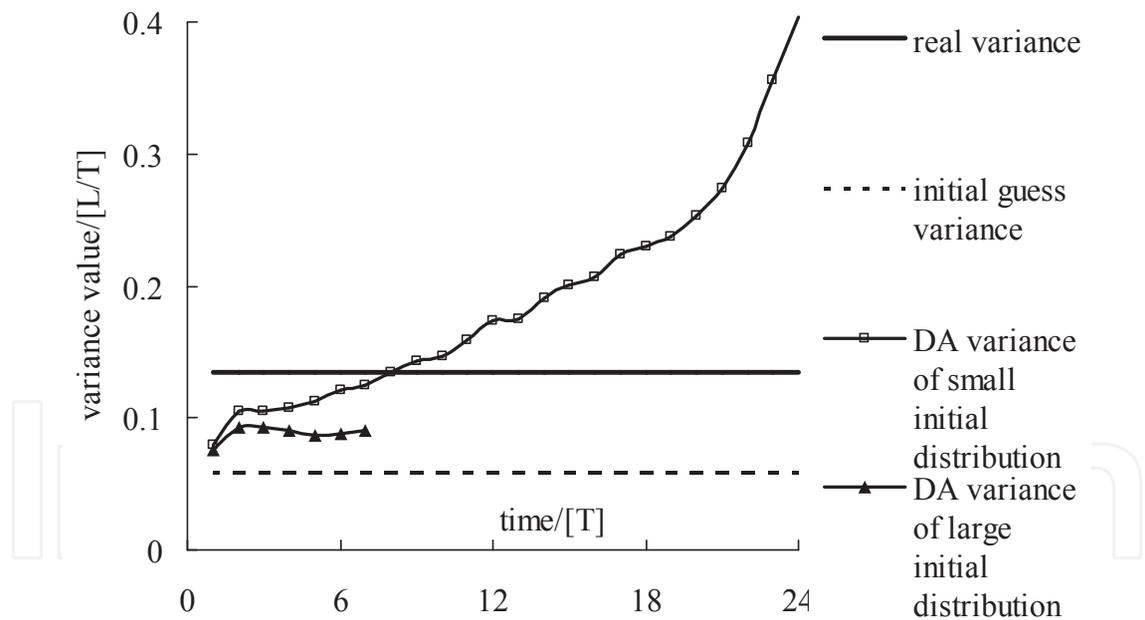


Figure 14. Variance values of hydraulic conductivity field: the real field, without and with data assimilation with small and large area of initial solute concentration under 81 observation wells

We plot the spatial distributions of the updated hydraulic conductivity field at various time steps in Figure 15, and the corresponding absolute hydraulic conductivity difference between the updated result and the real one is presented in Figure 16. The initial ensemble mean guess of hydraulic conductivity field (Figure 3(b)) is uniform, while the real field is heterogeneous

and shown in Figure 3(a). The spatial distribution of updated hydraulic conductivity field in Figure 15 have captured the feature of the real field, two low hydraulic conductivity zones in the middle and two high hydraulic conductivity zones at the right-top and left-middle areas.

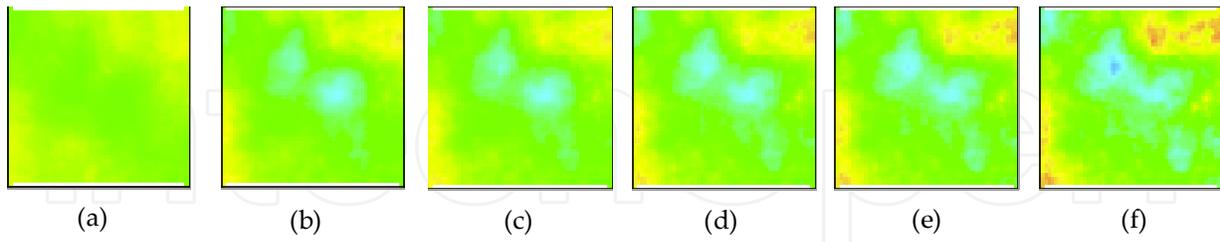


Figure 15. Assimilated $\log K$ field at different assimilation steps with small area of initial solute concentration under 81 observation wells: (a) the 1th step; (b) the 4th step; (c) the 5th step; (d) the 7th step; (e) the 8th step; (f) the 11th step

From Figure 16, it is can be seen that the updated hydraulic conductivity fields are better than the initial guess, and the hydraulic conductivity has been significantly improved in the middle area. This is explained as that the observed data are mainly obtained from the middle area during the simulation time. However, the constrained EnKF method overestimates the high hydraulic conductivity zone in the left-bottom and the right-top areas (Figures 15 and 16). The reason for the over estimation will be explained later, which leads to the deterioration of the results of the mean and variance values shown in Figure 13 and 14. It is also noticed that observed data of only 7 to 9 can be obtained from the observation wells because the solute does not spread to all domains at any time step, and solute duration time at any location is limited. Therefore, for any given well, if too many time step observations are used to update the conductivity field, the concentration values in some steps will be zero, such as in the time period before the solute reaches the well point or the period after the solute passes that point. If these void concentration data are used to update the conductivity field, the data will deteriorate the assimilation results. For the study case, the solute passes through some observation wells in a few time steps, so only the few time step data should be assimilated. In general, the EnKF can significantly improve the hydraulic conductivity field by assimilating 7 to 9 observed solute concentration data, which demonstrates that the constrained EnKF is an efficient method to update the hydraulic conductivity field via assimilating the observed solute concentration available.

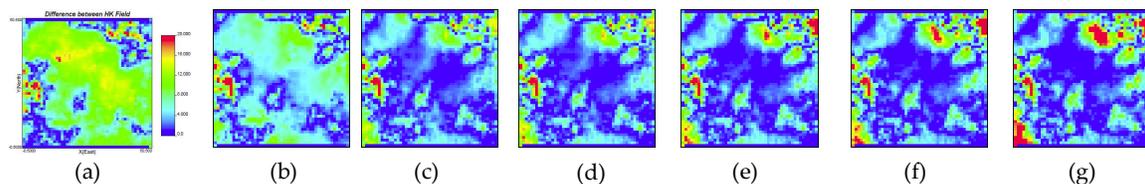


Figure 16. Absolute difference between real and updated or initial guess hydraulic conductivity field with small area of initial solute concentration under 81 observation wells: (a) initial guess; (b) 1th assimilation step; (c) 4th assimilation step; (d) 5th assimilation step; (e) 7th assimilation step; (f) 8th assimilation step; (g) 11th assimilation step

(Note: all the Figures of absolute difference of hydraulic conductivity fields for the conservative solute transport here and next have the same color scale)

Moreover, compared with 2601 grid elements in the study domain, the number of the observation data (7 to 9) is very small, which is very convenient in the reality. In part 4.1, From the study on updating conductivity field through assimilating hydraulic head data, it is found that 25 observed data (observation wells) are needed to capture major feature of the same conductivity field as the one used in this study. Therefore, the concentration data are more efficient than hydraulic head data. This is because the solute concentration in the middle part will be influenced by the groundwater velocity in the upstream and downstream areas. The hydraulic conductivity field will affect the hydraulic head field through the flow model, and the velocity is controlled by both the hydraulic head and conductivity fields through the Darcy's law (see Equation (2)). Thus, the groundwater velocity is directly influenced by the hydraulic conductivity field. In this study, the solute transport is dominated by the advection, so the solute concentration will be very sensitive to the hydraulic conductivity in the upstream and downstream areas, and the observed solute concentration data will contain the information of the hydraulic conductivity distribution. Furthermore, the dispersivity parameters are also related to the groundwater velocity closely in the solute transport modeling, which suggests that the dispersion part of the solute transport is also sensitive to the hydraulic conductivity indirectly. So only 7 to 9 observed solute concentration data in the middle part can be used to improve the hydraulic conductivity distribution not only in this middle part but can also in the left and right areas.

In part 4.1, it was already shown that the EnKF method by assimilating the hydraulic head data is not very efficient for the study domain with the mixed boundaries, no-flow boundary at the top and bottom borders and constant head boundary at the left and right borders. However, the updating results are very significant with the same hydraulic boundary conditions in this study. Therefore, it is can be concluded that updating a conductivity field through the constrained EnKF method via assimilating the solute concentration data is much more efficient than via assimilating the hydraulic head data.

4.2.2. Large initial distribution of solute concentration

To further explore the capability of the constrained EnKF to update the hydraulic conductivity field via assimilating solute concentrations, it is necessary to add more observation data. So the initial distribution of the solute area is extended from the one in Figure 12(a) to the one in Figure 12(b). In this way, the solute plume will cover a larger area and more observation wells can sample the solute concentration. The mean and variance values of the hydraulic conductivity field with and without data assimilation are plotted in Figures 13 and 14 at the first few and different assimilation steps, and compare them with the values in the real hydraulic conductivity field.

In Figure 13 and 14, it is shown that both the mean and variance values of the updated hydraulic conductivity field by the constrained EnKF become closer to the real ones than those of the initial guess field. Moreover, the *RMSE K* value at different assimilation time is less than that of the initial guess, which indicates that the updated hydraulic conductivity field is better than

the initial guess hydraulic field. The number of the wells from which the plume can be observed during the application of the constrained EnKF by assimilating the solute concentration varies from 13 to 16.

The spatial distributions of the updated hydraulic conductivity field and the corresponding absolute hydraulic conductivity difference between the updated result and the real field at some time steps are shown in Figures 17 and 18. From the figures, it can be seen that the updated hydraulic conductivity field captures well the main feature of the real field, and the updated results are much better than the initial guess. The low hydraulic conductivity zones of the real field are together in the middle area, while it is also indicated in the updated fields as shown in Figure 17. Like the real field, the high hydraulic conductivity zones can also be seen in the right-upper and the left-middle areas of the updated ones. In comparison with the 7th step updated results with a smaller initial plume size, shown in Figures 15(d) and 16(e), and with a larger initial plume size, Figures 17(c) and 18(c), the larger the initial plume size, the slightly better the updated results.

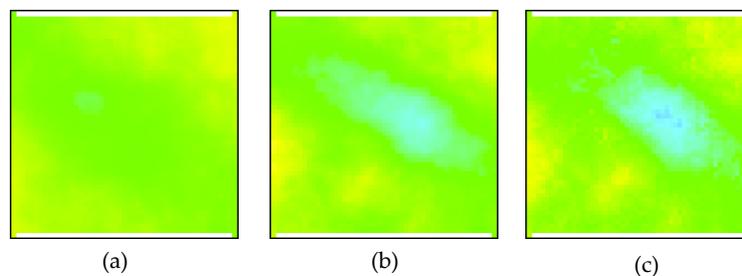


Figure 17. Assimilated $\log K$ field at different assimilation steps with large area of initial solute concentration under 81 observation wells: (a) the 1th step; (b) the 4th step; (c) the 7th step

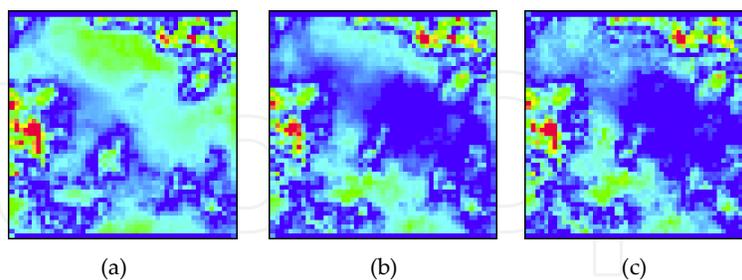


Figure 18. Absolute difference between real and updated hydraulic conductivity field with large area of initial solute concentration under 81 observation wells: (a) 1th assimilation step; (b) 4th assimilation step; (c) 7th assimilation step

What we want to mention is that the observation wells are uniformly distributed in space, which is very easy to conduct in practice, and the hydraulic conductivity field can be improved by obtaining solute concentration from these wells. If the concentration data are available in all the observation wells, it is not necessary to get the data at every observation well. They can be gotten selectively, that is to say that some observed data at these wells can be gotten at this

time, and some observed data at some different wells (not the same observation wells) can also be gotten at the next time. But how to determine the criteria which time is better to obtain observed data from which observation wells is an issue requiring further research.

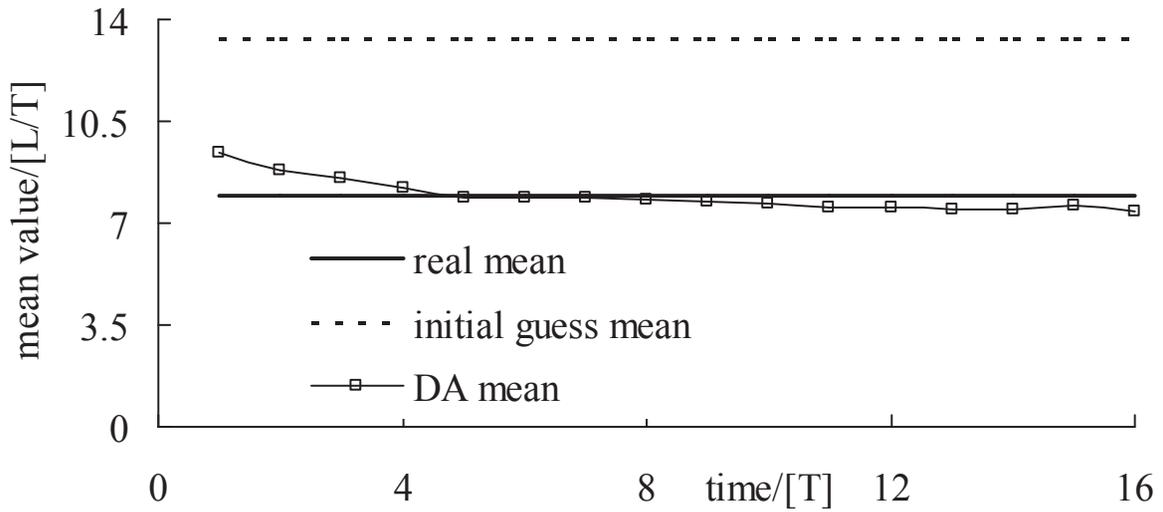


Figure 19. Mean values of hydraulic conductivity field: the real field, without and with data assimilation and large area of initial solute concentration under 81 observation is 5% of the solute concentration measurements

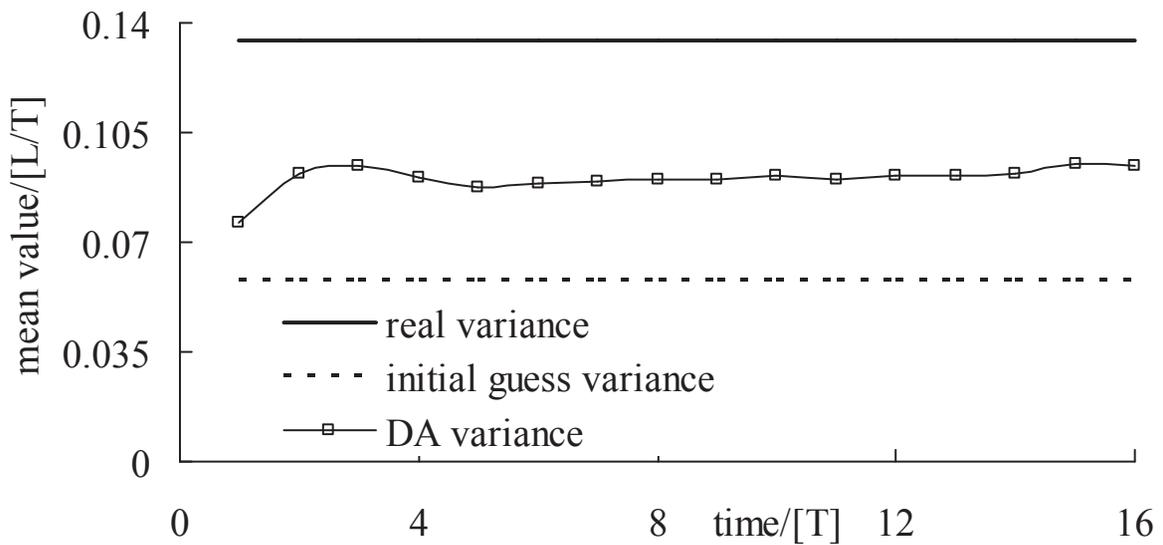


Figure 20. Variance values of hydraulic conductivity field: the real field, without and with data assimilation and large area of initial solute concentration under 81 observation wells while the standard deviation of the observation error is 5% of the solute concentration measurements

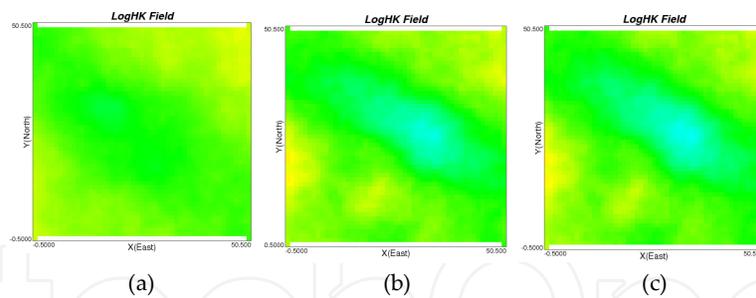


Figure 21. Assimilated $\log K$ field at different assimilation steps with large area of initial solute concentration under 81 observation wells while the standard deviation of the observation error is 5% of the solute concentration measurements: (a) the 1th step; (b) the 4th step; (c) the 7th step

However, maybe the standard deviation of the observation error as 1% of the solute concentration measurements is too small, the standard deviation of the observation error is increased to 5% of the solute concentration measurements. The calculated results of the mean and variance of the hydraulic conductivity are shown in Figures 19 and 20. From figures, it is found that the mean and variance of the updated hydraulic conductivity field via the data assimilation method is closer to that of the real hydraulic conductivity field than the initial guess conductivity field. The spatial distribution of the updated hydraulic conductivity field at different assimilated time steps are plotted in Figure 21. Comparing Figure 21 with Figure 17, it is found that these results are very similar. So the standard deviation of the observation error does not affect the simulation results very much here from the first view. Therefore, the absolute hydraulic conductivity difference between the updated result and the real field are not plotted for simplicity.

To have a further look at the influence of the standard deviation of the observation error on the updated results, the mean, variance and spatial distribution of the updated hydraulic conductivity field at different assimilated time steps with the standard deviation of the observation error be 10% of the solute concentration measurements are also plotted in Figures 22-24. From the mean and variance values of the updated hydraulic conductivity field in Figures 22 and 23, it is seen that the simulated time become longer and the results by the data assimilation method via the constrained EnKF is much closer to the real conductivity field than the initial guess even the standard deviation of the observation error is 10% of the observation measurements. Compared with Figures 13 and 14, it can be even seen that the results with the standard deviation of the observation error be 10% of the solute concentration measurements are much better than that of 1%, and the simulated time is also longer here. This phenomenon can be explained that maybe the constraints posed to the EnKF method have introduced too much potential error to the model, which lead to the avoidance of the filter divergence to a certain degree and longer simulation time, and also the increase of the observation error will give a tradeoff between the model error and the observation error. Furthermore, it is seen that the spatial distribution of the updated hydraulic conductivity fields still similar to that in Figure 17. Therefore, it is can proved that the standard deviation of the observation error does not affect the simulation results very much in a deeper step.

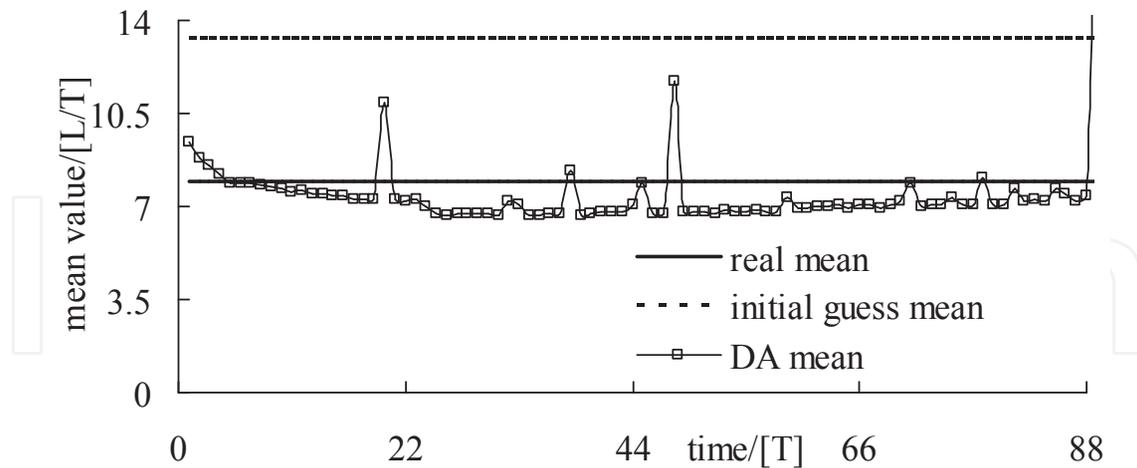


Figure 22. Mean values of hydraulic conductivity field: the real field, without and with data assimilation and large area of initial solute concentration under 81 observation wells while the standard deviation of the observation error is 10% of the solute concentration measurements

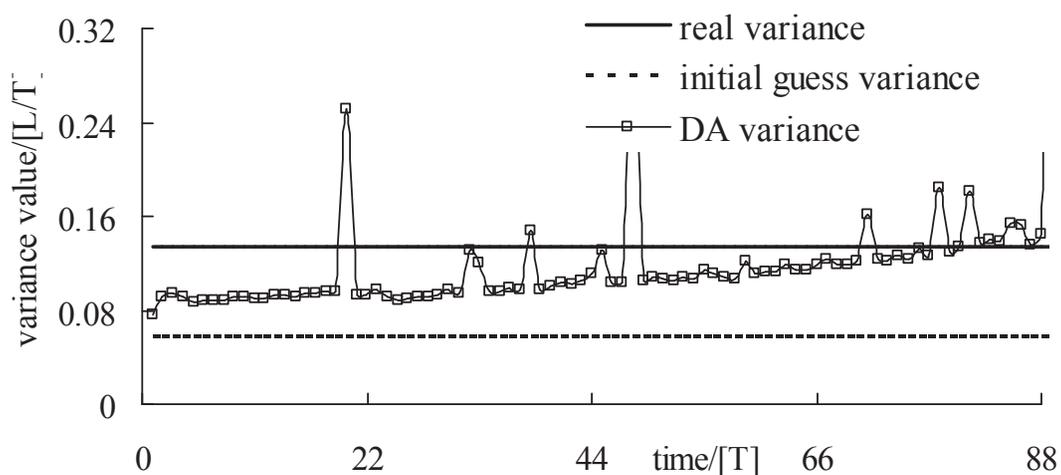


Figure 23. Variance values of hydraulic conductivity field: the real field, without and with data assimilation and large area of initial solute concentration under 81 observation wells while the standard deviation of the observation error is 10% of the solute concentration measurements

The *RMSE* values for the updated hydraulic conductivity field under different percentage (3%, 5%, 9%, 10%) standard deviation of the observation error of the solute concentration measurements are also plotted in Figure 25. From the figure, it can be seen that the *RMSE* values will increase as the assimilated time increase for different perturbations to the observation error. The larger perturbations added to the observation errors, the greater the *RMSE* values for the updated hydraulic conductivity fields. It is seen that if the standard deviation of the observation errors is greater than 5% of the measurements, the *RMSE* values for the updated hydraulic conductivity have the trend to increase. So it is better to use the standard deviation

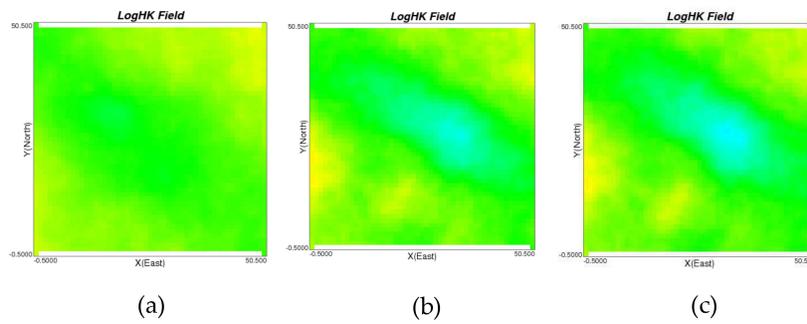


Figure 24. Assimilated $\log K$ field at different assimilation steps with large area of initial solute concentration under 81 observation wells while the standard deviation of the observation error is 10% of the solute concentration measurements: (a) the 1th step; (b) the 4th step; (c) the 7th step

of the observation errors no more than 5% of the measurements if I want to have less *RMSE*. Furthermore, it can also be found that the plots of the mean, variance and *RMSE* values for the updated hydraulic conductivity field with the standard deviation of the observation errors as 10% of the measurements in Figures 22, 23 and 25 are very identical, they almost fluctuate at the same time. This can also be seen from the former results in section 4.1, where the *RMSE* values become very high at the later assimilated time steps like the corresponding mean or variance values and they are not plotted there.

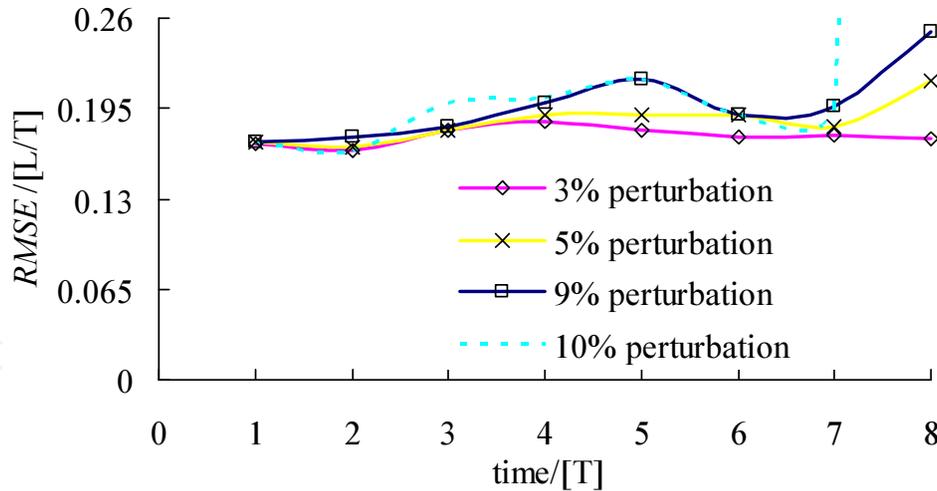


Figure 25. *RMSE* values of the assimilated hydraulic conductivity field with large area of initial solute concentration under 81 observation wells under different standard deviation of the observation errors

4.2.3. Conclusions

In this study, the constrained EnKF is used to estimate heterogeneous hydraulic conductivity distributions by assimilating solute concentration measurements for transient solute transport in a steady state flow field. The measurement error is considered, but the model error is not.

The forecast model is assumed to be known. A synthetic case of a rectangular field with the no-flow boundary at the upper and bottom borders and constant head boundary at the left and right borders is designed to investigate the capacity and effectiveness of the data assimilation method to identify a hydraulic conductivity distribution. The developed method is only demonstrated on 2-D transient solute transport with two different areas of initial instant solute source in the middle area of the field. The solute concentration measurements are used to identify the heterogeneous hydraulic conductivity field. Based on the study results, the following conclusions are obtained.

1. The constrained EnKF can be used to effectively calibrate a heterogeneous conductivity distribution by assimilating solute concentration measurements for a transient transport in a steady state flow field with the mixed no-flow boundary and constant head boundary. Only after a few (no more than 10) assimilation steps, the spatial distributions of hydraulic conductivity become significantly closer to the real field distribution than those without data assimilation.
2. The model predictability depends on the initial distribution of the solute concentration. The larger area of the initial distribution, the more observed data can be obtained, the better the calibration results.
3. Due to the strong sensitivity of solute concentration to the hydraulic conductivity field, less observation wells are needed to achieve the same calibration results for a hydraulic conductivity field through assimilating solute concentration data than assimilating hydraulic head data.
4. The data assimilation method can produce useful results in the first five or seven time step assimilation.
5. In study of part 4.1, the EnKF method was used to calibrate a conductivity field by assimilating hydraulic head data. Their study results indicate that the EnKF method cannot well capture a conductivity distribution in a field with constant head boundaries through assimilating hydraulic head data; while this study results indicate the EnKF method can through assimilating concentration data.
6. Maybe the constraints posed to the EnKF method have introduced too much potential error to the model, which lead to the avoidance of the filter divergence to a certain degree and longer simulation time, and also the increase of the observation error will give a tradeoff between the model error and the observation error. So the standard deviation of the observation error varies from 1% to 5% even to 10% of the solute concentration measurements, the simulated results by the data assimilation method are still very similar.
7. The assimilated simulation steps are not very long, this problem may be caused by ignoring the model error. Maybe in future, the model error (which is not known right now) can be added. Moreover, the data assimilation method just provides us the best estimate values for our study case, but it does not consider the mass balance during the simulation automatically.

Acknowledgements

This work is partly supported by the National Nature Science Foundation of China Major Research Project (Grant No. 91125024), the National Nature Science Foundation of China (Grant No. 51209187), and the Fundamental Research Funds for the Central Universities (Grant No. 2652011286).

Author details

Bill X. Hu^{1,2} and Juxiu Tong¹

1 Collage of Water Resources and Environmental Sciences, China University of Geosciences, Beijing, China

2 Department of Earth, Ocean and Atmospheric Sciences, Florida State University, Tallahassee, USA

References

- [1] Andreadis KM, Lettenmaier DP (2006) Assimilating remotely sensed snow observations into a macroscale hydrology model. *Adv. Water Resour* 29(6): 872-886
- [2] Arlen WH (2005) The U.S. Geological Survey Modular Ground-Water-the Ground-Water Flow Process. U.S. Geological Survey, Reston
- [3] Aubert D, Loumagne C, Oudin L (2003) Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model. *Journal of Hydrology* 280: 145-161
- [4] Bear J (1972) *Dynamics of Fluids in Porous Media*. New York, Elsevier Pub. Co., Inc.
- [5] Burgers G, van Leeuwen PJ, Evensen G (1998) Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review* 126: 1719-1724
- [6] Chen Y, Zhang D (2006) Data assimilation for transient flow in geologic formations via Ensemble Kalman Filter. *Adv. Water Resour.* 29: 1107-1122
- [7] Christakos G (2005) Methodological developments in geophysical assimilation modeling. *Reviews of Geophysics* 43: 2004RG000163, 1-10
- [8] 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

- [9] Cohn SE (1997) An introduction to estimation theory. *J. Meteor. Soc. Japan* 75(1B): 257-228
- [10] Constantinescu EM, Sandu A, Chai TF, Carmichael GR (2007) Assessment of ensemble-based chemical data assimilation in an idealized setting. *Atmospheric Environment* 41: 18-36
- [11] Daley R (1991) *Atmospheric Data Analysis*. New York: Cambridge University Press, pp. 457
- [12] Drecourt JP (2003) Kalman filtering in hydrologic modeling. DAIHM Technical Report, May 20
- [13] 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
- [14] Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* 99 (C5), 10.143-10.162
- [15] Evensen G (2003) The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics* 253: 343-367
- [16] Evensen G (2004) Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics* 54: 539-560
- [17] Fang F, Piggott MD, Pain CC, Gorman GJ, Goddard AJH (2006) An adaptive mesh adjoint data assimilation method. *Ocean Modeling* 15: 39-55
- [18] Fisher M (1998) Development of a simplified Kalman filter. ECMWF Research Department Tech. Memo. 260, 16 pp [Available from European Center for Medium-range Weather Forecast, Shin-field Park, reading, Berkshire, RG2 9AX, United Kingdom.]
- [19] Gelb A (1974) *Applied optimal estimation*. MIT Press, Cambridge, MA
- [20] Hemant AP, Oliver DS (2009) Data assimilation using the constrained ensemble Kalman filter (SPE 125101). In SPE Annual Technical Conference, 4-7 October 2009, New Orleans, Louisiana, U.S.A.
- [21] 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
- [22] Houtekamer PL, Mitchell HL (1998) Data assimilation using an ensemble Kalman Filter technique. *Monthly Weather Review* 126: 796-811
- [23] Houtekamer PL, Mitchell HL (2001) A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.* 129: 123-137

- [24] Hu BX, Meerschaert MM, Barrash W, Hyndman D, He C, Li X, Guo L (2009) Examining the influence of heterogeneous porosity fields on conservative solute transport. *J. Contaminant Hydrology* 108: 77-88.
- [25] Huang C, Bill XH, Li X, Ye M (2009) Using data assimilation method to calibrate a heterogeneous conductivity field and improve solute transport prediction with an unknown contamination source. *Stochastic Environment Research and Risk Assessment* DOI 10.1007/s00477-008-0289-4
- [26] Jazwinski AH (1970) *Stochastic Processes and Filtering Theory*. Elsevier, New York
- [27] Kalman RE (1960) A new approach to linear filtering and prediction problems, *Transactions of the ASME—Journal of Basic Engineering* 82 (Series D): 35-45
- [28] Kalman RE, Bucy RS (1961) New results in linear filtering and prediction theory. Retrieved 2008-05-03
- [29] LeDimet FX, Talagrand O (1986) Variational algorithms for analysis and assimilation of meteorological observations—theoretical aspects. *Tellus Series A—Dynamic Meteorology and Oceanography* 38(2): 97-110
- [30] Liu GS, Chen Y, Zhang DX (2008) Investigation of flow and transport processes at the MADE site using ensemble Kalman filter. *Advances in Water Resources* 31: 975-986
- [31] Liu Y, Gupta HV (2007) Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resour. Res.* 43, W07401, doi: 10.1029/2006WR005756
- [32] Maybeck PS (1979) *Stochastic models, estimation and control*. Academic Press, INC., London, LTD
- [33] Mclaughlin D, Townley LR (1996) A reassessment of the groundwater inverse problem. *Water Resour. Res.* 32(5): 1131-1161
- [34] Mclaughlin D (2002) An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering. *Adv Water Resour.* 25: 1275-1286
- [35] Natvik L-J, Evensen G (2003) Assimilation of ocean color data into a biochemical model of the North Atlantic Par 1. Data assimilation experiments. *Journal of Marine System* 40-41: 127-153
- [36] Oliver DS, Reynolds AC, Liu N (2008) Inverse theory for petroleum reservoir characterization and history matching. July 28
- [37] Poeter EP, Hill MC (1997) Inverse models: a necessary next step in ground-water modeling. *Ground water* 35(2): 250-260
- [38] Prakash J, Patwardhan SC, Shah SL (2008) Constrained state estimation using the ensemble Kalman filter. 2008 American Control Conference, Westin Seattle Hotel, Seattle, Washington, USA, June 11-13

- [39] Reichle RH (2008) Data assimilation methods in the Earth sciences. *Advances in Water Resources* doi:10.1016/j.advwatres.2008.01.001
- [40] Sridhar U, Dolence E, Li KY (2007) Constrained extended Kalman filter for nonlinear state estimation. In *Proceedings of the 8th International IFAC Symposium on Dynamics and Control of Process Systems*
- [41] Sun NZ (1994) *Inverse problems in groundwater modeling*. Kluwer Academic Publisher. Dordrecht, Netherlands
- [42] Sun NZ, Yeh W-G (1992) A stochastic inverse solution for transient groundwater flow: Parameter identification and reliability analysis. *Water Resour. Res.* 28(12): 3269-3280
- [43] Thacker WC (2007) Data assimilation with inequality constraints. *Ocean Modelling* 16(3-4): 264-276
- [44] Thomsen PG, Zlatev Z (2008) Development of a data assimilation algorithm. *Computers and Mathematics with Applications* 55: 2381-2393
- [45] Tipireddy R, Nasrellah HA, Manohar CS (2008) A Kalman filter based strategy for linear structural system identification based on multiple static and dynamic test data. *Probabilistic Engineering Mechanics* doi:10.1016/j.probenmech.2008.01.001
- [46] 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
- [47] Vrugt JA, Robinson BA, Vesselinov VV (2005a) Improved inverse modeling for flow and transport in subsurface media: Combined parameter and state estimation. *Geophys. Res. Lett.* 32: L18408, doi:10.1029/2005GL023940
- [48] Vrugt JA, Diks CGH, Gupta HV, Bouten W, Verstraten JM (2005b) 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
- [49] Vermeulen PTM, Stroet CBM, Heemink AW (2006) Model inversion of transient nonlinear groundwater flow models using model reduction. *Water Resour. Res.* 42: W09417, doi: 10.1029/2005WR004536
- [50] Wang DB, Chen YG, Cai XM (2009) State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. *Water Resources Research* 45, W11416, doi:10.1029/2008WR007401
- [51] Yeh T-C J, Liu S (2000) Hydraulic tomography: Development of a new aquifer test method. *Water Resour. Res.* 36(8): 2095-2105
- [52] Yeh T-C J, Zhang J (1996) A geostatistical inverse method for variably saturated flow in the vadose zone. *Water Resour. Res.* 32(9): 2757-2766

- [53] Zhang D, Lu Z, Chen Y (2007) Dynamic reservoir data assimilation with an efficient, dimension-reduced Kalman filter. *SPE Journal* 12(1): 108-117
- [54] Zheng C (1990) A Modular Three-dimensional transport model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems. S.S. Papadopoulos & Associates, Inc. Rockville, Maryland 20852
- [55] Zhu J, Yeh T-C J (2005) Characterization of aquifer heterogeneity using transient hydraulic tomography. *Water Resour. Res.* 41: W07028, doi:10.1029/2004WR003790
- [56] Zou YX, Testroet CBM, van Geer FC (1991) Using Kalman filtering to improve and quantifying the uncertainty of numerical groundwater simulation: 2. Application to monitoring network design. *Water Resour. Res.* 27(8): 1995-2006