

# Reservoir history matching and inversion using an iterative ensemble Kalman filter with covariance localization

Wang Yudou\* and Li Maohui

School of Physics Science & Technology, China University of Petroleum, Dongying, Shandong 257061, China

© China University of Petroleum (Beijing) and Springer-Verlag Berlin Heidelberg 2011

**Abstract:** Reservoir inversion by production history matching is an important way to decrease the uncertainty of the reservoir description. Ensemble Kalman filter (EnKF) is a new data assimilation method. There are two problems have to be solved for the standard EnKF. One is the inconsistency between the updated model and the updated dynamical variables for nonlinear problems, another is the filter divergence caused by the small ensemble size. We improved the EnKF to overcome these two problems. We use the half iterative EnKF (HIEnKF) for reservoir inversion by doing history matching. During the HIEnKF process, the prediction data are obtained by rerunning the reservoir simulator using the updated model. This can guarantee that the updated dynamical variables are consistent with the updated model. The updated model can nonlinearly affect the prediction data. It is proved that HIEnKF is similar to the first iteration of the EnRML method. Covariance localization is introduced to alleviate filter divergence and spurious correlations caused by the small ensemble size. By defining the shape and size of the correlation area, spurious correlation between the gridblocks far apart is alleviated. More freedom of the model ensemble is preserved. The results of history matching and inverse problem obtained from the HIEnKF with covariance localization are improved. The results show that the model freedom increases with a decrease in the correlation length. Therefore the production data can be matched better. But too small a correlation length can lose some reservoir information and this would cause big errors in the reservoir model estimation.

**Key words:** Half iterative ensemble Kalman filter, covariance localization, reservoir inversion, history matching, fluvial channel reservoir

## 1 Introduction

Most of the traditional automatic reservoir history matching methods are gradient methods. We have to use an adjoint method to calculate the sensitivity matrix which is the gradient of the objective function with respect to the model parameters. The adjoint method is difficult to apply to real reservoir history matching because it is very complicated and computational expensive (Li et al, 2003). The ensemble Kalman filter (EnKF) is a developing data assimilation method which was firstly introduced by Evensen in the context of ocean dynamics literature (Evensen, 1994). Since it was introduced into the petroleum engineering literature (Naevdal et al, 2003), EnKF has attracted great research interest and has now been successfully applied to automatic history matching of some practical reservoirs (Haugen et al, 2006; Evensen et al, 2007; Bianco et al, 2007; Seiler et al,

2009). Chinese scholars have started to apply the method to the inverse problem of the earth surface parameters and the estimation of rock mechanics parameters (Qin et al, 2005; Han and Li, 2008; Zhao et al, 2007; Jia et al, 2009). However, this method has not been used in the petroleum geology and reservoir engineering literature in China (Li et al, 2009). Although the standard EnKF has been used successfully in reservoir history matching and reservoir inversion, there are some urgent problems to be solved (Aanonsen et al, 2009). The standard EnKF is a time sequential history matching and reservoir inversion method. It requires that at each data assimilation time the estimated reservoir dynamic variables (pressure, saturation, dissolved gas-oil ratio, etc.) are consistent with the estimated reservoir model, that is, if we input the estimated reservoir model into the reservoir simulator, the predicted dynamic fields are consistent with the estimated dynamic fields. It has been proved that for linear problems this hypothesis is correct (Thulin et al, 2007). However, the results also show that for nonlinear problems consistency is not achieved. Most of reservoir flow problems are nonlinear, especially for the complex reservoirs (fluvial

\*Corresponding author. email: wangyd@upc.edu.cn

Received December 1, 2010

channel reservoirs, three-phase flow reservoir, etc.). In order to solve this problem, an iterative ensemble Kalman filter (IEnKF) which is based on a gradient method and has to use an adjoint method to calculate the sensitivity matrix is proposed (Li and Reynolds, 2009). Gu and Oliver (2007) also proposed an ensemble randomized maximum likelihood (EnRML) which used an average sensitivity matrix instead of the real matrix. Because the singular value decomposition (SVD) is used to solve the average sensitivity matrix, this method is more efficient (Oliver et al, 2008). However, all realizations of the ensemble have to rerun the simulator from time zero after each iteration, so the computation is still large. And because of the utilization of the average sensitivity matrix, we can not guarantee the objective function always decrease along the search direction in the subsequent iterations. The result is not improved when more iterations are done. Another iterative EnKF (HIEnKF) was used to estimate the depths of the fluid contacts and relative permeability curves (Wang et al, 2010). All realizations have to rerun the simulator from time zero using the updated model at each the data assimilation time. The result obtained by HIEnKF is similar to that obtained by EnRML. But the computational expense of HIEnKF is much less than that of EnRML.

EnKF is one type of Monte Carlo method. It uses an ensemble that contains  $N_e$  realizations to calculate the statistical parameters, such as the mean of the model and the covariance matrix (Evensen, 2007). Since the covariance matrix is calculated from the ensemble of  $N_e$  realizations, there are  $N_e$  degrees of freedom to estimate the actual model (Li and Reynolds, 2009). If the number of the realizations  $N_e$  is not very large and the error of the measurement data is small, the rank of the posteriori model covariance matrix will reduce when more and more data are assimilated. In the most extreme situation, the rank of the updated covariance matrix would be zero when all realization are same so that all predictions would be equal and we would not be able to change the predicted states to match the data. This situation is one type of the filter divergence that occurs commonly in practice (Agbalaka and Oliver, 2008). Furthermore, when a small ensemble size is used to represent covariance, there are significant sampling errors which tend to produce nonzero spurious long-distance correlation between parameters or between parameters and dynamical variables, whereas in reality, the variables at gridblocks far apart should be uncorrelated. Spurious correlations can result in changes in model parameters where none should occur because the predicted data (corresponding to the data to be assimilated) are insensitive to the model parameters. Not only does this cause changes during data assimilation in model parameters when no change should occur, but it can make it difficult to properly change these parameters to assimilate later time data which are actually sensitive to these parameters, thus, contributing to filter divergence (Agbalaka and Oliver, 2008). To improve the EnKF performance on history matching, Houtekamer proposed a standard EnKF method based on the covariance matrix localization, and made a preliminary application in atmospheric data assimilation (Houtekamer and Mitchell, 2001). Anderson proposed a covariance matrix inflation method. After data assimilation, magnification of

the covariance matrix can effectively restrain the filtering divergence (Anderson, 2007). But the method can not eliminate spurious correlation due to the smaller set of models. Agbalaka used covariance localization theory to localize the Kalman gain matrix and also improved the elimination of filtering divergence (Agbalaka and Oliver, 2008). But the method must store the Kalman gain matrix which is very large.

A fluvial channel reservoir is highly heterogeneous and complex. The fluid flow in this type of reservoirs is nonlinear and filtering divergence is quite serious. Inversion of the reservoir permeability field by using the standard EnKF is difficult (Li et al, 2009; Sarma et al, 2008; Jafarpour and McLaughlin, 2009). In this paper, we use the HIEnKF method to do reservoir history matching which can avoid the complex calculation of sensitivity coefficients and ensure the estimated dynamical fields are consistent with the estimated models. We proved that the HIEnKF is similar to the first iteration of the EnRML method. Covariance localization is used to eliminate filtering divergence and spurious correlation due to the small set of models. The HIEnKF with the covariance localization is used to estimate a fluvial channel reservoir. Better results of the history matching and estimation of the permeability field are obtained.

## 2 Ensemble Kalman filter method

We define a state vector

$$\mathbf{y}^n = \begin{bmatrix} \mathbf{m} \\ \mathbf{p}^n \\ \mathbf{d}^n \end{bmatrix} \tag{1}$$

where  $\mathbf{m}$  is a  $N_m$ -dimensional column vector which represents reservoir static model parameters, including discrete field of porosity and permeability;  $\mathbf{p}^n$  is a  $N_p$ -dimensional column vector which represents reservoir dynamic parameters (i.e., pressure, saturation, dissolved gas-oil ratio (GOR), etc.) at time  $t_n$ ; and  $\mathbf{d}^n$  is a  $N_d$ -dimensional column vector which represents the predicted data at time  $t_n$ . Therefore, the dimension of  $\mathbf{y}^n$ , denoted by  $N_y$ , is given by  $N_m + N_p + N_d$ . We define matrix

$$\mathbf{H} = [\mathbf{O} \quad \mathbf{I}_{N_d}] \tag{2}$$

where  $\mathbf{O}$  denotes the  $N_d \times (N_m + N_p)$  null matrix;  $\mathbf{I}_{N_d}$  denotes the  $N_d \times N_d$  identity matrix. Thus the predicted data is linearly related to the state vector

$$\mathbf{d}^n = \mathbf{H} \mathbf{y}^n \tag{3}$$

Actually the predicted data at time  $t_n$  satisfies the following equation

$$\mathbf{d}^n = \mathbf{g}^n(\mathbf{m}, \mathbf{p}^n) \tag{4}$$

where  $\mathbf{g}^n$  is the reservoir numerical model which shows that the production at time  $t_n$  is only related to the geologic model

and the dynamic field at this time. The observed data can be treated as a random vector

$$d_{obs}^n = d^n + \varepsilon_d^n = g^n(m, p^n) + \varepsilon_d^n \tag{5}$$

where  $\varepsilon_d^n$  is the measurement error. It is assumed that the measurement error has a Gaussian distribution with zero mean and covariance matrix given by  $C_{D_n}$  and there is no correlation between the measurement errors at different time.

The purpose of the EnKF is to generate a series of reservoir models that can match the production data by sampling the following conditional probability density function

$$f(y^n | d_{obs}^n \dots d_{obs}^1) = a \exp \left[ -\frac{1}{2} (y^n - \bar{y}^n)^T C_{Y_n}^{-1} (y^n - \bar{y}^n) - \frac{1}{2} (H y^n - d_{obs}^n)^T C_{D_n}^{-1} (H y^n - d_{obs}^n) \right] \tag{6}$$

By minimizing the following objective function we can obtain a sample of probability density function defined by Eq. (6)

$$O(y_j^n) = \frac{1}{2} (y_j^n - y_j^{n,p})^T C_{Y_n}^{-1} (y_j^n - y_j^{n,p}) + \frac{1}{2} (H y_j^n - d_{uc,j}^n)^T C_{D_n}^{-1} (H y_j^n - d_{uc,j}^n) \quad j = 1, \dots, N_e \tag{7}$$

where  $y_j^{n,p}$  represents the  $j$ th state vector at time  $t_n$ , which is obtained by running the reservoir simulator from time  $t_{n-1}$  to time  $t_n$ .

$$d_{uc,j}^n = d_{obs}^n + L_D Z_D \tag{8}$$

where  $C_{D_n} = L_D L_D^T$ . The  $N_d$ -dimensional random vector  $Z_D$  represents the normal random vector with zero mean and covariance given by the  $N_d \times N_d$  identity matrix  $I_{N_d}$ . Minimize  $O(y_j^n)$  defined by Eq. (7), we get standard EnKF analysis equation (Zafari et al, 2005)

$$y_j^{n,u} = y_j^{n,p} + C_{Y_n} H^T (C_{D_n} + H C_{Y_n} H^T)^{-1} (d_{uc,j}^n - d_j^{n,p}) \quad j = 1, \dots, N_e \tag{9}$$

The covariance matrix of state vector can be divided into

$$C_{Y_n} = \begin{bmatrix} C_{M_n} & C_{M_n P_n} & C_{M_n D_n} \\ C_{P_n M_n} & C_{P_n} & C_{P_n D_n} \\ C_{D_n M_n} & C_{D_n P_n} & C_{D_n D_n} \end{bmatrix} \tag{10}$$

where each sub-matrix represents a covariance between the vectors denoted by the subscript. In order to distinguish the covariance of the predicted data from the covariance of the data measurement error given by  $C_{D_n}$ , it is denoted by  $C_{D_n D_n}$ . Using Eqs. (2) and (10), we obtain

$$C_{Y_n} H^T = \begin{bmatrix} C_{M_n} & C_{M_n P_n} & C_{M_n D_n} \\ C_{P_n M_n} & C_{P_n} & C_{P_n D_n} \\ C_{D_n M_n} & C_{D_n P_n} & C_{D_n D_n} \end{bmatrix} \begin{bmatrix} O^T \\ I_{N_d}^T \end{bmatrix} = \begin{bmatrix} C_{M_n D_n} \\ C_{P_n D_n} \\ C_{D_n D_n} \end{bmatrix} \tag{11}$$

and

$$H C_{Y_n} H^T = \begin{bmatrix} O & I_{N_d} \end{bmatrix} \begin{bmatrix} C_{M_n D_n} \\ C_{P_n D_n} \\ C_{D_n D_n} \end{bmatrix} = C_{D_n D_n} \tag{12}$$

Substituting Eqs. (3), (11) and (12) into Eq. (9), we obtain

$$y_j^{n,u} = y_j^{n,p} + \begin{bmatrix} C_{M_n D_n} \\ C_{P_n D_n} \\ C_{D_n D_n} \end{bmatrix} (C_{D_n} + C_{D_n D_n})^{-1} (d_{uc,j}^n - d_j^{n,p}) \tag{13}$$

Eq. (13) can be divided into three updated analysis equations for a static field, dynamic field, and production data, respectively:

$$m_j^{n,u} = m_j^{n,p} + C_{M_n D_n} (C_{D_n} + C_{D_n D_n})^{-1} (d_{uc,j}^n - d_j^{n,p}) \tag{14}$$

$$p_j^{n,u} = p_j^{n,p} + C_{P_n D_n} (C_{D_n} + C_{D_n D_n})^{-1} (d_{uc,j}^n - d_j^{n,p}) \tag{15}$$

$$d_j^{n,u} = d_j^{n,p} + C_{D_n D_n} (C_{D_n} + C_{D_n D_n})^{-1} (d_{uc,j}^n - d_j^{n,p}) \tag{16}$$

In EnKF, we generally do not use Eq. (16) to update data. Corresponding covariance can be calculated by the following equations

$$C_{M_n D_n} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (m_j^{n,p} - \bar{m}^{n,p})(d_j^{n,p} - \bar{d}^{n,p})^T = \frac{1}{N_e - 1} (M^{n,p} - \bar{M}^{n,p})(D^{n,p} - \bar{D}^{n,p})^T \tag{17}$$

$$C_{P_n D_n} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (p_j^{n,p} - \bar{p}^{n,p})(d_j^{n,p} - \bar{d}^{n,p})^T = \frac{1}{N_e - 1} (P^{n,p} - \bar{P}^{n,p})(D^{n,p} - \bar{D}^{n,p})^T \tag{18}$$

$$C_{D_n D_n} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (d_j^{n,p} - \bar{d}^{n,p})(d_j^{n,p} - \bar{d}^{n,p})^T = \frac{1}{N_e - 1} (D^{n,p} - \bar{D}^{n,p})(D^{n,p} - \bar{D}^{n,p})^T \tag{19}$$

where the  $j$ th column of  $M^{n,p}$ ,  $P^{n,p}$ , and  $D^{n,p}$  represents the corresponding values of the  $j$ th model, respectively. And each column of  $\bar{M}^{n,p}$ ,  $\bar{P}^{n,p}$ , and  $\bar{D}^{n,p}$  is the mean of itself.

Using EnKF equations to assimilate the production data, we obtain the updated state vectors including the updated reservoir models (permeability field, porosity field, the depth of the initial fluid contacts, relative permeability curve, etc.) and the updated dynamic fields (pressure field, saturation field, dissolved GOR field, etc.). These updated fields are used to run the reservoir simulator to the next assimilation step to do the next data match. Therefore, the EnKF is a time-sequential data assimilation method, each ensemble model needs to run the simulator only once. Compared to the gradient-based history matching method, EnKF is more efficient.

### 3 Half iterative ensemble Kalman filter

As a time-sequential data assimilation algorithm, EnKF requires that the updated model should be consistent with the updated dynamic field at every data assimilation time, that is, if we input the updated model into the reservoir simulator and run it to the present time, the predicted dynamic field should be consistent with the updated dynamic field. However, for nonlinear problems consistency is not achieved. In order to overcome the inconsistency in the nonlinear problems, we use the half iterative EnKF (HIEnKF) to do the data assimilation (Wang et al, 2010). The state vector only includes model parameters, thus the HIEnKF analysis equation is

$$m_j^{n,u} = m_j^{n,p} + C_{M_n D_n} (C_{D_n} + C_{D_n D_n})^{-1} (d_{uc,j}^n - d_j^{n,p}) \quad j=1, \dots, N_e \quad (20)$$

where  $d_j^{n,p}$  is the predicted data by running the simulator from time zero to time  $t_n$  using model  $m_j^{n,p}$ . Because the effect of all updated parameters on the production data can be reflected by rerunning the simulator from time zero, the reservoir model can be updated reasonably.  $d_j^{n+1,p}$  at time  $t_{n+1}$  can be obtained by running the simulator from time zero to time  $t_{n+1}$  using the updated model. Just one simulator run from time zero at each assimilation time step is required, so the HIEnKF method is more efficient.

The HIEnKF method is similar to EnRML (Gu and Oliver, 2007). If we set iterative step length equal to one and use the prior model as the initial model of the iteration, the first iterative equation of EnRML can be written as

$$m_j^{n,u} = m_j^{n,p} + C_{M_n} \bar{G}^{nT} (C_{D_n} + \bar{G}^n C_{M_n} \bar{G}^{nT})^{-1} (d_{uc,j}^n - d_j^{n,p}) \quad j=1, \dots, N_e \quad (21)$$

where  $\bar{G}^n$  is the approximation of the average sensitivity matrix at time  $t_n$ , and can be calculated by

$$\Delta D^{n,p} = \bar{G}^n \Delta M^{n,p} \quad (22)$$

The  $j$ th column of  $\Delta M^{n,p}$  represents the deviation between the  $j$ th model and the mean of all the models in the ensemble. The dimension of  $\Delta M^{n,p}$  is  $N_m \times N_e$ . The  $j$ th column of  $\Delta D^{n,p}$  represents the deviation between predicted data of the  $j$ th

model and the average predicted data of all the models in the ensemble. The dimension of  $\Delta D^{n,p}$  is  $N_d \times N_e$ . In general,  $\Delta M^n$  is irreversible. We can do a singular value decomposition to it

$$\Delta M^{n,p} = U \Lambda V^T \quad (23)$$

where  $U$  and  $V$  are both orthogonal matrices.

$$U U^T = U^T U = I_{N_m} \quad (24)$$

$$V V^T = V^T V = I_{N_e} \quad (25)$$

$\Lambda$  is a  $N_m \times N_e$  matrix. If  $N_m > N_e$ ,  $\Lambda$  can be written as

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_{N_e} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} A_{N_e} \\ \mathbf{0} \end{bmatrix} \quad (26)$$

$\lambda_i$  is the  $i$ th non-zero singular value of  $\Delta M^{n,p}$ .  $A_{N_e}$  is  $N_e \times N_e$  diagonal matrix composed by  $\lambda_i$ . Therefore

$$\Delta M^{n,p} = U_{N_e} A_{N_e} V_{N_e}^T \quad (27)$$

$U_{N_e}$  is a  $N_m \times N_e$  matrix. Its  $j$ th column is the left singular vector corresponding to the  $j$ th non-zero singular value  $\lambda_i$ .  $V_{N_e}$  is a  $N_e \times N_e$  matrix. Its  $j$ th column is the right singular vector corresponding to the  $j$ th non-zero singular value  $\lambda_i$ . The average sensitivity matrix can be calculated by

$$\bar{G}^n = \Delta D^{n,p} V_{N_e} A_{N_e}^{-1} U_{N_e}^T \quad (28)$$

From the calculation of  $C_{M_n}$

$$C_{M_n} = \frac{1}{N_e - 1} \Delta M^{n,p} \Delta M^{n,pT} \quad (29)$$

we can obtain

$$\begin{aligned} C_{M_n} \bar{G}^{nT} &= \frac{1}{N_e - 1} \Delta M^{n,p} \Delta M^{n,pT} \bar{G}^{nT} \\ &= \frac{1}{N_e - 1} \Delta M^{n,p} V_{N_e} A_{N_e} U_{N_e}^T U_{N_e} A_{N_e}^{-1} V_{N_e}^T \Delta D^{n,pT} \\ &= \frac{1}{N_e - 1} \Delta M^{n,p} V_{N_e} A_{N_e} A_{N_e}^{-1} V_{N_e}^T \Delta D^{n,pT} \\ &= \frac{1}{N_e - 1} \Delta M^{n,p} V_{N_e} V_{N_e}^T \Delta D^{n,pT} \\ &= \frac{1}{N_e - 1} \Delta M^{n,p} \Delta D^{n,pT} = C_{M_n D_n} \end{aligned} \quad (30)$$

Analogously, we can obtain

$$\bar{G}^n C_{M_n} \bar{G}^{nT} = C_{D_n D_n} \quad (31)$$

Substituting Eqs. (30) and (31) into Eq. (21), the first iteration of the EnRML analysis equation is same to the HIEnKF equation (Eq. (20)). However, due to the application of the average sensitivity matrix, EnRML can not guarantee the search direction is always downhill. In the real case, the better result could not be obtained by doing more iteration (Wang et al, 2010). Compared to EnRML method, the computation of the HIEnKF is more efficient.

### 4 EnKF with covariance localization

In general, the parameters at two gridblocks far apart should be uncorrelated. When small ensemble size is used to represent the covariance, there are significant sampling errors which tend to produce nonzero spurious long-distance correlation between parameters or between parameters and dynamical variables. Spurious correlations can result in changes in model parameters where none should occur because the predicted data (corresponding to the data to be assimilated) are insensitive to the model parameters. It would contribute to filter divergence. Covariance localization can both eliminate long-distance spurious correlation and expand the degree of freedom available to assimilate data by setting the correlation between data measured at a particular location and model parameters or data corresponding to a location far away from the measurement to zero. We use the Schur product to filter the approximation of the forecast covariance at large distances and to increase the effective rank of the ensemble. The Schur product is an element-wise multiplication of two matrices  $A$  and  $B$  resulting in a matrix  $C$  of the same dimensions, defined as

$$C = A \circ B \tag{32}$$

where  $\circ$  represents Schur product, and

$$C_{i,j} = A_{i,j} B_{i,j} \tag{33}$$

Subscripts  $i$  and  $j$  imply the entry of  $i$ th row and  $j$ th column of the corresponding matrix. We can define the correlation coefficient matrix  $\rho$  as

$$\rho = \begin{bmatrix} \rho_M & \rho_{MP} & \rho_{MD} \\ \rho_{PM} & \rho_P & \rho_{PD} \\ \rho_{DM} & \rho_{DP} & \rho_{DD} \end{bmatrix} \tag{34}$$

Each sub-matrix represents the correlation coefficient matrix of the variables corresponding to the subscript. The value of the elements of the matrix is computed from a distance-dependent correlation function that varies from one at the observation location to zero at some pre-specified radial distance. The EnKF analysis equation with covariance localization can be written as

$$y_j^{n,u} = y_j^{n,p} + (\rho \circ C_{Y_n}) H^T [C_{D_n} + H(\rho \circ C_{Y_n}) H^T]^{-1} (d_{uc,j}^n - d_j^{n,p}) \tag{35}$$

$j = 1, \dots, N_e$

Using Eqs. (2), (10), and (34), we have

$$(\rho \circ C_{Y_n}) H^T = \begin{bmatrix} \rho_M \circ C_{M_n} & \rho_{MP} \circ C_{M_n P_n} & \rho_{MD} \circ C_{M_n D_n} \\ \rho_{PM} \circ C_{P_n M_n} & \rho_P \circ C_{P_n} & \rho_{PD} \circ C_{P_n D_n} \\ \rho_{DM} \circ C_{D_n M_n} & \rho_{DP} \circ C_{D_n P_n} & \rho_{DD} \circ C_{D_n D_n} \end{bmatrix} \begin{bmatrix} O \\ I_{N_d} \end{bmatrix} \tag{36}$$

$$= \begin{bmatrix} \rho_{MD} \circ C_{M_n D_n} \\ \rho_{PD} \circ C_{P_n D_n} \\ \rho_{DD} \circ C_{D_n D_n} \end{bmatrix}$$

Similarly,

$$H(\rho \circ C_{Y_n}) H^T = \begin{bmatrix} O & I_{N_d} \end{bmatrix} \begin{bmatrix} \rho_{MD} \circ C_{M_n D_n} \\ \rho_{PD} \circ C_{P_n D_n} \\ \rho_{DD} \circ C_{D_n D_n} \end{bmatrix} \tag{37}$$

$$= \rho_{DD} \circ C_{D_n D_n}$$

Substituting Eqs. (36) and (37) into Eq. (35), we have

$$y_j^{n,u} = y_j^{n,p} + \begin{bmatrix} \rho_{MD} \circ C_{M_n D_n} \\ \rho_{PD} \circ C_{P_n D_n} \\ \rho_{DD} \circ C_{D_n D_n} \end{bmatrix} [C_{D_n} + \rho_{DD} \circ C_{D_n D_n}]^{-1} (d_{uc,j}^n - d_j^{n,p}) \tag{38}$$

$j = 1, \dots, N_e$

By doing covariance localization, the effects of the data on the state vector are localized.

Note that, if we use the standard EnKF with covariance localization, we need to update both the model parameters and dynamical variables to keep the updated model consistent with the updated dynamic field. If the HIEnKF with covariance localization is used, we only need to update the model parameters using Eq. (38). The dynamical fields are obtained by running the reservoir simulator from time zero to the present time using the updated model.

The element of the correlation coefficient matrix represents the correlation coefficient between two grids with a certain distance and can be calculated by (Gaspari and Cohn, 1999)

$$\begin{cases} \phi = -\frac{1}{4} \left(\frac{\delta}{L}\right)^5 + \frac{1}{2} \left(\frac{\delta}{L}\right)^4 + \frac{5}{8} \left(\frac{\delta}{L}\right)^3 - \frac{5}{3} \left(\frac{\delta}{L}\right)^2 + 1 & 0 \leq \delta \leq L \\ \phi = \frac{1}{12} \left(\frac{\delta}{L}\right)^5 - \frac{1}{2} \left(\frac{\delta}{L}\right)^4 + \frac{5}{8} \left(\frac{\delta}{L}\right)^3 + \frac{5}{3} \left(\frac{\delta}{L}\right)^2 - 5 \frac{\delta}{L} \\ \quad + 4 - \frac{2}{3} \left(\frac{\delta}{L}\right)^{-1} & L \leq \delta \leq 2L \\ \phi = 0 & \delta > 2L \end{cases} \tag{39}$$

where  $\delta$  is the distance between the gridblock and the



measurement point;  $L$  is the correlation length which represents the size of correlated area to the production data. For an anisotropic reservoir, the correlation length along each direction is different. Here we consider the shape of the correlation area is an ellipse. The area and orientation can be determined by the major correlation length, minor correlation length and the direction of the major correlation direction.

If we know the size and shape of the correlated area, the correlation length along each direction can be obtained from the location of the grids and measurement points. The correlation coefficient at different distance is shown as Fig. 1. When the correlation coefficient matrix affects the covariance matrix, the spurious correlation between the gridblocks far apart is alleviated.

### 5 Examples

We apply the standard EnKF, HIEnKF, and HIEnKF techniques with covariance localization to a fluvial reservoir model to history match the production. The reservoir model is discretized over a  $45 \times 45 \times 1$  grid block system. The size of each gridblock is  $10\text{m} \times 10\text{m} \times 10\text{m}$ . The real permeability field is shown as Fig. 2(a). The permeability of the sand is 5,000 mD and the background permeability is 500 mD. Four producers and four injectors are completed in this reservoir. 200 prior models were generated by using multi-point geostatistical software Snesim with the hard data (Strebelle, 2002). The average prior model is shown as Fig. 2(b) and the training image is shown as Fig. 2(c). The reservoir model was water flooded for 1,080 days. The bottom hole pressure of the injectors are measurement data of the injectors. The producer measurement data are the oil and water production. The permeability field is estimated by history matching these production data.

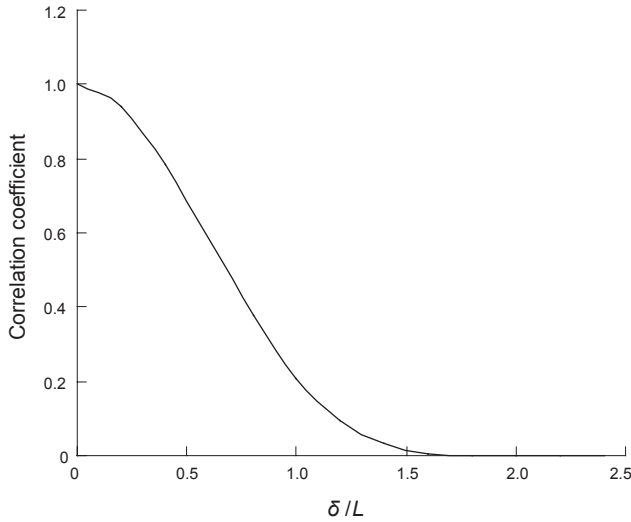


Fig. 1 Correlation coefficient

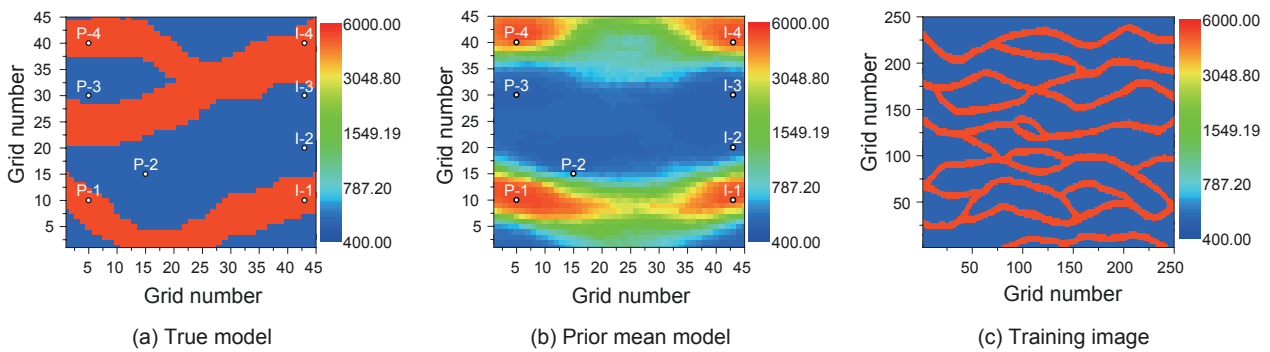


Fig. 2 Basic information of the reservoir

#### 5.1 Comparison of the standard EnKF and the HIEnKF

First we use the standard EnKF to estimate the permeability field by history matching the production data. The average of the posteriori models is shown in Fig. 3(a).

Compared with the real permeability field (shown in Fig. 2(a)), we find that the result produces the main structure of the reservoir. The border of the sand and the background is blurred, and the location of the channel is inconsistent with the truth. The permeability in the same facies is not

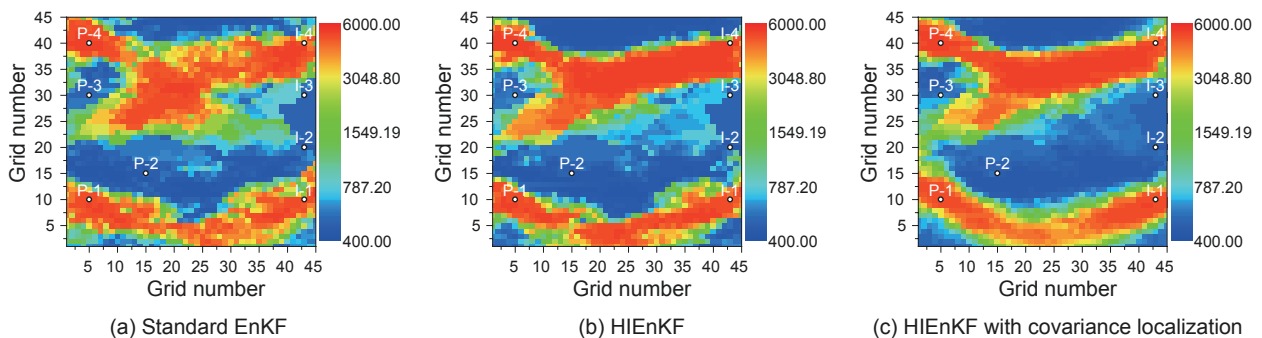
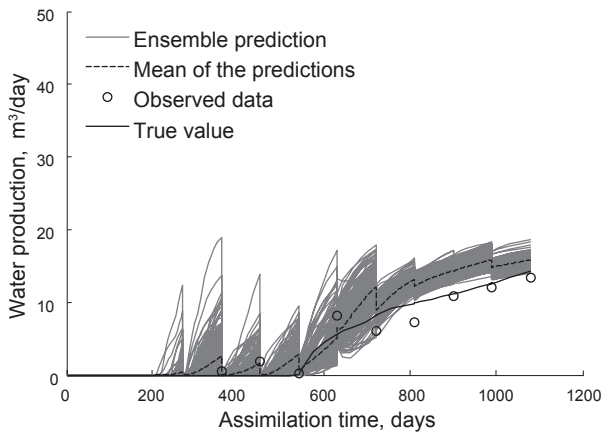


Fig. 3 The estimation of the permeability field

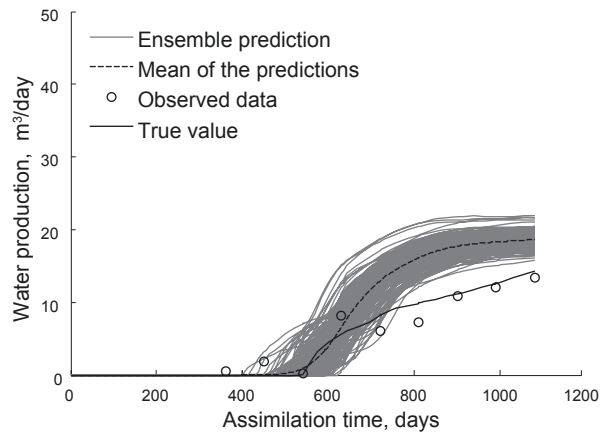
uniform. There are two reasons for the result. The first one is that the updated permeability field is inconsistent with the updated dynamic field due to the nonlinearity. The second is the large sampling errors cause a great spurious correlation between far apart grids. The HIEnKF method can guarantee the updated permeability field is consistent with the updated dynamic field. The estimated permeability field (as shown in Fig. 3(b)) is better than that obtained from the standard EnKF. The border is relatively clear. But the permeability in the same facies is still not uniform. The permeability in some gridblocks is too large.

Fig. 4(a) shows the water production prediction for well P-2 during the data assimilation procedure with the standard EnKF. During the early production stage the water production can be matched. But as more and more data were assimilated, the distribution becomes narrow which implies that the models in ensemble trend towards each other and the degree of freedom becomes small. It is more difficult to assimilate

more data. The data mismatch between the predicted data and the measurement data appeared during the later time. This finally leads to filter divergence. Fig. 4(b) shows the water production prediction of well P-2 from running the simulator from time zero using the final updated model obtained from the standard EnKF. Obviously, at the later time, the predicted data deviates heavily from the measurement data. The reason is the inconsistency between the updated dynamic field and the updated reservoir model caused by the nonlinearity of the fluids flow in the fluvial reservoir. Filter divergence also appears in the results of the HIEnKF (shown in Fig. 5(a)). However, since the HIEnKF overcomes the inconsistency between the updated dynamic field and the updated static field, the result from rerunning the simulator from time zero using the final updated model (shown in Fig. 5(b)) is similar to that obtained during the data assimilation procedure. Compared with the standard EnKF the result is a substantial improvement.

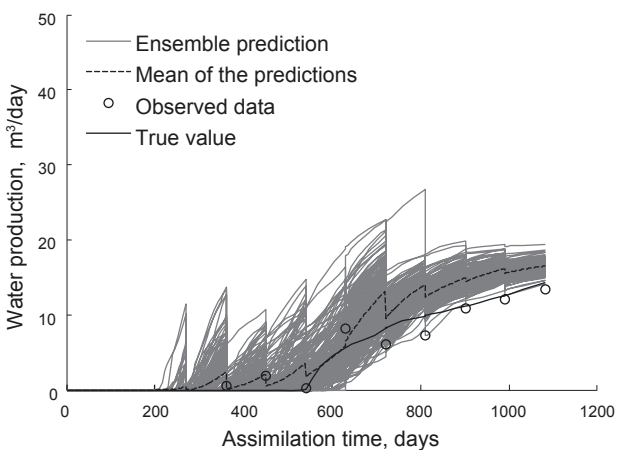


(a) During the data assimilation procedure

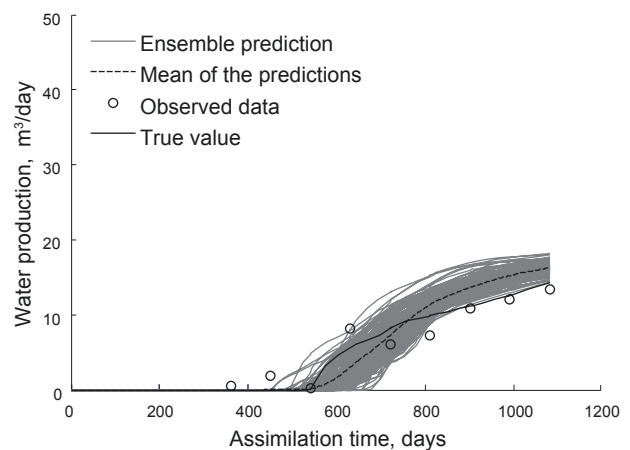


(b) Rerunning from time zero using the final estimated model

**Fig. 4** Water production prediction at well P-2 from the standard EnKF



(a) During the data assimilation procedure



(b) Rerunning from time zero using the final estimated model

**Fig. 5** Water production prediction at well P-2 from the HIEnKF

**5.2 Improvement of the HIEnKF with covariance localization**

We use the HIEnKF with covariance localization to match the production data. We set the major correlation length to 400 m along the *x*-axis. The minor correlation length is 200 m along the *y*-axis. Since the covariance localization reduces

the spurious correlation of gridblocks far apart, the estimated permeability field is improved (as shown in Fig. 3(c)). The location and border of the fluvial channel are closer to the real permeability distribution. The permeability in the same facies is uniformly distributed. However some information about the channel is lost. The possible reason is: 1) the production

data is not sensitive to the parameters at these gridblocks; or 2) the information is lost by covariance localization.

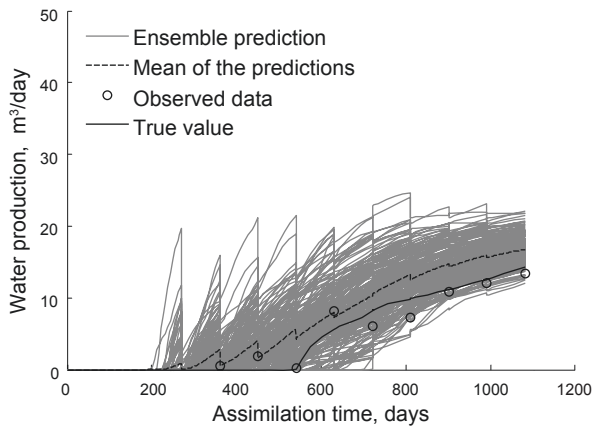
Fig. 6 shows the water production prediction at well P-2 obtained from the HIEnKF with covariance localization during the data assimilation procedure and the results obtained by rerunning from time zero using the final updated models. During the data assimilation procedure the distribution of the model predicted curves becomes wide and makes the measurement data within the band of predictions from the ensemble. The filter divergence is amended. The covariance localization makes sure that only the correlation area is adjusted by the measurement data and increases the degrees of freedom available to match data. Because the

HIEnKF overcomes the nonlinearity, the data match obtained by rerunning the simulator from time zero is similar to that obtained during the data assimilation procedure.

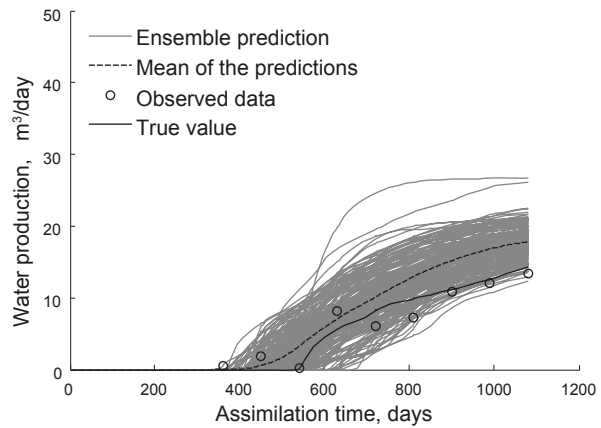
The distribution of the predictions of the ensemble represents the freedom of the ensemble to match data. The distribution of the water production predictions of the ensemble can be determined from the standard deviation of the water production predictions of the ensemble, defined by

$$\sigma_{q_w^n} = \sqrt{\frac{1}{N_e - 1} \sum_{j=1}^{N_e} (q_{w,j}^n - \bar{q}_w^n)^2} \tag{40}$$

where  $\bar{q}_w^n$  represents the mean of the water production predictions of the ensemble at the  $n$ th time.



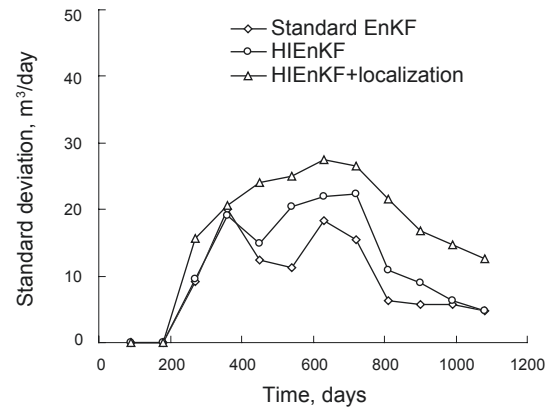
(a) During the data assimilation procedure



(b) Rerunning from time zero using the final estimated model

**Fig. 6** Water production prediction at well P-2 from the HIEnKF with covariance localization

Fig. 7 shows the standard deviation of predicted water production at well P-2 from the standard EnKF, HIEnKF, and the HIEnKF with covariance localization, respectively. We find that the standard deviations of the standard EnKF and HIEnKF are close. The standard deviations of the HIEnKF are larger than the standard EnKF only at the middle production time. These two methods have the same standard deviation when data assimilation is completed. However the standard deviation of the HIEnKF with covariance localization is substantially higher. The large standard deviation after data assimilation is completed implies that the HIEnKF with covariance localization can maintain large degrees of freedom to match additional future data.



**Fig. 7** The standard deviation of the water production predictions at well P-2

### 5.3 Effect of the ensemble size

Each individual posteriori realization of the model parameters is a linear combination of the associated prior ensemble of models for the EnKF. The sampling error is determined by the number of prior realizations. A big sampling error will lead to filter divergence. Fig. 8 shows the average estimation of the model using the standard EnKF when the ensemble size is 100 and 200, respectively. Larger sampling errors result in poorer estimation of the permeability field when the ensemble size is 100. The location and border of the channel are not clear. The distribution of the permeability is not uniform in one facies. The sand between wells P-2 and P-3 is lost. When the ensemble size increases to 200, the result is significantly improved. However,

compared with the real permeability field, the result is not good. Fig. 9 shows the water production prediction at well P-2 obtained by rerunning the simulator from time zero using the final updated models estimated from the standard EnKF. Although the band of the prediction curves of ensemble size 200 is wider than that of ensemble size 100, divergence appears in both cases. This implies that the estimations of the permeability field from the standard EnKF are worse whatever the ensemble size is. Fig. 10 shows the average estimation of the model using the HIEnKF with covariance localization. No matter if the model number is 100 or 200, the estimation of the permeability field is better than



that obtained from the standard EnKF. The permeability field estimation is very close to the true field when the ensemble size is 200. The result of 100 ensemble member obtained from the HIEnKF with covariance localization is better than that estimated from the standard EnKF with 200 ensemble member. This point can also be obtained from the water production prediction curves

(Fig. 11(a)). Although the band of the prediction curves of the ensemble size of 100 obtained from the HIEnKF with covariance localization is similar to that of ensemble size of 100 obtained from the standard EnKF, the data match is good and the filter divergence is alleviated. When ensemble member increases to 200, the result of the HIEnKF with covariance localization is

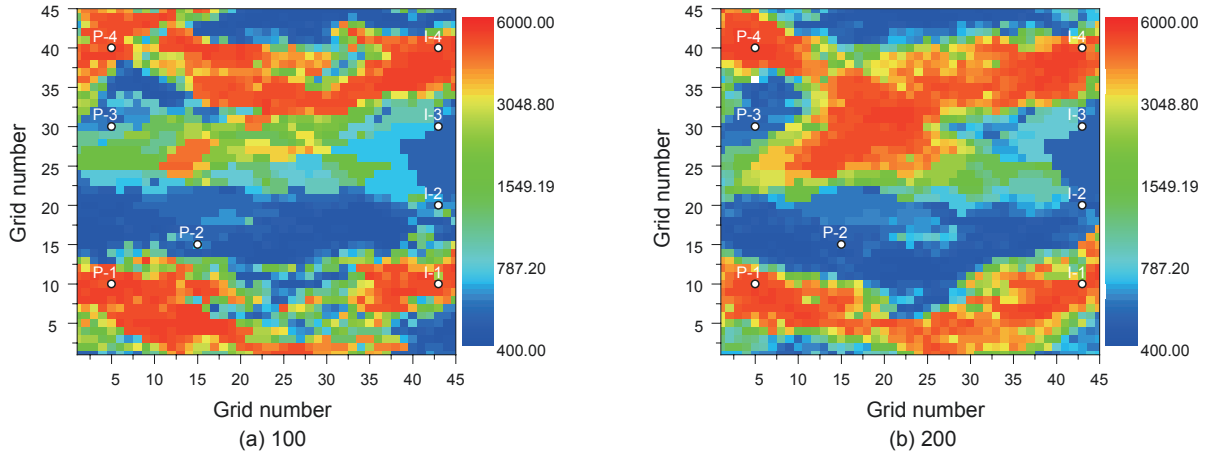


Fig. 8 The effect of the ensemble size on the estimation of the permeability (EnKF)

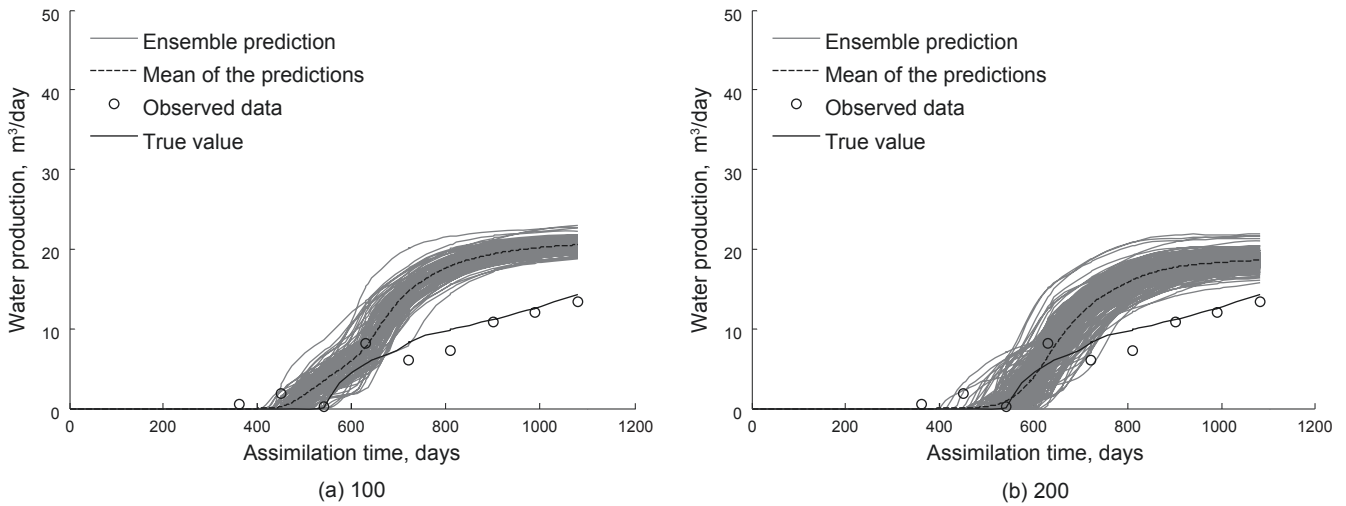


Fig. 9 The effect of the ensemble size on the production predictions (EnKF)

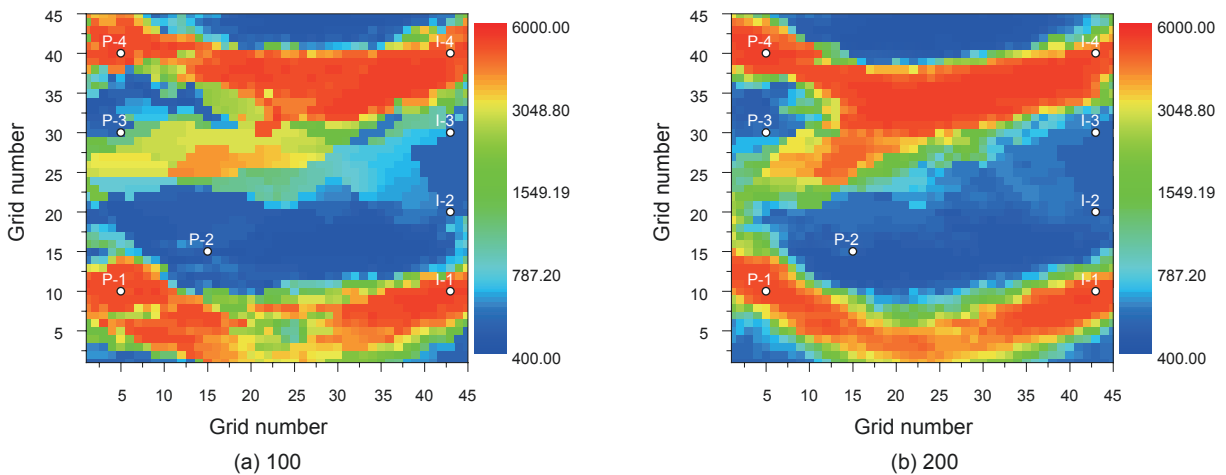
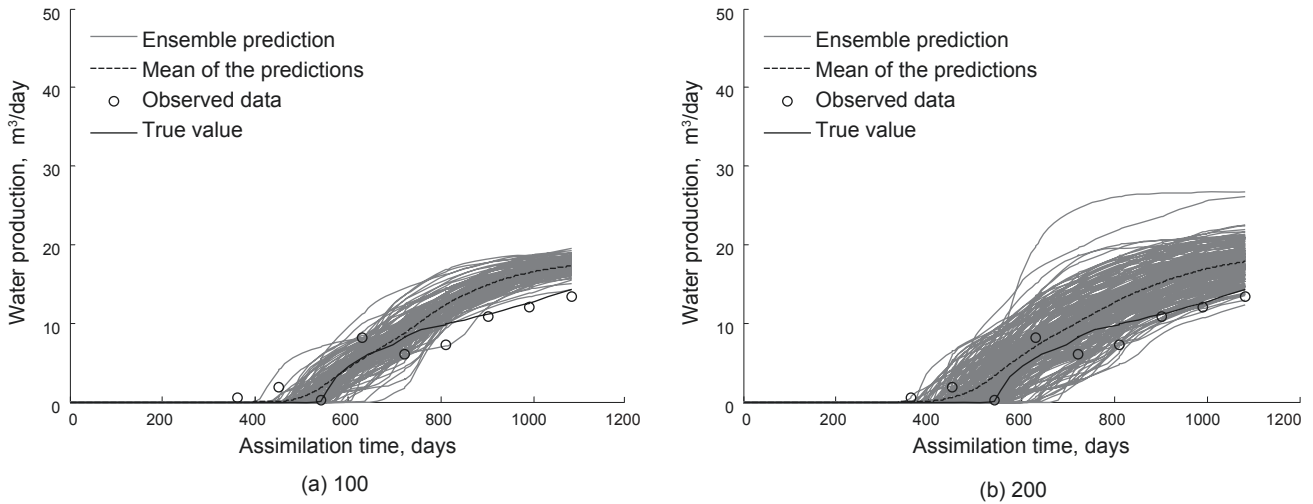


Fig. 10 The effect of the ensemble size on the estimation of the permeability (HIEnKF with covariance localization)



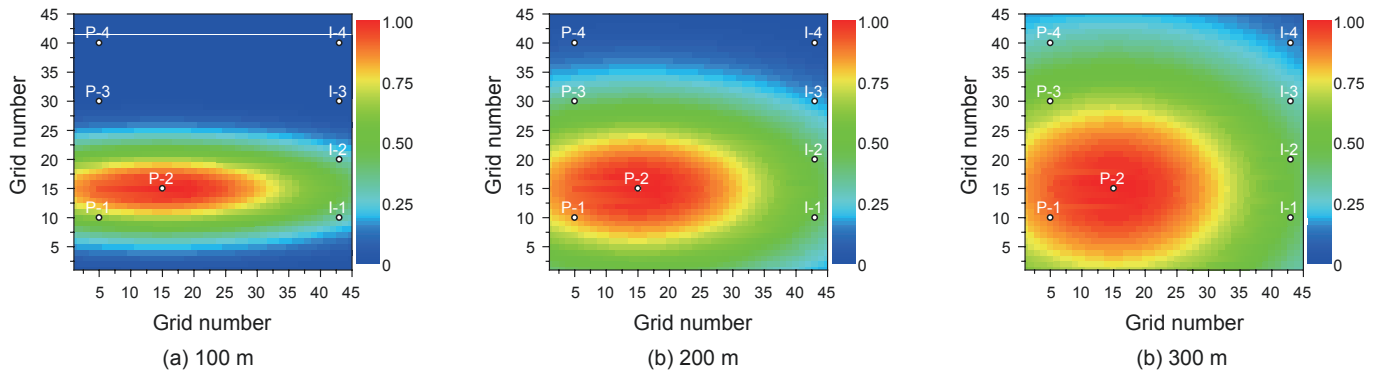
**Fig. 11** The effect of the ensemble size on the production predictions (HIEnKF with covariance localization)

greatly improved. Therefore, although the covariance localization can reduce the spurious correlation and sampling error due to a small ensemble size, the geological information of the reservoir cannot be fully reflected by the small ensemble size.

**5.4 Effect of the correlation length**

We use 200 ensemble members to discuss the effect of the correlation length on the inversion of the permeability field. The major correlation length is 400 m along the *x*-axis which is a constant when the inversion is done. The correlation coefficient between measurement data of well P-2 and grid gridblock is shown in Fig. 12 when the minor main correlation length is 100, 200, and 300 m, respectively. The area correlated to the measurement data at well P-2 increases with an increase in the correlation length. Fig. 13 shows the estimation of the permeability field using different correlation lengths. Fig. 14 shows the predicted water production curves for well P-2. With an increase in the correlation length, the correlation area to the measurement data becomes larger and more information can be obtained

during the data assimilation procedure. However this will in turn result bigger spurious correlation. As shown in Fig. 11, the estimation of the permeability field is closer to the true field with increasing correlation length. It can be deduced that the estimation of the permeability will closer to the result obtained from the HIEnKF (shown in Fig. 3(b)) if we increase the length further. From Fig. 14 we find that the band of the predicted water production curve becomes narrow although the data match is similar with the increase in the correlation length. Fig. 15 shows the standard deviation of the predicted water production from different correlation lengths. With increasing correlation length, the correlation area of the production data increases which results in an increase in the scope and scale of the model adjustment. This can lead to lower standard deviation of the predicted curve which means the degrees of the freedom reduces. Actually, the correlation length is affected by many factors, such as geological model, production data type, perforation location, etc. The relationship between the correlation length and these factors require further studies.

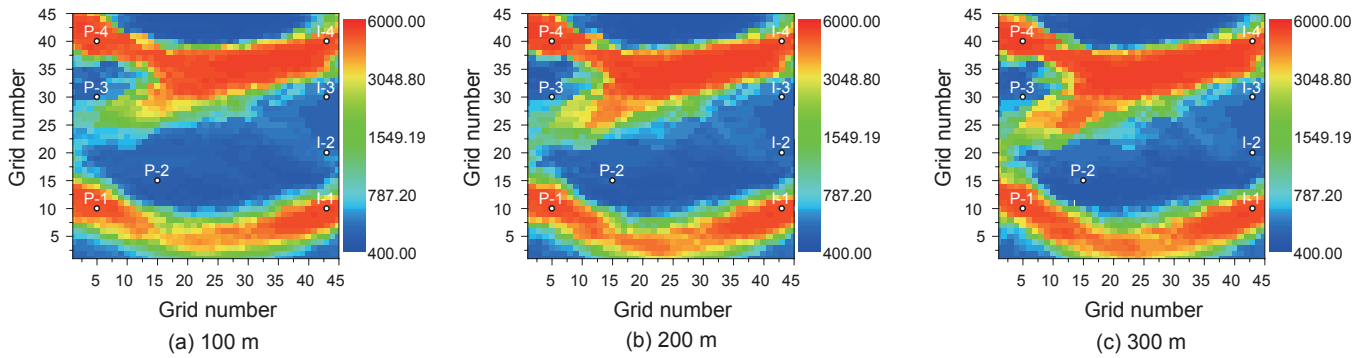


**Fig. 12** The effect of the correlation length on the correlation matrix

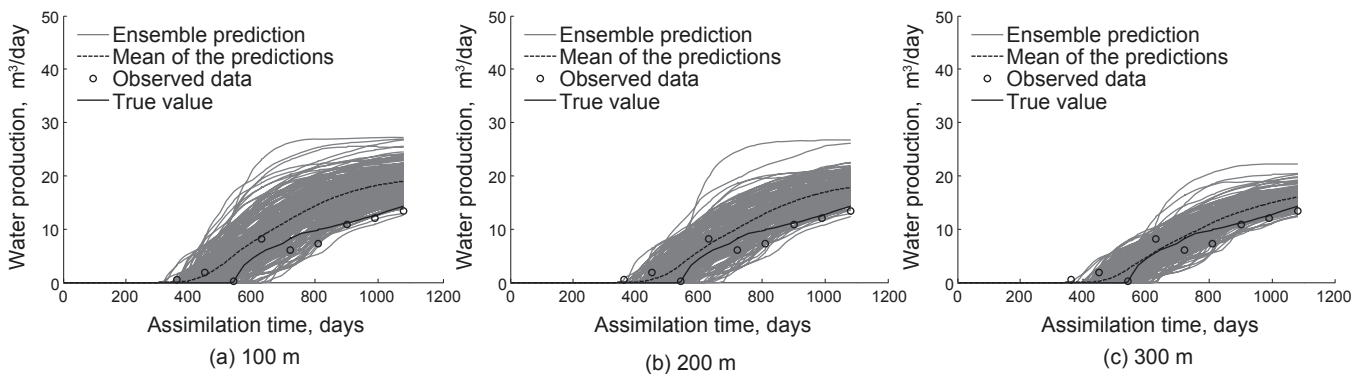
**6 Conclusions**

Two problems of the standard EnKF method need to be

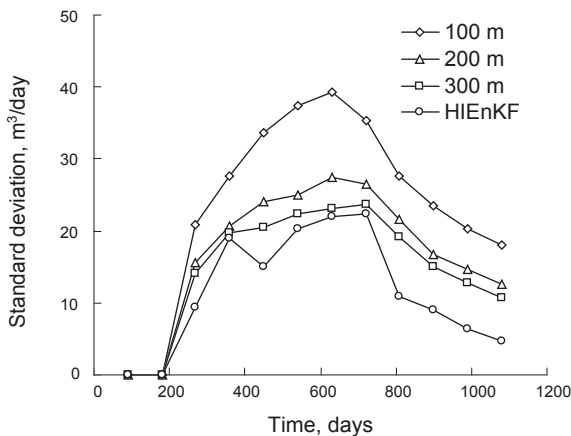
solved. The first is the inconsistency between the updated dynamic field and the updated model for the nonlinear flow problem. Another one is filter divergence and spurious



**Fig. 13** The effect of the correlation length on the estimation of the permeability



**Fig. 14** The effect of the correlation length on the production predictions



**Fig. 15** The effect of the correlation length on the standard deviation of the water production predictions at well P-2

correlation resulting from the small ensemble size. These two problems will result in poor estimation and data matching.

1) HIEnKF can solve the nonlinear problem of flow in porous media. During the data assimilation procedure, only model parameters are updated. The predicted data is obtained by rerunning the simulator from time zero using the updated model. The effect of the updated parameters on production data can be reflected by doing this procedure. The inconsistency problem is solved by this method although the expenses of the computation increased compared to the

standard EnKF. It is proved that HIEnKF is similar to the first iteration of the EnRML.

2) We introduce covariance localization technology to eliminate the filter divergence and spurious correlation due to the small size of the ensemble. By defining the shape and the size of the correlation zone, the spurious correlation of grids far apart is avoided. Covariance localization can maintain the freedom degrees of the ensemble and get a good data match and model estimation.

3) As the correlation length decreases, the degree of freedom increases and the ensemble can match more future data. However too small a correlation length loses some reservoir information.

4) The shape and size of the correlation zone are affected by many factors, such as heterogeneity of the geological model, production data type, perforation location, etc. The relationship between the correlation length and these factors require further studies.

### Acknowledgements

The authors are grateful for financial support from the Shandong Natural Science Foundation (Grant No. ZR2010EM053) and the Fundamental Research Funds for the Central Universities (Grant No. 10CX04042A)

### References

Aanonsen S I, Nævdal G, Oliver D S, et al. The ensemble Kalman filter

- in reservoir engineering: A review. *SPE Journal*. 2009. 14(3): 393-412 (paper SPE 117274)
- Agbalaka C C and Oliver D S. Application of the EnKF and localization to automatic history matching of facies distribution and production data. *Mathematical Geostatistics*. 2008. 40: 353-374
- Anderson J L. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus*. 2007. 59A: 210-224
- Bianco A, Cominelli A, Dovera L et al. History matching and production forecast uncertainty by means of the ensemble Kalman filter: A real field application. Paper SPE 107161 presented at EUROPEC/EAGE Conference and Exhibition, June 11-14, 2007, London
- Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*. 1994. 99(C5): 10143-10162
- Evensen G. *Data Assimilation: The Ensemble Kalman Filter*. Springer, Berlin. 2007
- Evensen G, Hove J, Meisingset H C, et al. Using the EnKF for assisted history matching of a North Sea reservoir. Paper SPE 106184 presented at SPE Reservoir Simulation Symposium, February 26-28, 2007, Houston, Texas
- Gaspari G and Cohn S E. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*. 1999. 125(554): 723-757
- Gu Y and Oliver D S. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE Journal*. 2007. 12(4): 438-446 (Paper SPE 108438)
- Han X J and Li X. Review of the nonlinear filters in the land data assimilation. *Advances in Earth Science*. 2008. 23(8): 813-820 (in Chinese)
- Haugen V, Natvik L J, Evensen G, et al. History matching using the ensemble Kalman filter on a North Sea field case. Paper SPE 102430 presented at SPE Annual Technical Conference and Exhibition, September 24-27, 2006, San Antonio, Texas
- Houtekamer P L and Mitchell H L. A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*. 2001. 129(1): 123-137
- Jafarpour B and McLaughlin D B. Efficient permeability parameterization with the discrete cosine transform. *SPE Journal*. 2009. 14(1): 182-201 (Paper SPE 106453)
- Jia B H, Xie Z H, Tian X J, et al. A soil moisture assimilation scheme based on the ensemble Kalman filter using microwave brightness temperature. *Science in China Ser. D Earth Sciences*. 2009. 52: 1835-1848 (in Chinese)
- Li G and Reynolds A C. An iterative ensemble Kalman filter for data assimilation. *SPE Journal*. 2009. 14(3): 496-505 (paper SPE 109808)
- Li R, Reynolds A C and Oliver D S. History matching of three-phase flow production data. *SPE Journal*. 2003. 8(4): 328-340 (paper SPE 66351)
- Li X J, Cheng S Q, Li Y Z, et al. Review on reservoir description and optimization problems with ensemble Kalman filter. *Inner Mongolia Petrochemical Industry*. 2009. 4: 1-5 (in Chinese)
- Naevdal G, Johnsen L M, Aanonsen S I, et al. Reservoir monitoring and continuous model updating using ensemble Kalman filter. Paper SPE 84372 presented at SPE Annual Technical Conference and Exhibition October 5-8, 2003, Denver, Colorado
- Oliver D S, Reynolds A C and Liu N. *Inverse Theory for Petroleum Reservoir Characterization and History Matching*. Cambridge: Cambridge University Press. 2008
- Qin J, Yan G J, Liu S M, et al. Application of ensemble Kalman filter to geophysical parameters retrieval in remote sensing: A case study of kernel-driven BRDF model inversion. *Science in China Ser. D Earth Sciences*. 2005. 35 (8): 790-798 (in Chinese)
- Sarma P, Durlowsky L J and Aziz K. Kernel principal component analysis for efficient differentiable parameterization of multipoint geostatistics. *Mathematic Geosciences*. 2008. 40: 3-32
- Seiler A, Evensen G, Skjervheim J A. Advanced reservoir management workflow using an EnKF based assisted history matching method. Paper SPE 118906 presented at SPE Reservoir Simulation Symposium, February 2-4, 2009, Woodlands, Texas
- Strebelle S. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*. 2002. 34: 1-21
- Thulin K, Li G, Aanonsen S I, et al. Estimation of initial fluid contacts by assimilation of production data with EnKF. Paper SPE 109975 presented at SPE Annual Technical Conference and Exhibition, November 11-14, 2007, Anaheim, California
- Wang Y, Li G and Reynolds A C. Estimation of depths of fluid contacts and relative permeability curves by history matching using iterative ensemble Kalman smoothers. *SPE Journal*. 2010. 15(2): 509-525 (Paper SPE 119056)
- Zafari M. Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter. Master Thesis. University of Tulsa, 2005
- Zhao H L, Feng X T, Zhang D X, et al. Spatial variability of geomechanical parameter estimation via ensemble Kalman filter. *Rock and Soil Mechanics*. 2007. 28(10): 2219-2223 (in Chinese)

(Edited by Sun Yanhua)