Journal of Computational Physics 229 (2010) 8011–8030

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

History matching of facies distribution with the EnKF and level set parameterization

Haibin Chang [a], Dongxiao Zhang [a,b,*], Zhiming Lu [c]

[a] Department of Energy and Resources Engineering, College of Engineering, Peking University, Beijing 100871, China
[b] The Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, Los Angeles, CA 90089, USA
[c] Computational Earth Sciences Group, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

ARTICLE INFO

Article history:
Received 14 August 2009
Received in revised form 3 April 2010
Accepted 1 July 2010
Available online 7 July 2010

Keywords:
Ensemble Kalman filter
Level set parameterization
Facies distribution

ABSTRACT

In this work, we develop a methodology to combine the Ensemble Kalman filter (EnKF) and the level set parameterization for history matching of facies distribution. With given prior knowledge about the facies of the reservoir geology, initial realizations are generated by commonly used software as the prior guesses of the unknown field. Furthermore, level set functions are used to reparameterize these initial realizations. In the reparameterization process, a representing node system is set up, on which the values of level set functions are assigned using Gaussian random numbers. The mean and the standard deviation of the Gaussian random numbers are designed according to the facies proportion, and the sign of the random numbers depends on the facies type at the representing nodes. The values of the level set functions at the other grid nodes are obtained by linear interpolation. The level set functions on the representing nodes are the model parameters of the EnKF state vector and are updated in the data assimilation process. On the basis of our numerical examples for two-dimensional reservoirs with two or three facies, the proposed method is demonstrated to be able to capture the main features of the reference facies distributions.

1. Introduction

The rock properties of a reservoir are crucial factors for determining its production behavior. A typical reservoir consists of several distinct rock units (which are usually called facies), each of which may have relatively uniform properties such as permeability and porosity. Owing to the significantly different properties between facies, the fluid flow is strongly impacted by the facies distribution. If the properties of each facies are known, a good estimation of the facies boundary is extremely important for predicting or optimizing the future reservoir production.

Information of a reservoir usually comes from production wells, which may include static facies type observations and dynamic production records. There exist two different types of methods for estimating the facies distribution by assimilating the observation data: Gradient-based method and Monte Carlo method. The level set method is a typical gradient-based method [2,14,17–19,21], in which facies distribution is represented by level set functions and the facies boundaries are evolved to reduce an objective function. In these optimization procedures, gradient calculations are needed and obtained by either Jacobian method or adjoint method. The level set method was originally used for tracing the evolution of interfaces [18,19]. Using the level set method for permeability estimation can also be found in the literature [2,17,21].

* Corresponding author at: Department of Energy and Resources Engineering, College of Engineering, Peking University, Beijing 100871, China. Tel.: +86 010 6275 7432; fax: +86 010 6275 7426.
E-mail addresses: donzhang@usc.edu, dxz@pku.edu.cn (D. Zhang).

0021-9991/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2010.07.005

Ensemble Kalman filter (EnKF) is a Monte Carlo approach for history matching of production data [6,7,9]. In the EnKF, a large number of initial realizations are generated as the prior guesses of the unknown field. At the data assimilation steps, each realization will be updated based on the correlation between the observation and the parameters. The EnKF is only optimal for linear problem with Gaussian statistics. The Gaussian statistics limitation is a major challenge for the application of the EnKF to the estimation of facies distributions, because the facies distributions are usually multimodal. Truncated plurigaussian is one of the approaches to parameterize the facies distribution [1,8,12,13,23]. That is, the facies distribution is generated by truncating two or more Gaussian random fields into separate regions that denote different facies types. In [13], Liu and Oliver suggested using two Gaussian random fields in the state vector of EnKF instead of the rock properties because it is more likely to satisfy the Gaussian distribution assumption. The application of truncated plurigaussian method is extended to a 3D reservoir in [1] and the authors also proposed to decouple the assimilation of production and facies data where production data are assimilated first and facies data afterward in order to avoid improper weighting of the production data that could lead to large changes in the updated state variables. In [23], a different approach to ensure the ensemble members honor facies observations at wells is proposed and the difficulty of their work is the larger nonlinearity in the data. Truncated plurigaussian is suitable for geometries with two-point statistics, which can honor the facies proportions and indicator variograms. For curvilinear geometries, however, truncated plurigaussian is not a proper option because a representation of such geometries may require multiple-point statistics [20].

Other approaches on using EnKF to estimate facies distribution include: Gaussian mixture models [5], discrete cosine transform [10], and level set methods [16]. In [5], Dovera and Della Rossa proposed a new analysis scheme in which they derived expressions for the conditional means, conditional covariance, and weights for the Gaussian mixture model to make the EnKF formulation suitable for Gaussian mixture model. In [10], Jafarpour and McLaughlin used discrete cosine transform (DCT) to parameterize both the state variables and model parameters and the coefficients of the retained cosine basis functions were updated with EnKF. The updated model parameters and state variables were then constructed from these updated coefficients. In [16], Moreno and Aanonsen proposed to combine the level set method with EnKF. In their work, they assumed there exist a "best guess" model and the facies boundaries be implicitly modeled through a level set function. The boundary velocity was modeled as a Gaussian random field and updated with the EnKF. At each updating time step, the new updated velocity field was applied to the same original level set function to get new realizations. The results of this method depend on the quality of the "best guess" model: If it cannot represent the true geology properly it is difficult to get good results.

For the EnKF method, the quality of the initial realizations is an important factor that will impact the history matching results. The initial realizations should honor the statistical feature of the geologic formation and also obey the Gaussian statistics limitation of EnKF, which is especially difficult when dealing with curvilinear geometries. Regardless of the Gaussian statistics limitation, generating random geometries with two-point or multiple-point statistics can be performed with the aid of existing software [3,4,20]. However, the outputs of these software packages are simple indicator fields, which are definitely not Gaussian random fields. How to utilize these initial representations of the unknown random field to perform history matching is the main objective of the present work. In this study, we combine the EnKF with the level set parameterization for history matching of the facies distribution. With the prior knowledge about the facies distributions, initial realizations are generated by commonly used software. Reparameterization of these realizations is performed with the aid of level set functions. We first design a representing node system and generate random numbers for the level set function values at these nodes. For each realization, we generate a set of Gaussian random numbers. The mean and the standard deviation of the Gaussian random numbers are designed according to the facies proportion and the sign of the random numbers depends on the facies type at the representing nodes. The values of level set functions at the grid nodes other than the representing nodes are obtained by linear interpolation. The level set functions on the representing nodes are the model parameters of the EnKF state vector and will be updated by assimilating both the static facies type observations and dynamic production records. When the first EnKF update is finished, a second EnKF update may be performed if needed.

This paper is organized as follows: We introduce the forward model and the EnKF methodology in Sections 2 and 3, respectively. In Section 4, we discuss the parameterization procedure in detail. In Section 5, we introduce the model for updating the facies distribution using the EnKF. The case studies are given in Section 6. Some conclusions are drawn in Section 7.

2. Forward model

We consider a two-phase system with water and oil, whose basic flow equations are given as:

$$C_1\nabla\cdot\left[\frac{k_{ro}}{\mu_o B_o}k(\nabla P_o(x,y,z,t)-\gamma_o\nabla Z(x,y,z))\right]=\frac{\phi}{C_2}\frac{\partial}{\partial t}\left(\frac{S_o}{B_o}\right)+q_o(x,y,z,t), \tag{1}$$

$$C_1\nabla\cdot\left[\frac{k_{rw}}{\mu_w B_w}k(\nabla P_w(x,y,z,t)-\gamma_w\nabla Z(x,y,z))\right]=\frac{\phi}{C_2}\frac{\partial}{\partial t}\left(\frac{S_w}{B_w}\right)+q_w(x,y,z,t), \tag{2}$$

subject to the constraint $S_o + S_w = 1$. The subscripts o and w stand for oil and water phases, respectively; C_1 and C_2 are unit conversion factors for converting Eqs. (1) and (2) to oil field units; B_o and B_w are the oil and water formation volume factors; μ is the viscosity; γ is the specific density; ϕ is the porosity; Z is the vertical distance from a datum level; q is the sink or source term; k_r is the relative permeability; k is the intrinsic permeability; P is the pressure; and S is the saturation.

The rock property fields such as permeability and porosity are crucial input parameters for this forward model. The measurements of these fields are usually obtained from wells. Owing to the limited number of measurements, the rock properties fields can be only depicted with high uncertainty. In this study, the rock property fields are represented as zones with unknown boundaries. The purpose of this study is to identify the boundary while assuming that the rock properties for each facies are known. The details will be given in the following sections.

3. Ensemble Kalman filter

The ensemble Kalman filter is a Monte Carlo method and it propagates an ensemble of initial reservoir realizations along time to assimilate data. The ensemble of state vectors, Ψ, denoted by

$$\Psi = \{\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots, \boldsymbol{y}^{N_e}\} \in \mathbb{R}^{N_y \times N_e}, \tag{3}$$

where N_e denotes the total number of realizations; $\boldsymbol{y}^i$, $i = 1, \ldots, N_e$, are state vectors. A typical state vector consists of model parameters ($\boldsymbol{m}$), state variables ($\boldsymbol{u}$), and observations ($\boldsymbol{d}$). At a certain time step t_k for $k = 1, \ldots, N_t$, the state vector is expressed as

$$\boldsymbol{y}_k = \begin{Bmatrix} \boldsymbol{m} \\ \boldsymbol{u}_k \\ \boldsymbol{d}_k \end{Bmatrix} \in \mathbb{R}^{N_m + N_u + N_d}. \tag{4}$$

Model parameters are variables that are not varying in time, which may include rock properties such as permeability and porosity. State variables are time-dependent variables, which may include pressure and saturation for all phases. The observations may include production records and facies type observations.

The main procedure for EnKF contains two sequential steps: the forecast step and the update step. The forecast model is performed on each ensemble member independently:

$$\boldsymbol{y}_k^{p,j} = F(\boldsymbol{y}_{k-1}^{a,j}), \quad j = 1, \ldots, N_e, \tag{5}$$

where $F(\cdot)$ is a forecast operator, representing the reservoir simulator in our study; superscripts p and a denote the "predicted" and "analyzed" states, respectively. When the observation data are available, the update step is done via

$$\boldsymbol{y}_k^{a,j} = \boldsymbol{y}_k^{p,j} + \boldsymbol{K}_k(\boldsymbol{d}_k^j - \boldsymbol{H}_k \boldsymbol{y}_k^{p,j}), \quad j = 1, \ldots, N_e, \tag{6}$$

where $\boldsymbol{K}$ is the Kalman gain matrix; $\boldsymbol{H}$ is the observation operator, which represents the relationship between the state vector and the observation vector:

$$\boldsymbol{H}_k = [0 \ \mathbf{I}], \tag{7}$$

where $\mathbf{0}$ is an $N_d \times (N_m + N_u)$ matrix with all 0's as its entries; $\mathbf{I}$ is an $N_d \times N_d$ identity matrix, and $\boldsymbol{d}_k^j$ is the observation data at current time plus observation noises

$$\boldsymbol{d}_k^j = \boldsymbol{d}_{obs,k} + \varepsilon_k^j, \quad j = 1, \ldots, N_e. \tag{8}$$

The ensemble Kalman gain takes the following form,

$$\boldsymbol{K}_k = \boldsymbol{C}_{y_k^p} \boldsymbol{H}_k^T \left(\boldsymbol{H}_k \boldsymbol{C}_{y_k^p} \boldsymbol{H}_k^T + \boldsymbol{C}_{d_k} \right)^{-1}, \tag{9}$$

where $\boldsymbol{C}_y$ denotes the state cross-covariance matrix, $\boldsymbol{C}_d$ is the error covariance matrix, and the superscript T stands for matrix transpose. The cross-covariance matrix can be calculated using standard statistical formula,

$$\boldsymbol{C}_{y_k^p} \approx \frac{1}{N_e - 1} \sum_{n=1}^{N_e} \left\{ \left[\boldsymbol{y}_k^{p,n} - \langle \boldsymbol{y}_k^p \rangle \right] \left[\boldsymbol{y}_k^{p,n} - \langle \boldsymbol{y}_k^p \rangle \right]^T \right\}. \tag{10}$$

Since only partial entries of the cross-covariance are needed in Eq. (9), one does not need to calculate the entire matrix.

The EnKF is only optimal for linear problem with Gaussian statistics in the limit of infinitely many model realizations. The major difficulty in applying the EnKF to history matching of the facies distribution is the highly non-Gaussian property of the parameter fields. The parameter fields considered here, such as permeability and porosity, are discontinuous, zonally constant functions with the assumption that the properties inside each facies are constant. Reparameterization of the facies distribution is necessary to make the EnKF algorithm applicable under this condition.

4. Reparameterization of the parameter field

4.1. Level set method and the representation of the parameter field

The idea of the level set method is briefly introduced here and more detailed descriptions of the method can be found elsewhere [18,19]. Let Γ be a closed curve in domain Ω and it partitions Ω into two regions D and D^C ($D^C = \Omega \setminus D$ is the complement of D). We define a function Φ that satisfies the following properties:

$$\begin{aligned} &\Phi(\boldsymbol{x}) > 0 \quad \text{for } \boldsymbol{x} \in D; \\ &\Phi(\boldsymbol{x}) < 0 \quad \text{for } \boldsymbol{x} \in D^C; \\ &\Phi(\boldsymbol{x}) = 0 \quad \text{for } \boldsymbol{x} \in \Gamma, \end{aligned} \tag{11}$$

Γ is the zero level set of the function Φ, and Φ is called the level set function for Γ. The unit normal to the surface, $\boldsymbol{n}$, can be given by

$$\boldsymbol{n} = \frac{\nabla \Phi}{|\nabla \Phi|}, \tag{12}$$

where $|\nabla \Phi| = \sqrt{\nabla \Phi \cdot \nabla \Phi}$. The conventional level set methods have been popular in tracking, modeling and simulating the motion of dynamic interfaces. The motion of the interface is determined by a velocity field, $\boldsymbol{v}$, which is a function of position, and the evolution can be described as following

$$\frac{\partial \Phi}{\partial t} + \boldsymbol{v} \cdot \nabla \Phi = 0, \quad \Phi(\boldsymbol{x}, 0) = \Phi_0(\boldsymbol{x}), \tag{13}$$

where $\Phi_0(\boldsymbol{x})$ denotes the initial level set function and t denotes the artificial time. According to Eq. (12), Eq. (13) can be rewritten using the normal velocity, v_n as

$$\frac{\partial \Phi}{\partial t} + v_n \cdot |\nabla \Phi| = 0, \quad \Phi(\boldsymbol{x}, 0) = \Phi_0(\boldsymbol{x}), \tag{14}$$

where $v_n = \boldsymbol{v} \cdot \frac{\nabla \Phi}{|\nabla \Phi|}$. Eq. (14) is called the level set equation and the solution of this equation over artificial time describes the evolution of the interface.

The level set method has also been used for inverse problems dealing with the reconstruction of shapes [2,14,17,21]. For this kind of problems, an objective function needs to be designed first and the gradient method is used to determine the velocity of evolving boundary such that the objective function will be reduced sequentially. In this study, different from the conventional level set method we only use the level set function to represent the parameter fields without involving solution of the let set equation.

For a reservoir with two facies, the parameter field $p(\boldsymbol{x})$ can be written as the following form

$$p(\boldsymbol{x}) = p_1 H(\Phi(\boldsymbol{x})) + p_2 (1 - H(\Phi(\boldsymbol{x}))), \tag{15}$$

where H is the Heaviside function, the coefficients p_1 and p_2 are the parameter values of the facies.

If $p(\boldsymbol{x})$ has more than two different constant values, more level set functions are needed to represent the parameter field. With n level set functions, the domain can be partitioned into 2^n regions, and the parameter field $p(\boldsymbol{x})$ can be represented as [2]:

$$p(\boldsymbol{x}) = \sum_{j=1}^{2^n} p_j \Psi_j(\Phi_1, \ldots, \Phi_n), \tag{16}$$

where

$$\Psi_j(\boldsymbol{x}) = \prod_{r=1}^{n} R_j(\Phi_r(\boldsymbol{x})), \quad R_j(\Phi_r) = \begin{cases} H(\Phi_r) & \text{if } b_r^j = 0, \\ 1 - H(\Phi_r) & \text{if } b_r^j = 1, \end{cases} \tag{17}$$

and b_r^j is the rth element in the binary expansion of $(j-1)$, written in the form of $(b_1^j, b_2^j, \ldots, b_n^j)$.

4.2. Parameterization of the level set functions

In this study, we assume the rock properties of each facies, such as permeability and porosity, are constant and known. Under this assumption, the fluid flow will be only impacted by the spatial distribution of the different facies presented in the reservoir. As mentioned in the pervious subsection, the spatial distribution of the facies can be described by level set functions. Thus the level set functions, defined by their values on all grid nodes, can be used as the model parameters instead of the rock property fields.

Because of the limitations of the EnKF methodology on Gaussian statistics of parameter values, the first issue that needs to be addressed is how to construct initial realizations of level set functions that follow Gaussian statistics, which is required

in updating the level set functions with the EnKF. A natural idea is to assume that the values of level set functions at all grid nodes, $\{\Phi(\boldsymbol{x}_j)\}$, be Gaussian random variables. However, this assumption would lead to three difficulties: (1) how to alleviate the non-uniqueness property of the level set function because the facies type at each grid node is only determined by the sign of the level set function value at that node and the magnitude of the function value is irrelevant; (2) how to build the correlation of the random variables to make the parameter field be jointly Gaussian distribution; (3) how to honor the prior knowledge of the unknown field such as the facies proportion, mean length, and the indicator correlation. The prior knowledge of the unknown field can be obtained by geostatistics.

In order to alleviate the non-uniqueness of the level set function, we use a treatment similar to the idea of master point technique [11,22]. Instead of using the level set function values at all the grid nodes, we only choose part of the grid nodes as representing nodes and take the values of level set functions at these representing nodes as model parameters. The level set function values at the representing nodes are assumed to be Gaussian random variables and the function values at the other nodes are obtained by numerical interpolation. The prerequisite of this scheme is that the numerical grid size is much smaller than the average facies size, which is usually the case for a numerical model. A 2D example is given here to demonstrate the idea. We consider a rectangular domain of lengths L_x and L_y in x and y directions, respectively, and there are N_x and N_y grid blocks in each direction. Suppose we have M_x and M_y representing nodes in x and y axis, respectively, with the constraints of $M_x \leqslant N_x$ and $M_y \leqslant N_y$. Accordingly, there are $M_x \times M_y$ representing nodes in the 2D domain. With the given values at the representing nodes, the level set function values at the remaining nodes can be obtained by using the following equation

$$\Phi(x,y) = \sum_{i=1}^{M_x}\sum_{j=1}^{M_y} \Phi(x_i,y_j)\mathcal{I}_i(x)\mathcal{I}_j(y), \quad x \in [0,L_x] \text{ and } y \in [0,L_y], \tag{18}$$

where $\{\mathcal{I}_j\}$ are interpolation basis functions. Here we choose linear interpolation because of the following considerations. First, the representing nodes are relatively dense in the domain, which will be seen in the later section. In addition, linear interpolation will significantly simplify the relationships among the model parameters, the level set function values on the representing nodes, and the parameter field. This simplification will be useful when performing the EnKF update.

By the linear interpolation, the magnitude of the level set function value at each representing node will affect the facies types at the surrounding grid nodes. That is, the level set function at each representing node has an impact region. This will help alleviate the non-uniqueness behavior. Also, if the distance between the representing nodes is properly designed such that the level set function values on nearby representing nodes are uncorrelated or weakly correlated, we can simply treat the level set function on the representing nodes independently without incurring too much error. How to choose the representing node system and how to honor the prior knowledge will be discussed later.

5. Updating Facies distributions with EnKF

We first consider a two-facies reservoir where the facies distribution can be described by one level set function, $\Phi(x)$. The model parameters are the function values at the representing nodes, $\{\Phi(x_j)\}_{j=1}^{M_x\times M_y}$, which are assumed to be Gaussian random variables. The state variables contain pressure ($\boldsymbol{P}$) and saturation ($\boldsymbol{S}_w$) at all grid nodes. Then the state vector takes the following form

$$\boldsymbol{y}_j = \left[\{\Phi(x_j)\}^T, \boldsymbol{P}^T, \boldsymbol{S}_w^T, \boldsymbol{d}^T\right]_j^T, \tag{19}$$

here the observations, $\boldsymbol{d}$, include the production data from the wells such as oil production rate and bottom hole pressure and also include the facies type observations at the well locations.

The facies type observations at well locations are non-numerical data. One solution to this problem is to introduce a proxy, f, which is defined as [1,13]

$$f = \begin{cases} 0, & F_{sim} = F_{obs}, \\ 1, & F_{sim} \neq F_{obs}, \end{cases} \tag{20}$$

where F denotes the facies type, and subscripts *sim* and *obs* denote simulation and observation, respectively.

At the data assimilation steps, a facies matching loop is usually needed to make sure that each ensemble member matches the facies type observations at the well locations [13]. Here we take the method described in [1]: At the data assimilation steps, we first assimilate the production data and then iterate on the facies constraint if the facies type observations are violated after the EnKF update. In performing the assimilation of the facies observation data, the state vector is redefined as [1]

$$\boldsymbol{y}_j = \left[\{\Phi(\boldsymbol{x}_j)\}^T, \boldsymbol{F}^T\right]_j^T. \tag{21}$$

The advantage of doing this is that the enforcement of facies constraint will not affect the updated state variables. This kind of facies constraint enforcement procedure does not need to rerun the forward model, thus resulting in high efficiency.

If the reservoir has more than two facies, the above procedure is the same except for a need of more level set functions.

6. Case studies

6.1. Simulation model

A synthetic two-dimensional reservoir model is used here to test the performance of the methodology described above. It is a five-spot pattern with an injector located at the center and four producers located at the corners. The reservoir is a square with side length 1500 ft, and is evenly divided into 50 grid blocks in both directions. The reservoir has a thickness of 60 ft and located at a depth of 4000 ft. The producer is controlled by the BHP constraint of 3000 psi. The injector is controlled by the surface flow rate target of 5000 stb/day. The injection starts at the first day of the production. The simulation lasts for 900 days. The observation data include facies type observations at the well locations and production data (which are oil production rate and water cut of the producers as well as BHP of the injector). The observations are available every 30 days.

The perturbation of observation by adding additive noise is necessary to avoid divergence of the filter [6,7]. The measurement errors are drawn from a zero-mean Gaussian distribution. For the production data, the standard deviation is set to 1% of the actual measurement. No noise has been added for facies type observations.

6.2. Case 1: Two facies case with given facies proportion and indicator correlation

In this case, we consider a two-facies reservoir. The permeability of facies 1 and facies 2 are 300 md and 50 md, respectively. The porosity of facies 1 and facies 2 are 25% and 20%, respectively. About the prior knowledge of the facies distribution, we assume that the facies proportion and the correlation of the indicator random variables are known. Here, the indictor function is a traditional tool used for depicting facies distribution. The definition of indicator function, $I_i(x)$, is

$$I_i(x) = \begin{cases} 1 & \text{if facies } i \text{ occurs at location } x \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

Unlike the continuous level set functions, the indicator function uses one or zero to denote whether or not a specific facies occurs. An indicator variogram is usually used for depicting a facies distribution.

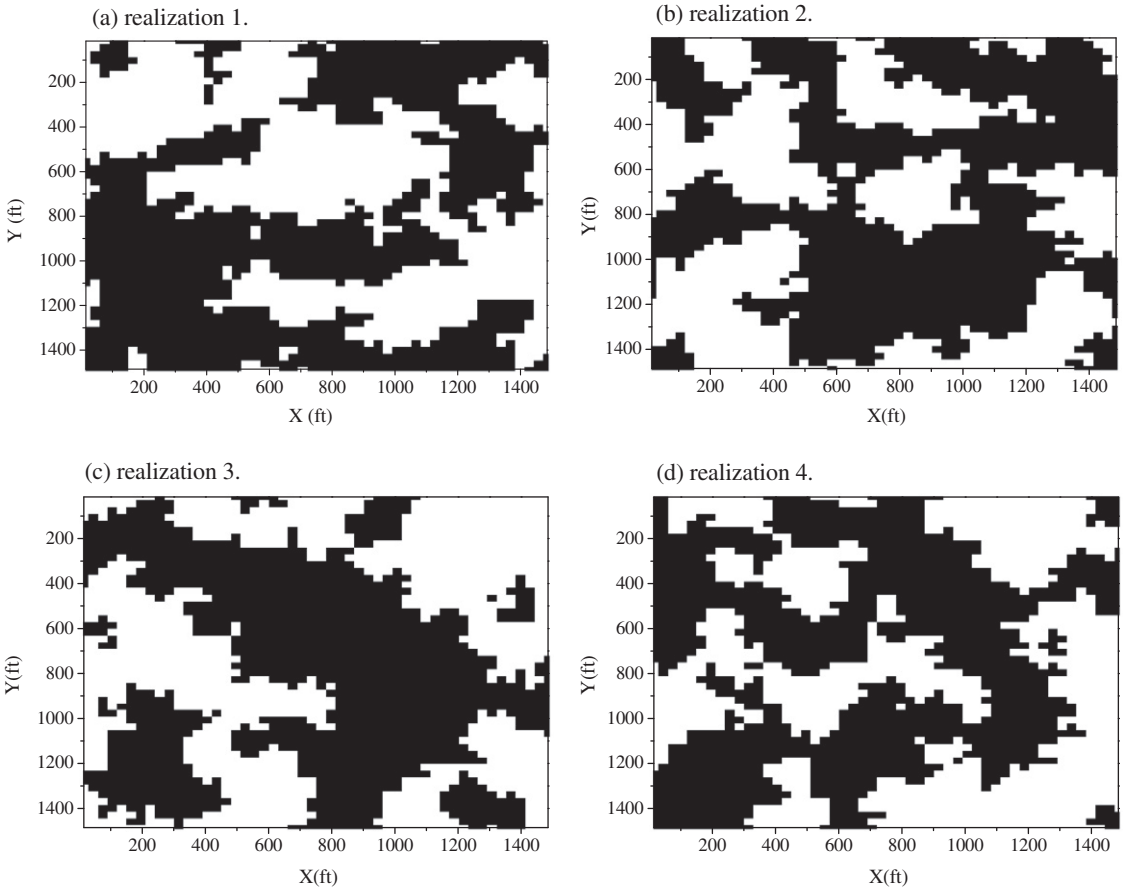

Fig. 1. Initial realizations generated by SISIM program (white for facies 1 and black for facies 2).

In this example, we assume that two facies have equal proportion and the indicator variogram model is Gaussian. With the given knowledge, initial realizations of the facies distribution can be generated using existing software. For example, the SISIM program of the GSLIB Fortran library [4] and Transition Probability Geostatistical Software (T-PROGS) [3] can be used to generate geometries with two-point statistics and the SNESIM program [20] is suitable to generate geometries with multiple-point statistics. In this study, we choose the SISIM program to generate the initial realizations and some realizations are illustrated in Fig. 1. It is seen that the realizations are quite different with each other even though they follow the same statistical characteristics. The initial model parameter ensemble of the EnKF process is obtained by reparameterization of these realizations. The reparameterization procedure is illustrated in Fig. 2, where Fig. 2a shows one realization generated by SISIM. As discussed before, the first step of the reparameterization is to design a representing node system for the level set functions. The representing node system should be designed in accord with the characteristics of the facies distribution. Here we use a constant interval representing node system and the interval width is an important parameter that needs to be chosen wisely. The representing nodal interval determines the total number of representing nodes, which is the degree of freedom of the model parameters as shown in Eq. (19). A dense representing node system will help the reconstruction of the initial realization more accurately, but it will increase the degree of the freedom of the model parameters and also decrease the impact region of each representing node, which will increase the difficulties in the data assimilation process. Also in order to simplify the problem, we would like the correlation with respect to the representing nodes interval to be small. Then the random variables on the representing nodes can be treated independently. As investigated in [15], under some conditions, a two-facies reservoir with equal mean length, the correlation length of the indicator random variables is half of the facies mean length. Although the correlation is not strictly zero when the separation distance equals the correlation length, the random variables are only weakly correlated with that separation distance. Thus we set the interval of the representing nodes as half of the facies mean length in each direction. If the mean length of each facies is different, we use the smallest mean length as the reference for the choosing the representing nodes interval. For the case considered here, the two facies have the same mean length of about 300 ft in both x and y directions and the representing nodal interval is chosen to 150 ft in both x and y directions. The representing node system is shown in Fig. 2b.

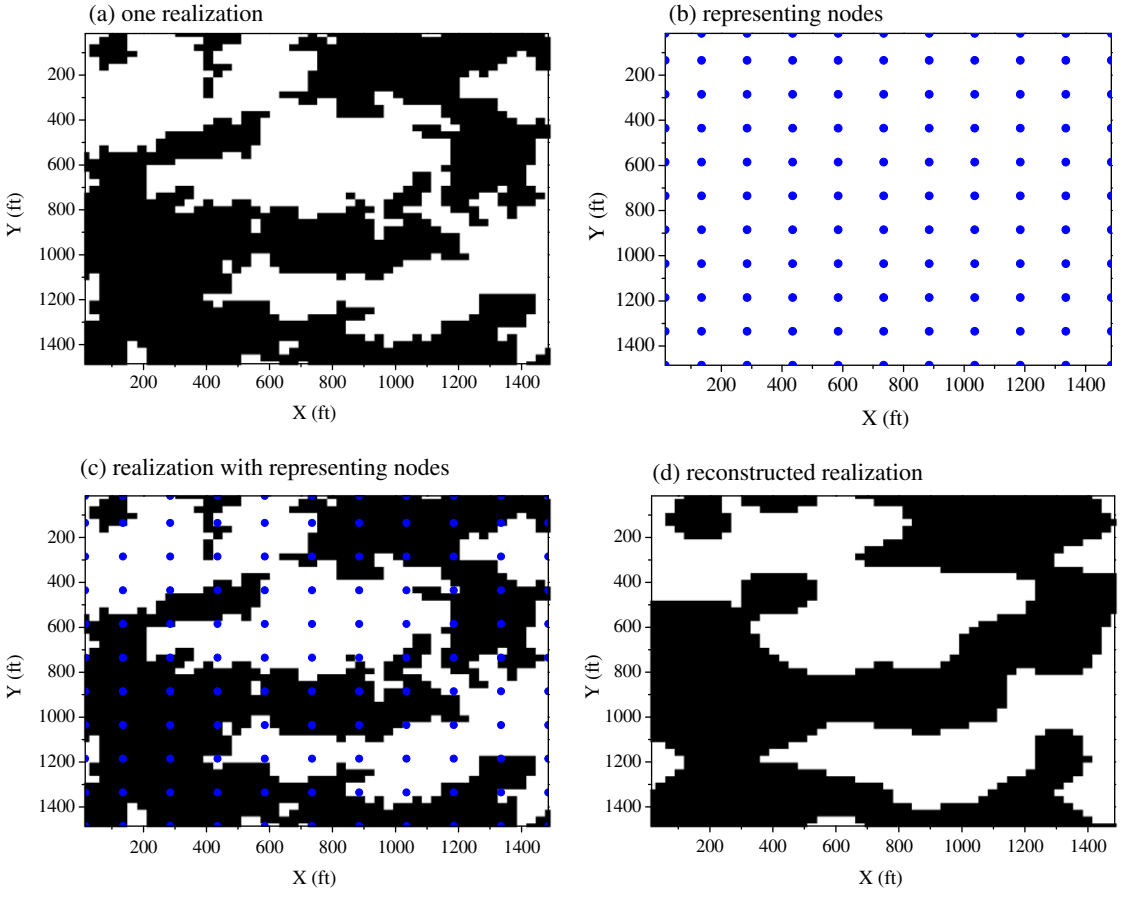

Fig. 2. The procedure of reparameterization: (a) the SISIM realization; (b) the representing node system; (c) the facies distribution with representing nodes; and (d) the reconstructed facies distribution.

After the representing nodes are designed, the next step is generating the level set function values on these representing nodes. We generate Gaussian random numbers for all representing nodes, $\{\Phi(x_j)\}_{j=1}^{M_x \times M_y}$, independently. The mean and the standard deviation of the Gaussian random numbers should be designed according to the facies proportion and the sign of the random numbers depends on the facies type at the representing nodes as shown in Fig. 2c. Actually we perform the following

$$\begin{cases} \text{generate } \Phi(x_j) \text{ that satisfy } \Phi(x_j) \in N(u,\sigma) \text{ and } \Phi(x_j) < 0 & \text{if facies 1 occurs at location } x_j \\ \text{generate } \Phi(x_j) \text{ that satisfy } \Phi(x_j) \in N(u,\sigma) \text{ and } \Phi(x_j) > 0 & \text{if facies 2 occurs at location } x_j, \end{cases} \tag{23}$$

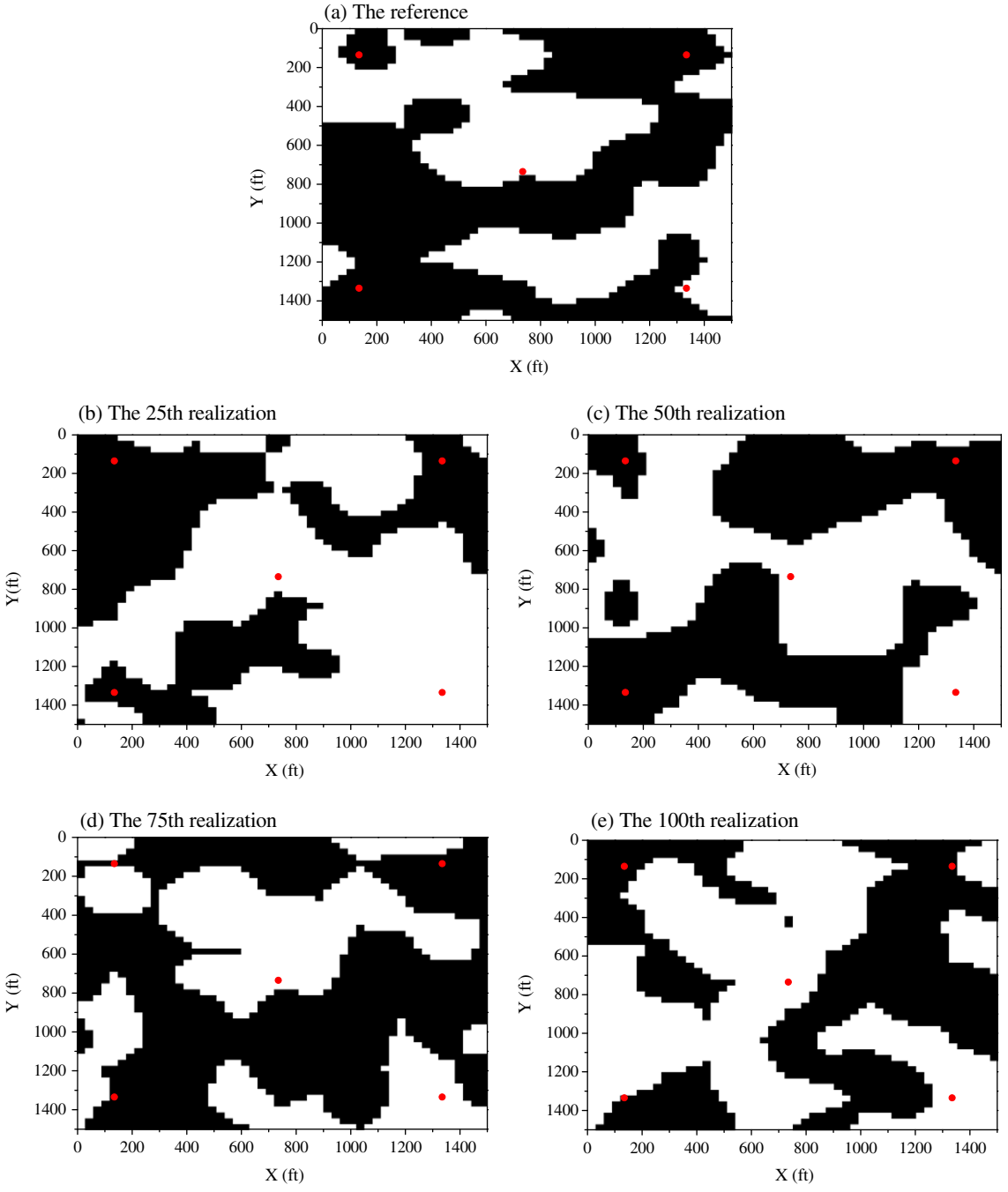

Fig. 3. The reference (a) and four selected conditional realizations (b, c, d, and e). The red circles denote well locations, and injector is at the center and the producers at the corners. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where u and σ denote the mean and the standard deviation of the Gaussian random variables.

Because the facies proportion of the two facies is the same in this case, the mean of the Gaussian random variables is zero. We set the standard deviation of the Gaussian variables to be 0.5. The standard deviation of the Gaussian variables should not be too large because otherwise some of the generated random numbers may be far away from zero and it is not easy to change sign after update, which means difficult to adjust facies type for the large random numbers. Also, if the standard deviation is too small the facies type is too easy to be adjusted even if there is only a small innovation. After we get the values for $\{\Phi(x_j)\}_{j=1}^{M_x\times M_y}$, the level set function values on all the grid nodes, $\{\Phi(x_j)\}_{j=1}^{N_x\times N_y}$, are interpolated using Eq. (18). With the level set function, the facies distribution of the reservoir will be determined as shown in Fig. 2d. By comparing Fig. 2d with Fig. 2a, we can see that the reconstructed facies distribution looks quite similar to the original facies distribution. Since the realizations generated by SISIM can honor the prior knowledge, the reconstructed realizations inherit this property. With such a reconstruction, we can parameterize the facies distribution by the level set function values on the representing nodes. In the history matching process, the observation data are assimilated to update the level set function values on the representing nodes to adjust the facies distribution.

The facies distribution shown in Fig. 2d is selected as the reference field. And 100 more realizations generated by conditional SISIM simulations are reconstructed to make up the initial ensemble of the EnKF. Here we use conditional SISIM simulations to generate the initial realizations that honor the facies type observations. Four conditional realizations are illustrated in Fig. 3, from which we can see that although these realizations have the same facies type at the well locations as the reference, they still look quite different to the reference. Then we perform the reservoir simulation and assimilate the production data when they are available. At each data assimilation step, if the facies types are mismatched at the well locations for some realizations after assimilating the production data, the facies matching loop will be performed. Four updated realizations are shown in Fig. 4. By comparing with Fig. 3, we can see that great improvements can be found after assimilating the production data and the updated realizations resemble the major patterns in the reference. It is seen that, the production data is more useful than the facies data. This is because the impact region of a facies data is limited by the facies size and is thus usually much smaller than that of a production data, especially the oil production rate that provides average information between the injector and producer. Fig. 5 shows the match to the oil production rate from the initial ensemble

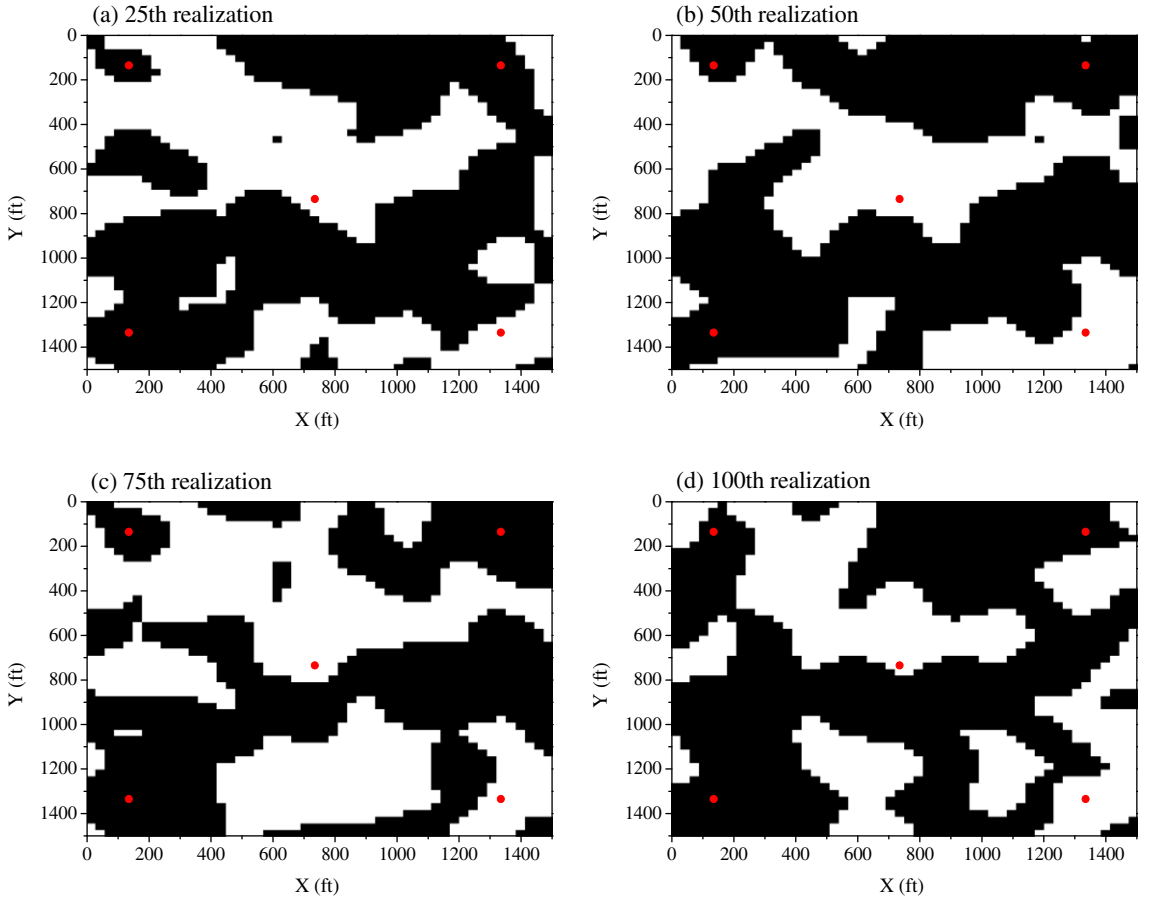

Fig. 4. The four realizations (shown in Fig. 3) after assimilating the production data.

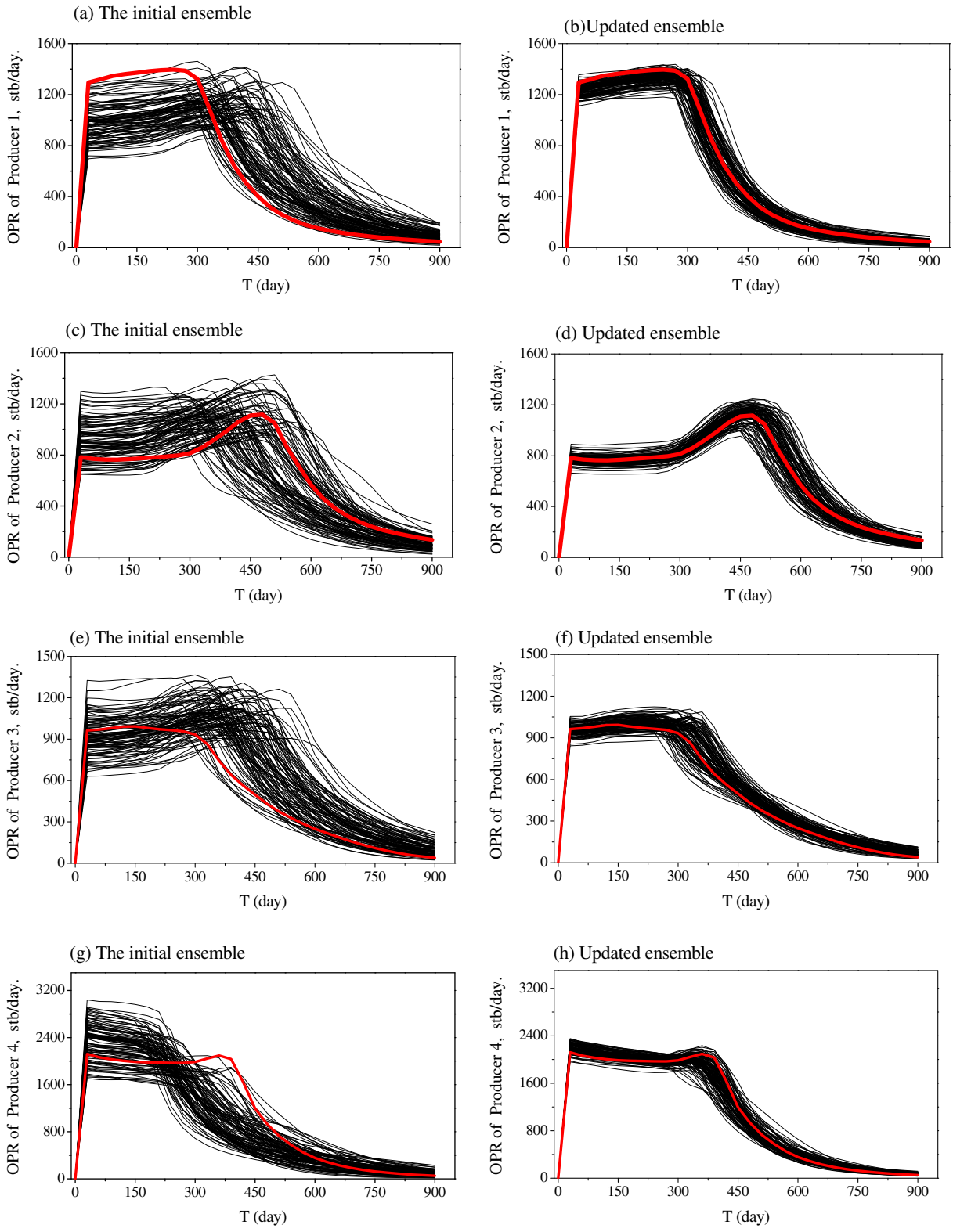

Fig. 5. The match to the oil production rate from the initial ensemble and the updated ensemble. The red curve depicts the responses computed using the reference parameter field, and the black curves represent the responses from the ensemble. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and the updated ensemble. We can see that the matches are improved significantly after the EnKF update. Fig. 6 shows the mean and standard deviation of the initial and updated level set function realizations at the representing nodes along the profile y = 15 ft. From Fig. 6a, we can see that the mean of the initial realizations is nearly zero and the standard deviation is around 0.5, which are exactly the statistics we prescribed for the random variables on the representing nodes. The statistics of the updated realizations is shown in Fig. 6b. We can see that the mean of the level set function realizations are no longer zero because the updated realizations will try to match the facies types on the representing nodes in the reference field. The standard deviation of the updated realizations decreases significantly because all these updated realizations become close to the true field. Fig. 7 illustrates the probabilistic map for facies 1, computed from all updated realizations. In this map, the white shade corresponds to a probability of 1 and the black shade to zero probability that facies 1 is observed at a specific location in the domain. The different shades of gray represent probability values between zero and one. For this case, the true probabilistic map for facies 1 is the same as Fig. 3a. By comparison, we can see that the estimated probabilistic map is able to capture the main feature of the true probabilistic map.

In designing the representing node system, we suggested in Section 4 that the system should not be too dense. In order to test this, here we design a case with a denser representing node system. The representing nodal interval for this case is set to be half of the previous case, being 75 ft. Then the same conditional SISIM realizations are reparameterized to get the model parameters for EnKF. We next perform the EnKF update. Four updated realizations from EnKF are shown in Fig. 8. We can see that, the results are not satisfactory and there are too much small scale features. One reason for this behavior is that we treat the random variables on the representing nodes independently, which is not reasonable for a dense representing node system. Also a dense representing node system will decrease the impact region of each representing node.

6.3. Case 2: Two facies case with given facies proportion and mean length

In the previous subsection, we tested the performance of the proposed method using a case with facies proportion and indicator correlation given as prior knowledge. Such prior knowledge is traditionally obtained by geostatistics and the infor-

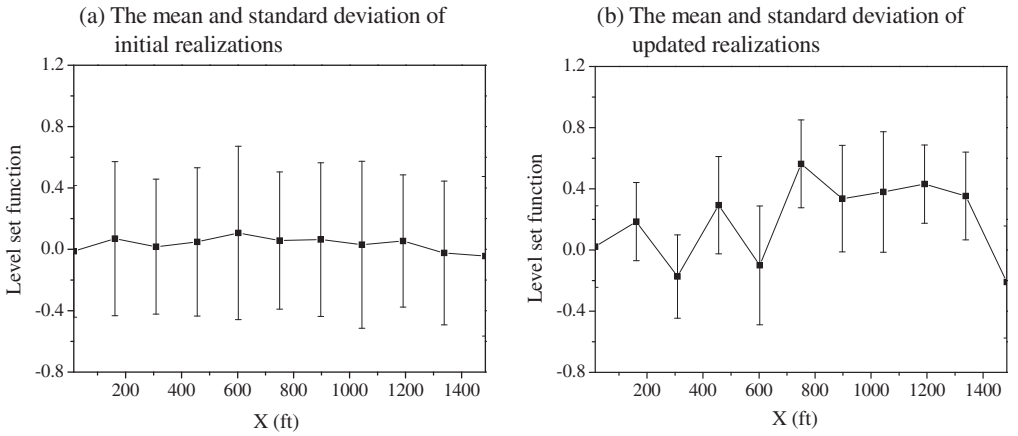

Fig. 6. The mean and standard deviation of the level set function realizations on the representing nodes at y = 15 ft: (a) from the initial realizations and (b) from the updated realizations.

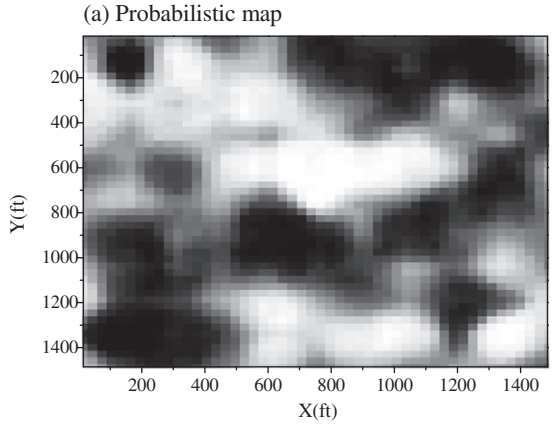

Fig. 7. The probabilistic map for facies 1 computed from all updated realizations.

mation of a reservoir is usually from well logs. Since the wells in a reservoir are usually very sparse, it is hard to have a good understanding of the reservoir characteristics. In this subsection, we would like to relax the prior knowledge. That is, we investigate a case with given facies proportion and mean length, but no prior information about the indicator correlation function. This is the condition when we do not know the indicator variogram or when the facies geometries cannot be characterized by a two-point statistics. The software, SISIM, used in previous subsection is not applicable here because it needs to input the indicator variogram information. Although some other software can be chosen as alternatives, here we do not use

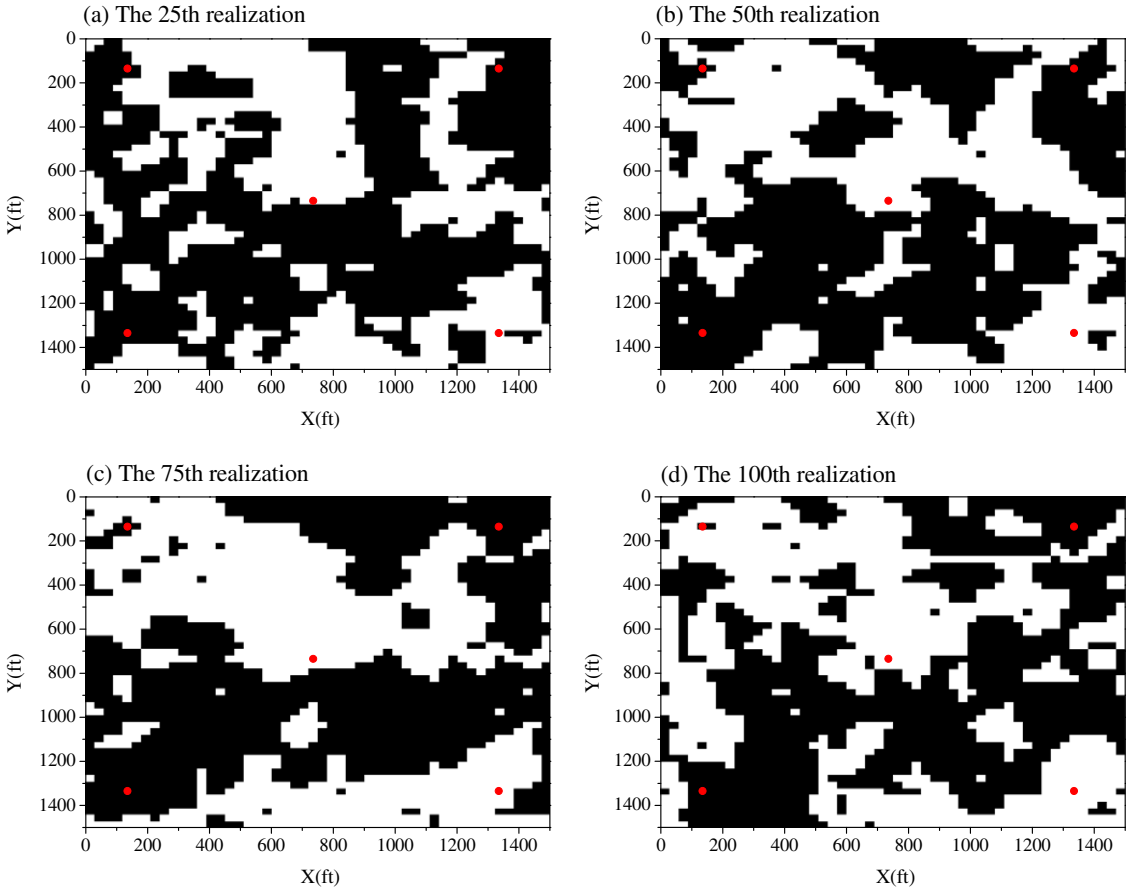

Fig. 8. Four updated realizations with dense initial representing nodes.

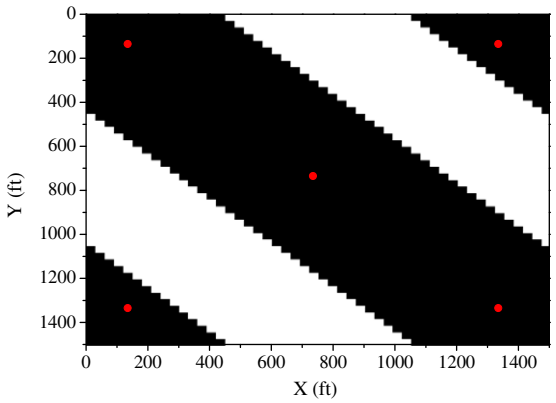

Fig. 9. The reference field for case 2 (white for facies 1 and black for facies 2).

the reconstruction procedure to generate the reference and the initial ensemble. We want to investigate the performance of the proposed method under the condition that the reference and the initial ensemble are obtained from different procedures that will result in large differences between them. About the prior knowledge of this case, the facies mean length in a specific

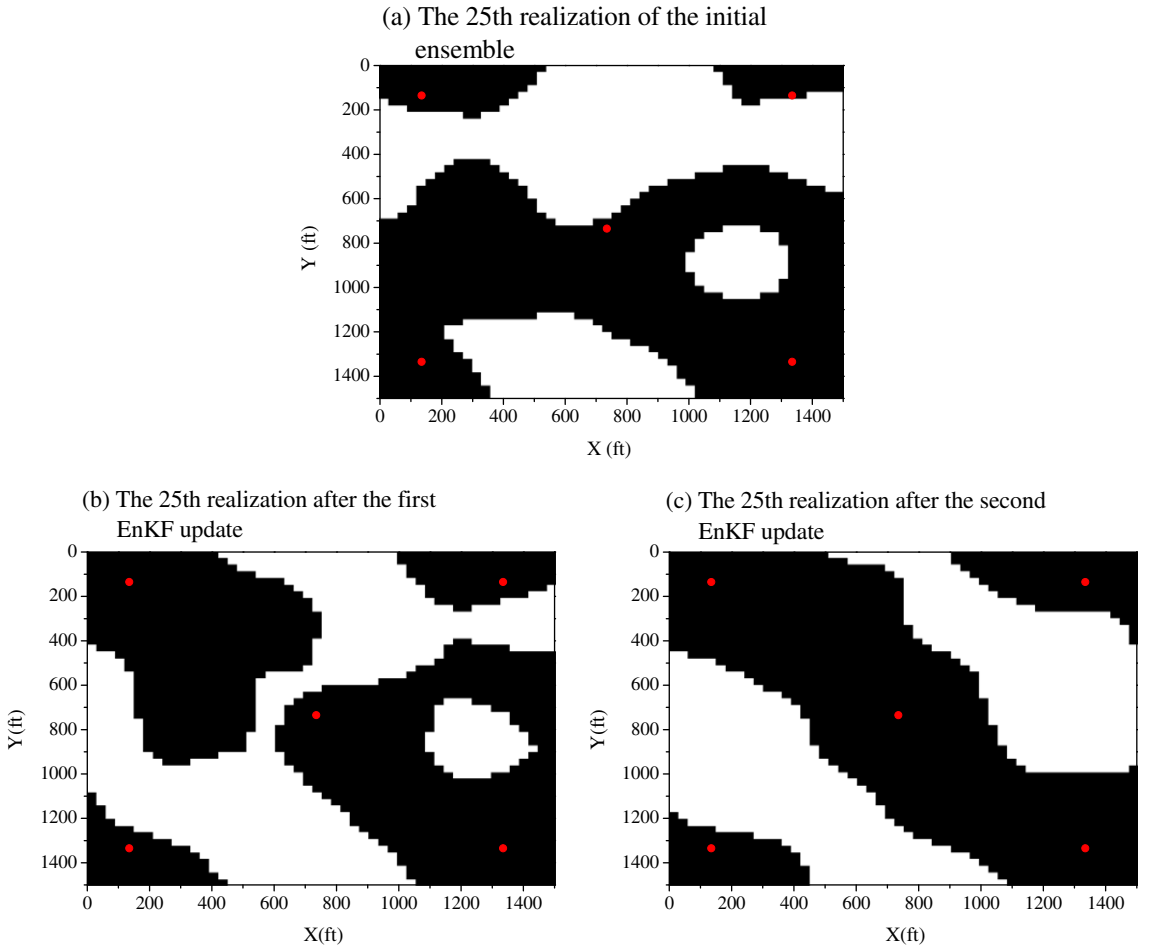

Fig. 10. The 25th realization: (a) initial, (b) after the first EnKF update, and (c) after the second EnKF update.

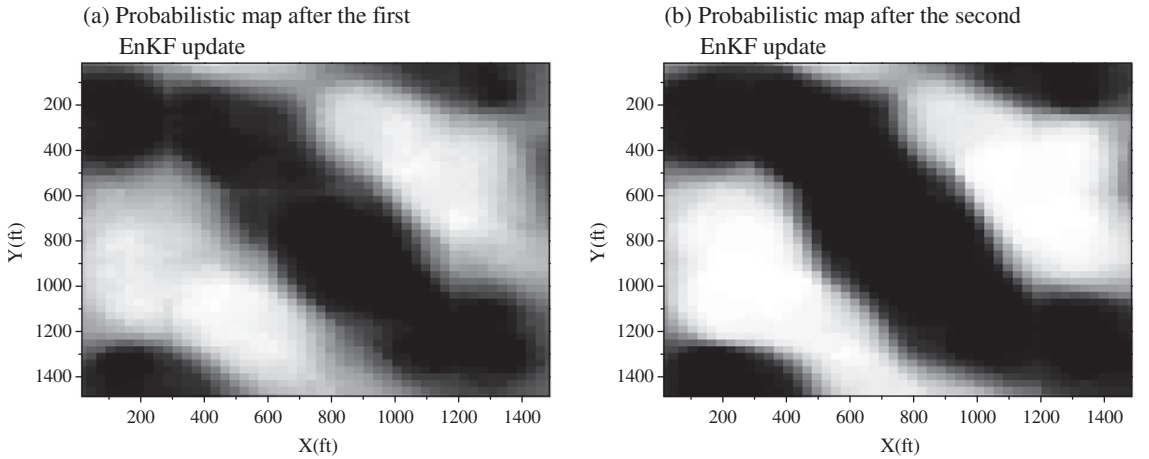

Fig. 11. The probabilistic map for facies 1 computed from all realizations after (a) the first and (b) the second EnKF update.

direction is the average length of each occurrence of a facies, and the facies proportion can be computed from the mean lengths. How to generate the initial ensemble that can honor the prior knowledge is the first issue to be addressed here.

The procedure of generating initial ensemble is similar to that described for the previous case. That is, we need to design the representing node system first and then to generate level set function values on these representing nodes. After calculating the level set function values on the rest grid nodes by linear interpolation, the facies distribution is determined. In this case, we do not reconstruct any existing realizations. Instead, we try to honor the prior knowledge of statistics. We first discuss how to generate the level set function values on the representing nodes following the same proportion as the reference. The probability of each facies occurring at each representing node depends on the mean and standard deviation of the Gaussian random variable. Since we use linear interpolation to get the level set function values on the other grid nodes, the proportions of each facies in the fields depend on the statistics of these Gaussian random variables on the representing nodes. If the proportion is given as prior knowledge, the mean and standard deviation should be designed accordingly.

In Fig. 9, a synthetic facies distribution with proportions of facies 1 and facies 2 being 40.8% and 59.2%, respectively, is illustrated. If a negative level set function value denotes facies 1, we have

$$\begin{cases} P\{\Phi(x_j) < 0\} = 40.8\%, \\ P\{\Phi(x_j) > 0\} = 59.2\% \end{cases} \tag{24}$$

Denote the mean and standard deviation of $\Phi(x_j)$ as μ and σ, respectively. Then we have

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{0} e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt = 0.408 \tag{25}$$

If we set $\sigma = 0.5$, μ can be obtained by interpolation, which is approximately 0.116 for this case.

The second issue is how to design a representing node system to honor the mean length. Because we do not reconstruct any existing realizations, in each direction, the facies mean length, l, only depends on the representing nodes interval, h and $l(h)$ is a monotonic function. We first let $h = h_0$, with this initial choice of h, we generate N set of Gaussian random numbers for the representing nodes and interpolate them to get the facies distribution for each realization. In each direction, the facies mean length of each realization is calculated by statistics and denoted by $l_1(h_0),\ldots,l_N(h_0)$. Let $l_{mean} = \frac{1}{N}\sum_{i=1}^{N} l_i(h_0)$, and denote the mean length of the reference field and the error tolerance as l_{true} and ε, respectively, if $|l_{mean} - l_{true}| < \varepsilon$, then h_0 is the interval value we need. Otherwise, if $l_{mean} >$ (or<) l_{true}, we decrease (or increase) h_0 by Δh step by step. At step n, if we have $|l_{mean}(h_0 - n\Delta h) - l_{true}| < \varepsilon$ (or $|l_{mean}(h_0 + n\Delta h) - l_{true}| < \varepsilon$), then $h_0 - n\Delta h$ (or $h_0 + n\Delta h$) is our choice. Otherwise, continue this procedure until getting $l_{mean}(h_0 - n\Delta h) > l_{true}$ and $l_{mean}(h_0 - (n+1)\Delta h) < l_{true}$ (or $l_{mean}(h_0 + n\Delta h) < l_{true}$ and $l_{mean}(h_0 + (n+1)\Delta h) > l_{true}$). Then we get the region that contains the solution as $[h_0 - (n+1)\Delta h,\ h_0 - n\Delta h]$ (or $[h_0 + n\Delta h,\ h_0 + (n+1)\Delta h]$). In the following, we can only check the midpoint of the region to get a new solution region, which is half of the previous region. With this procedure, according to the convergence criteria, we can refine the solution.

As shown in Fig. 9, the smaller mean length (the mean length of the white shade) is about 2L/5. Following the above algorithm, the representing nodal interval is set to a half of mean length, L/5. We design this reference field purposely, which makes it different from the initial ensemble. In the reference field, these two facies are obviously extended in the NW-SE direction. This fact is assumed not known *a priori* and thus not accounted for when generating the initial ensemble. The 25th initial realization is shown in Fig. 10a, the white shade extends in the x direction, which is different from the reference field. Fig. 10b shows the 25th updated realization, from which we can see that the updated realization is closer to the reference but the result is not satisfactory. In order to get a better result we perform a second EnKF updated. The initial realizations of the second EnKF update are the updated realizations of the first EnKF update and the update process goes from time zero to the end of the simulation. After the second EnKF update, the 25th realization is shown in Fig. 10c, from which we

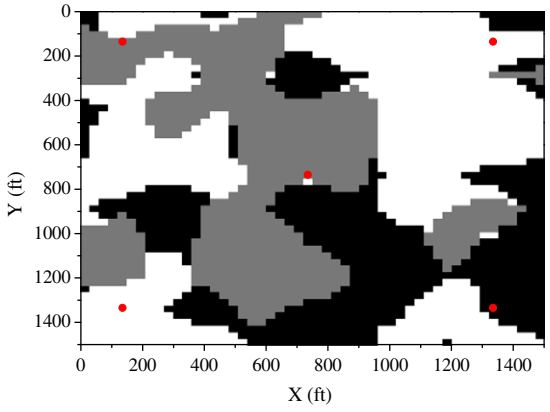

Fig. 12. The reference field for case 3 (white for facies 1, gray for facies 2, and black for facies 3).

can see that the updated realization has a feature very close to the reference. So simply repeating the EnKF update can help to get better result, but it doubles the computational effort at the same time. The probabilistic maps for facies 1 of the updated

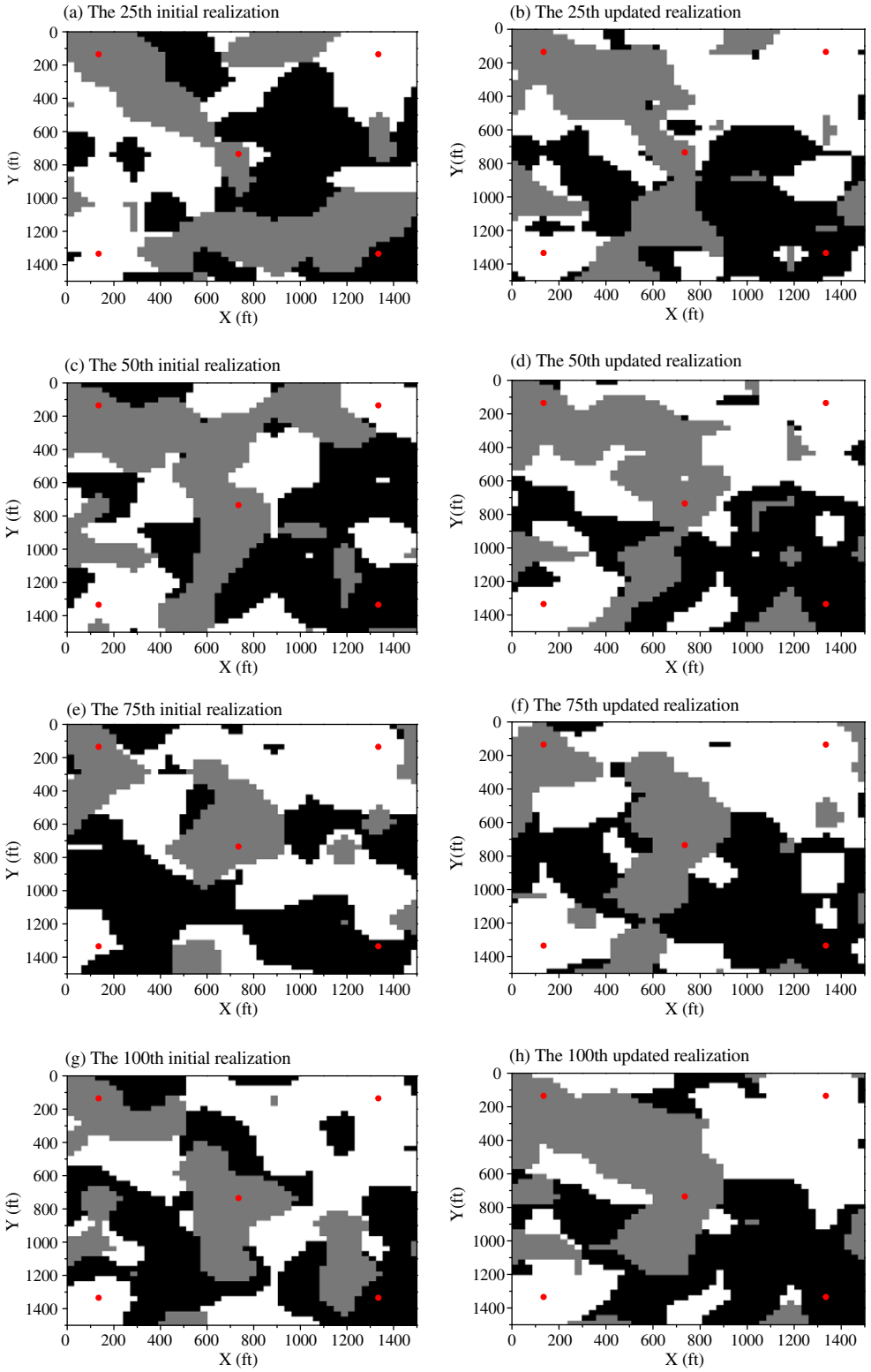

Fig. 13. The left column (a, c, e, and g) shows four initial realizations; the right column (b, d, f, and h) the corresponding realizations after two EnKF updates.

realizations after the first and the second EnKF update are shown in Fig. 11. We can see that, after the first EnKF update the probabilistic map captures the main feature in the true probabilistic map and the second EnKF update helps to improve the result.

6.4. Case 3: Three facies case with given proportion and indicator correlation

The prior knowledge of the facies distribution in this case is the same as in case 1. The difference is that there are three facies, which makes the geometry more complex. The permeability of facies 1, facies 2, and facies 3 are 900 md, 450 md, and

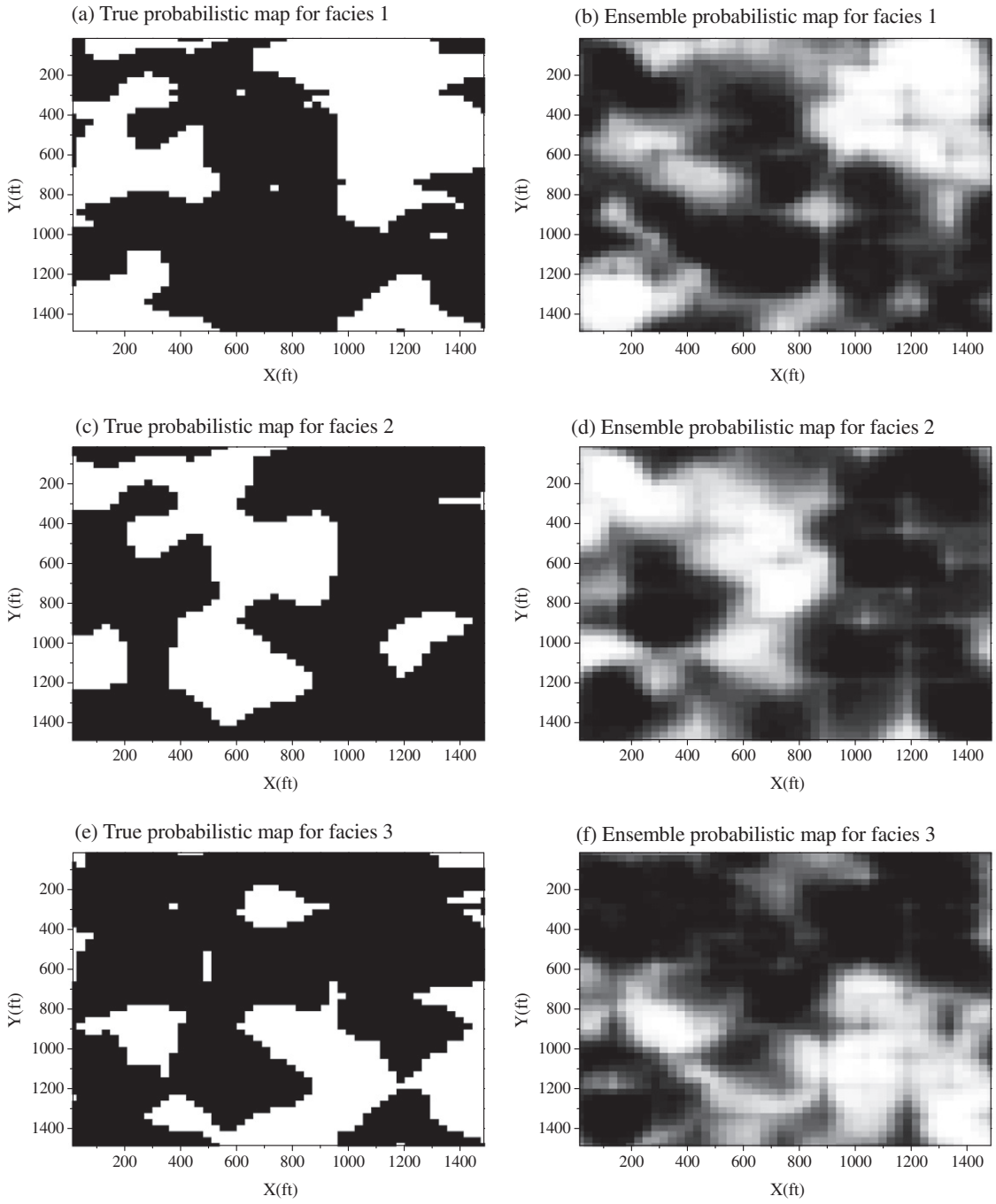

Fig. 14. The true probabilistic maps for facies 1, facies 2, and facies 3 (a, c, and e) and the corresponding probabilistic maps of updated realizations (b, d, and f).

50 md, respectively. The porosity of facies 1, facies 2, and facies 3 are 30%, 22%, and 15%, respectively. Two level set functions, $\Phi_1(x)$ and $\Phi_2(x)$, are needed to describe the facies distribution with the following definition,

$$\begin{cases} \Phi_1(x) > 0, \ \Phi_2(x) > 0, & \text{if facies 1 occurs at location } x; \\ \Phi_1(x) > 0, \ \Phi_2(x) < 0, & \text{if facies 2 occurs at location } x; \\ \Phi_1(x) < 0, & \text{if facies 3 occurs at location } x. \end{cases} \tag{26}$$

We use the same reconstruction procedure to obtain the reference and initial ensemble as described in case 1. The representing nodes are selected to include all well location, which simplifies the iterative enforcement of the facies constraint. The reference of this case is shown in Fig. 12. All the prior knowledge can be honored using the reconstruction procedure. In the data assimilation step, the state vector takes the following form,

$$y_j = \left[\{\Phi_1(x_j)\}^T, \{\Phi_2(x_j)\}^T, P^T, S_w^T, d^T\right]_j^T \tag{27}$$

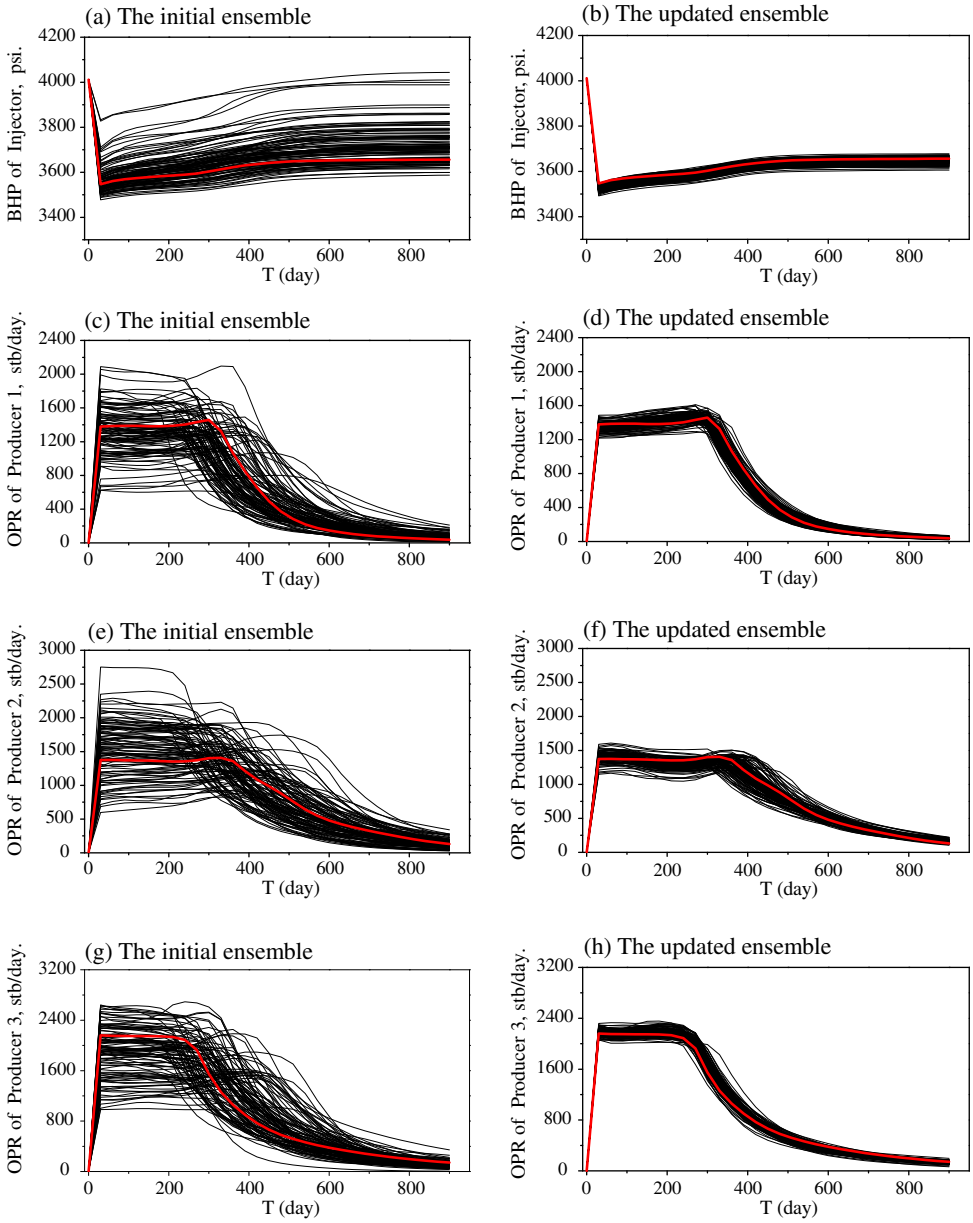

Fig. 15. The match to the production data from the initial ensemble (left column) and the updated ensemble (right column).

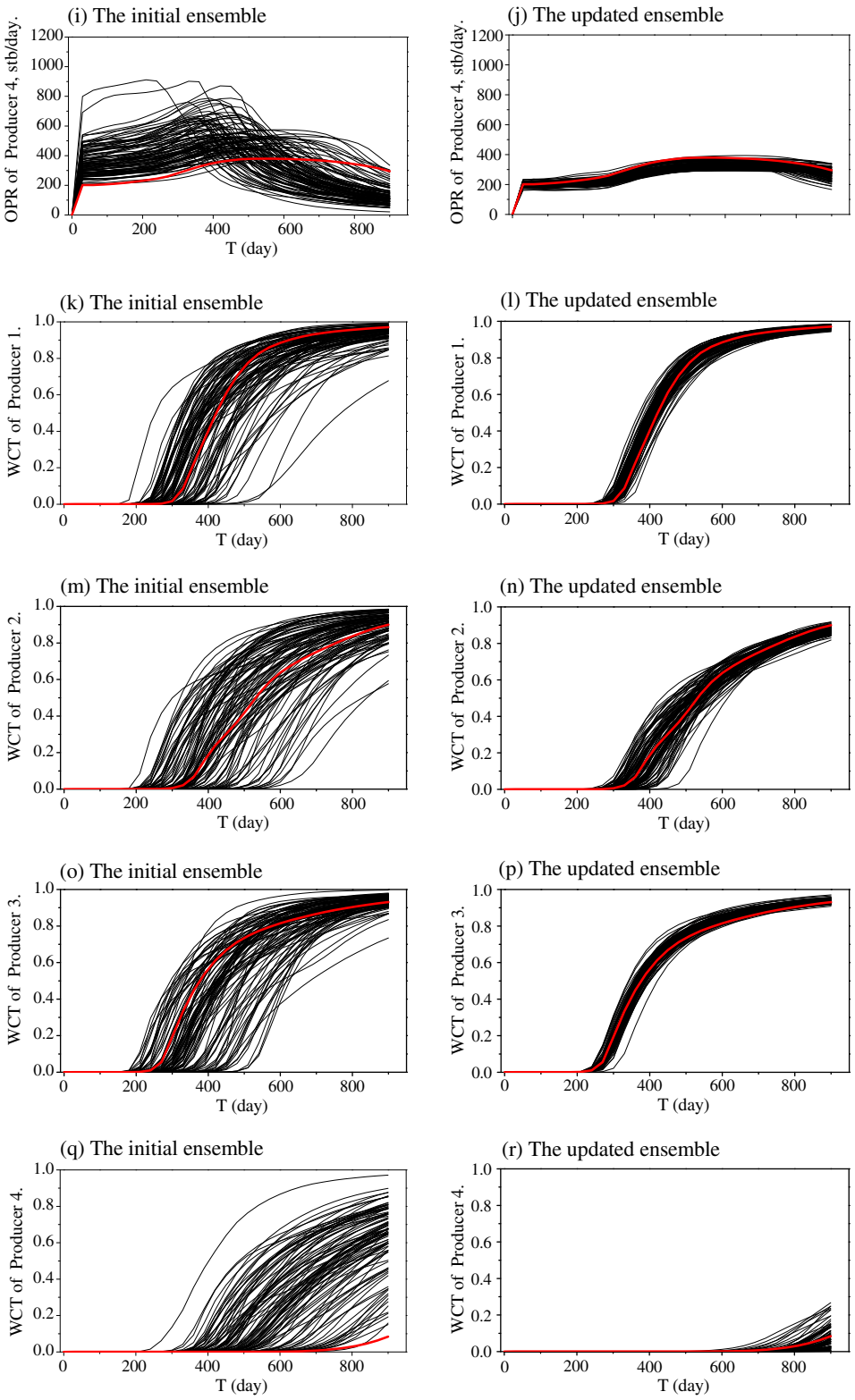

Fig. 15 (*continued*)

Two level set function values on the representing nodes are updated at each data assimilation step. The sensitivity between each level set function and the observations will not be as strong as in the two facies case, which increases the difficulty in the updating process. Also the increased number of facies presented in a reservoir will aggravate the non-unique relationship between the facies distribution and the reservoir production behavior.

Fig. 13 shows four initial realizations (the left column) and their corresponding updated realizations (the right column) after two EnKF updates. It is seen that these initial realizations are conditioned on the facies type observations and thus have the same facies types at the well locations as in the reference field. Besides the well locations, there are not many features in common, reflecting a large variability of the initial realizations. After assimilating the production data, these realizations have some similar features and look more like the reference field. Fig. 14 shows the true probabilistic maps for facies 1, facies 2, and facies 3 and the corresponding probabilistic maps computed from updated realizations. It is seen that the probabilistic maps of updated realizations captures the main features of the true maps. The match to the production data from the initial ensemble and the updated ensemble are shown in Fig. 15. It can be seen that the production profiles from the initial ensemble have large variability, and such variability is significantly reduced after assimilating the production data.

7. Conclusions

In this study, we developed an approach for combining the EnKF and level set parameterization for history matching of the facies distribution. With the prior knowledge about the facies distribution, the initial realizations can be generated using existing software, and level set functions are used to reparameterize these realizations with a properly designed representing node system. In the EnKF update process, the level set function values on the representing nodes are updated to adjust the facies distributions. In the case studies, we investigated the proposed methodology for a two-dimensional reservoir with two or three facies. The choice of the representing nodes is important for our method. We discussed how to design a proper representing node system in the reconstruction procedure. The numerical results show that the production data can be matched well and the main features of the reference fields can be captured by the realizations. Simply repeating the EnKF update process is showed to be able to help further reduce the variability of the model parameters.

Acknowledgements

This work is partially funded by National Science Foundation through grant OCI-0904754, National Natural Science Foundation of China through grant 50688901, and the Chinese National Basic Research Program through grant 2006CB705800. The first author would also like to acknowledge the support from China Scholarship Council through grant 2007100458.

References

[1] C.C. Agbalaka, D.S. Oliver, Application of the EnKF and localization to automatic history matching of facies distribution and production data, Mathematical Geosciences 40 (4) (2008) 353–374.
[2] I. Berre, M. Lien, T. Mannseth, A level-set corrector to an adaptive multiscale permeability prediction, Computational Geosciences 11 (1) (2007) 27–41.
[3] S.F. Carle, T-PROGS: Transition Probability Geostatistical Software, University of California, Davis, CA, 1999.
[4] C.V. Deutsch, A.G. Journel, GSLIB Geostatistical Software Library and User's Guide, Oxford University Press, New York, NY (EUA), 1998.
[5] L. Dovera, E. Della Rossa, Ensemble Kalman filter for gaussian mixture models, in: Petroleum Geostatistics, 10–14 September 2007, Cascais, Portugal, A16, Extended Abstracts Book, EAGE Publications BV, Utrecht, The Netherlands.
[6] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Research 99 (C5) (1994) 10143–10162.
[7] G. Evensen, The ensemble Kalman filter: theoretical formulation and practical implementation, Ocean Dynamics 53 (2003) 343–367.
[8] A. Galli, H. Beucher, G. Le Loc'h, B. Doligez, H. Group, The Pros and Cons of The Truncated Gaussian Method, Geostatistical Simulations, Kluwer Academic, Dordrecht, 1994. pp. 217–233.
[9] P.L. Houtekamer, H.L. Mitchell, Data assimilation using an ensemble Kalman filter technique, Monthly Weather Review 126 (3) (1998) 796–811.
[10] B. Jafarpour, D.B. McLaughlin, History matching with and ensemble Kalman filter and discrete cosine parameterization, in: Paper SPE 108761 presented at the SPE Annual Technical Conference and Exhibition, Anaheim, California, USA, 11–14 November 2007, doi:10.2118/108761-MS.
[11] J.J. Gómez-Hernánez, A. Sahuquillo, J.E. Capilla, Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data: I. Theory, Journal of Hydrology 203 (1–4, 31) (1997) 162–174.
[12] G. Le Loc'h, A. Galli, Truncated plurigaussian method: Theoretical and practical points of view, in: E.Y. Baafi, N.A. Schofield (Eds.), Geostatistics wollongong '96, vol. 1, Kluwer Academic, Dordrecht, 1997, pp. 211–222.
[13] N. Liu, D.S. Oliver, Ensemble Kalman filter for automatic history matching of geologic facies, Journal of Petroleum Science and Engineering 47 (2005) 147–161.
[14] Z. Lu, B.A. Robinson, Parameter identification using the level set method, Geophysical Research Letters 33 (2006) L06404, doi:10.1029/2005GL025541.
[15] Z. Lu, D. Zhang, On stochastic modeling of flow in multimodal heterogeneous formations, Water Resources Research 38 (10) (2002) 1190, doi:10.1029/2001WR001026.
[16] D. Moreno, S.I. Aanonsen, Stochastic facies modeling using the level set method, in: Petroleum Geostatistics, 10–14 September 2007, Cascais, Portugal, A16, Extended Abstracts Book, EAGE Publications BV, Utrecht, The Netherlands.
[17] L.K. Nielsen, H. Li, X. Tai, S.I. Aanonsen, M. Espedal, Reservoir description using a binary level set model, Computing and Visualization in Science (2008), doi:10.1007/s00791-008-0121-1.
[18] S. Osher, F. Santosa, Level set methods for optimization problems involving geometry and constraints I. Frequencies of a two-density inhomogeneous drum, Journal of Computational Physics 171 (2001) 272–288.
[19] S. Osher, J.A. Sethian, Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations, Journal of Computational Physics 79 (1) (1988) 12–49.
[20] S. Strebelle, Conditional simulation of complex geological structures using multiple-point statistics, Mathematical Geology 34 (1) (2002).

[21] R. Villegas, O. Dorn, M. Moscoso, M. Kindelan, Reservoir characterization using stochastic initializations and the level set method, Computers and Mathematics with Applications 56 (2008) 697–708.
[22] X. Wen, C.V. Deutsch, A.S. Cullick, High-resolution reservoir models integrating multiple-well production data, SPEJ 3 (4) (1998) 344–355.
[23] Y. Zhao, A.C. Reynolds, G. Li, Generating facies maps by assimilating production data and seismic data with the ensemble Kalman filter, in: Paper SPE 113990 presented at the SPE/DOE Symposium on Improved Oil Recovery, Tulsa, April 20–23, 2008, doi:10.2118/113990-MS.