CSUG/SPE 137480

Application of the Ensemble Kalman Filter for Characterization and History Matching of Unconventional Oil Reservoirs

S.B. Chitralekha, SPE, J.J. Trivedi, SPE, S.L. Shah, University of Alberta

This paper was prepared for presentation at the Canadian Unconventional Resources & International Petroleum Conference held in Calgary, Alberta, Canada, 19–21 October 2010.

Abstract

Recently, the Ensemble Kalman Filter (EnKF) has emerged as an effective tool for performing continuous updating of petroleum reservoir simulation models. The method is firmly grounded on the theory of Kalman filters and sequential Monte Carlo techniques. The ability of the method to sequentially update the spatial properties in petroleum reservoir models, such as permeability and porosity, by integrating the dynamic production data makes it a very attractive approach. Moreover, the method takes into account the production uncertainty in the reservoir models by using error covariance matrices for the measurement vector (Production and injection rates, Gas-Oil ratio, Steam-Oil ratio, etc.) and the state vector (pressure, saturation, permeability, porosity). Similar to the traditional Kalman filter, the covariance matrices have to be tuned to reflect the uncertainty in the model and the measurements. We consider two unconventional oil reservoir models: 1) highly heterogeneous black-oil reservoir model, and 2) heterogeneous SAGD reservoir model. The results will demonstrate the advantage of using the localized EnKF for effective history matching using ensemble sizes relatively lower than what otherwise would be required with the ordinary EnKF. The results will also show the advantages of using prior knowledge available from the wells (permeability and porosity measurements) to generate initial realizations. One of the main practical advantages of history matching using the EnKF over traditional optimization based approaches is its low computational effort. The computational cost is dominated by Monte Carlo simulation of the ensemble of models only. Thus, significant computational time saving is possible by running each of the ensemble simulations on independent processors in a parallel mode. Moreover, the method can be easily integrated with any commercial reservoir simulation software.

1. Introduction

History matching is the art of calibrating reservoir simulation models to past production data available from the field. From a systems engineering point of view, this can be defined as identifying models that are consistent to data collected from the process. In the context of petroleum engineering, the process of interest is the flow of multi-component fluids (mainly mixture of oil, gas and water) through the heterogeneous geological layers of the reservoir. The reservoir simulation models are a set of highly nonlinear partial differential equations in space and time. Typically, the differential equations are functions of several model parameters - this will include the spatially varying rock properties (permeability, porosity), initial conditions of the reservoir, relative permeability curves and other physicochemical properties of the reservoir strata and fluids. We collectively call these properties of the reservoir as model parameters. Thanks to the modeling effort of reservoir engineers and physicists, today there are several sophisticated reservoir simulation packages available off-the-shelf. These packages can integrate the complex physics along with efficient numerical algorithms and simulate the reservoir, given the values of all the model parameters in the differential equations. From the solutions of the differential equations, we can obtain simulated production data. The predictions that we obtain from the simulations will depend on these model parameters. If the objective of modeling the reservoir is to get reliable predictions of the future performance of the reservoir, then it is required to have a good knowledge of the values of these model parameters. However, in a practical scenario many of these model parameters will be uncertain or completely unknown because of our inability to directly measure them in the true reservoir. It is a well accepted fact that all reservoirs are

heterogeneous with different degrees of heterogeneity exhibited by different properties. For example, the geological properties such as the permeability and porosity can vary in a spatial scale of a few meters. However, these properties may only be known at around 1% of the reservoir volume through core samples taken at well locations Romeu 2010. The geological uncertainty is inevitable in any reservoir simulation based decision making exercise McLennan and Deutsch 2006. Through the use of sophisticated geostatistical algorithms, such as GSLIB Deutsch and Journel 1998, SGeMS Remy et al. 2009 etc., we can capture the spatial variation to some extent. However, the exact geological heterogeneity is never known and can only be approximately predicted in between wells through these stochastic realizations McLennan and Deutsch 2010. The geological uncertainty is usually represented by generating multiple stochastic realizations which honor the spatial variogram models. The traditional way of validating a particular geological realization is to manually compare the corresponding flow simulations with the production data from the true reservoir.

History matching is essentially an iterative model validation procedure - given some historical production data as the target, the objective is to find a set of model parameters which when plugged into the model equations will result in a simulated production data close to the historical production data of the wells. In other words, it is an iterative task of tuning the model parameters which will yield simulation results that are consistent with production history of the wells. The production data will include measurements that are routinely available from the wells such as oil/gas production rates, steam-to-oil ratios, well water cuts etc. It is well known that such production variables are sensitive to the geological heterogeneity. Traditionally, history matching was done by manually adjusting the model parameters directly or indirectly through the geostatistical models. Evidently, this is a trial-and-error approach and is known to be too time-consuming and ineffective Romeu 2010. More importantly, the problem is known to be underdetermined and hence there will exist multiple plausible solutions Oliver et al. 2008. A single history matched model will be biased towards a few model parameters. Hence, modern history matching techniques are aimed at arriving at multiple set of history matched realizations. Such solutions will satisfy the needs of decision making in the face of uncertainty in the production forecasts. Moreover, the advancement in sensor technologies have resulted in the use of permanent sensors which can provide reservoir production data more frequently than that was possible with the legacy sensors. There is a growing need to develop fast history matching algorithms that can continuously update the reservoir models by integrating the most recent information available through such modern day sensors (Seiler et al. 2009).

Several approaches have been proposed in the literature with a focus on the modern day needs of continuous history matching of reservoir simulation models in the face of uncertainty. These techniques broadly fall into two categories: stochastic global optimization methods (Schulze-Riegert et al. 2001) and Ensemble Kalman filter based methods (Lorentzen et al. 2001, Evensen 2007, Aanonsen et al. 2009). The stochastic optimization methods are based on the idea of minimizing an objective function which quantifies the mismatch/error between model predictions and the observed data. The objective function is usually chosen as a quadratic function, such as the sum of square of the error (SSE). Due to the highly nonlinear nature of the reservoir simulation models and the large number of optimization variables, gradient-free evolutionary optimization algorithms are used. The numerical cost associated with such optimization algorithms become unacceptably high because of the large number of forward simulations required to iteratively converge to the optimal solution, especially when there are large number of unknown parameters. Modern distributed computing technologies are employed to arrive at a solution within reasonable time frames for decision making in a history matching project (Schulze-Riegert and Ghedan 2007).

Recently, the Ensemble Kalman filter (EnKF) has emerged as a very promising alternative because of its non-iterative nature and requirement of only *O(50)* forward simulations for estimating model parameters in real reservoir case studies (Evensen et al. 2007, Haugen et al. 2008, Zhang and Oliver 2009). The method is based on the idea of making probabilistic inference using the classical Bayes' theorem. It is based on the classical Kalman filter which was proposed in the 1960's for estimating unmeasured variables in linear systems using a model and data collected from the real process Kalman 1960. The EnKF is an adaptation of the Kalman filter for handling systems characterized by nonlinear models. The history matching step in the algorithm is a one-step linear update rule proposed in the Kalman filter and does not require any iterative solution. The history matching step is given by the following equation:

$$\boldsymbol{\theta}_{k|k} = \boldsymbol{\theta}_{k-1|k-1} + \mathbf{K}_{gain}(\mathbf{y}_k^{obs} - \mathbf{y}_k^{pred}) \qquad (1)$$

where, $\boldsymbol{\theta}_{.|.}$ is the vector of unknown parameters, $\mathbf{y}_k^{obs}$ is the vector of newly observed production data and $\mathbf{y}_k^{pred}$ is the vector of predicted reservoir production based on the most recently history matched models. The subscript k indicates the update time index and '$k|k$' means the value of the parameter at k^{th} update time index given all the measurements until the k^{th} time index. The vector $\mathbf{K}_{gain}$ is the Kalman gain proposed by Rudolph Kalman Kalman 1960, the sole variable to be calculated, whose calculation is adapted for nonlinear systems in the EnKF. As seen in Eq. (1), the update magnitude is proportional to the magnitude of error between the predictions and measured data. The Kalman gain applies a weighted updating to the unknown parameters in an 'intelligent fashion', and depends on the degree of uncertainty in the models and the measurements. Based on Eq. (1), we can see another advantage of the method in that it is sequential in nature, i.e., the history matching step depends only on the new

measurements and the most recently history matched set of models. This makes this method well suited for continuous updating of reservoir simulation models. More details about the EnKF algorithm steps will be discussed in the next section.

Most of the applications of EnKF in the current petroleum engineering literature have concentrated on conventional black-oil reservoir models. In this article, we investigate the application of the EnKF algorithm for history matching of unconventional reservoir models. We present two simulated case studies: (1) a highly heterogeneous black-oil reservoir model (2-dimensional), and (2) a heterogeneous SAGD reservoir simulation model (50x10x5). We consider the heterogeneous grid block permeabilities as the unknown model variables in the history matching problem. In both cases, a set of realistically placed well-core measurements combined with a geostatistical variogram model is used to generate the initial ensembles for the EnKF algorithm. These application studies will show that using a set of geostatistical realizations as the initial ensemble is an important step to obtain realistic results when the heterogeneity information is available a priori. Otherwise, since the history matching problem is an under determined problem, there is a high chance for the ensembles to move into domains that are unrealistic. This is because the solution space of the EnKF algorithm is limited by the span of the ensembles (Evensen 2009).

This paper is organized as follows. In Section 2, we explain the basic methodology of the Ensemble Kalman filter algorithm as applied to history matching. Some modifications to the basic algorithm, based on previous work in the literature that we have employed in this work, are briefly discussed. Some insights on the critical tuning factors which can significantly affect the quality of the results are provided. Two quality measures for measuring the efficacy of the EnKF algorithm in any history matching application are proposed. In Section 3, we present two synthetic case studies on the application of the methodology for history matching and characterizing heterogeneous reservoirs, followed by some concluding remarks in Section 4.

2. Methodology: Ensemble Kalman Filter

The Ensemble Kalman filter was first proposed as a novel method for estimating unmeasured variables, techniques more widely known as data assimilation algorithms, in highly nonlinear ocean models (Evensen 1994). Numerous applications of this method for nonlinear state estimation, especially for large scale systems, have been shown in the oceanography and meteorology literature (Evensen 2003). The methodology was introduced as an alternative method for history matching in petroleum engineering by Lorentzen et al. 2001. Here, we briefly explain the basic formulation and a few modifications proposed in earlier works in the literature. The EnKF is a Monte Carlo simulation based variant of the Kalman filter designed for handling nonlinear systems. In order to explain the methodology, consider the following stochastic model of the reservoir:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \boldsymbol{\theta}) \qquad (2)$$

$$\begin{aligned} \mathbf{y}_k &= \mathbf{g}(\mathbf{x}_k) + \mathbf{v}_k \\ \mathbf{v}_k &\sim N(\mathbf{0}, \mathbf{R}) \end{aligned} \qquad (3)$$

where, $\mathbf{x}_k$ denotes the vector of variables which define the *reservoir state* at the k^{th} instant. This will include the dynamic variables which directly represent the internal conditions at all the grid block locations of the reservoir such as pressure, saturations, temperature. The function **f** is the *state transition function* which will relate the current state of the reservoir to the previous state. The partial differential equations which define the dynamics of the reservoir simulation can be considered to be represented by **f**, i.e.,

$$\mathbf{f} \; \square \int_{k-1}^{k} \dot{\mathbf{x}} dt$$

In real reservoirs, the variables in $\mathbf{x}_k$ are not measured directly since it is not possible to sample/place online sensor devices at all the grid block locations of the true reservoir. However, all these variables will have an effect on the *production data* that we measure from the reservoir. The variable $\mathbf{y}_k$ represents all these dynamic production variables that can be measured at well locations such as oil/gas/water flow rates, monthly/cumulative production, gas-oil ratio, steam-oil ratio etc. The production data that we obtain from the real reservoir will depend on the reservoir states through the *measurement function* **g** and will be contaminated with some sensor noise $\mathbf{v}_k$. We assume that $\mathbf{v}_k$ follows a zero-mean Gaussian distribution with covariance matrix **R**. The variable $\boldsymbol{\theta}$ is the vector of all the static model parameters which are unknown such as permeability and porosity (3-D grids unfolded into a 1-D vector). Since we are uncertain about the true value of $\boldsymbol{\theta}$, we consider them to be a random variable following some distribution.

Both **f** and **g** are part of any commercial reservoir simulator which can be considered as a 'black-box' in the EnKF algorithm (Wen and Chen 2007). In the EnKF algorithm it is not required to explicitly know any of the functions **f** and **g**. It is only required to obtain the variables $\mathbf{x}_k$ and $\mathbf{y}_k$ through a forward simulation of the reservoir starting from the initial conditions $\mathbf{x}_{k-1}$. The functions **f** and **g** are usually highly nonlinear due to the coupled nature of the partial differential equations and the nonlinear mass balance correlations (relating the well production data to the grid block saturations). Note that one of the basic requirements for the Kalman filter computations is that the state transition function and the measurement function be linear functions. Also, the

uncertainty in the reservoir parameters and the measurement noise are required to follow a Gaussian distribution. We can relax these requirements in the EnKF computations. The main idea is to use a bank of Kalman filters on an *ensemble* of Monte Carlo simulations of the reservoir. The sample covariance/cross-covariance of the simulated ensemble of $\mathbf{x}_k$ and $\mathbf{y}_k$ are used in the calculation of the Kalman gain in the EnKF algorithm.

In the presence of uncertain model parameters and unmeasured state variables, the history matching problem will have to be considered as a combined state and parameter estimation problem using measured data. For this, we augment the state vector with the uncertain parameter vector as follows:

$$\mathbf{z}_k \equiv \begin{bmatrix} \mathbf{x}_k \\ \boldsymbol{\theta}_k \end{bmatrix} = \begin{bmatrix} \mathbf{f}(\mathbf{x}_{k-1}, \boldsymbol{\theta}_{k-1}) \\ \boldsymbol{\theta}_{k-1} \end{bmatrix} \qquad (4)$$

where the augmented state vector $\mathbf{z}_k$ has to be estimated using the measurements. Note that the augmented state vector will contain both the dynamic and static variables. For the static variables, it is a common practice to assume some slow artificial dynamics in the form of a random walk/Brownian motion (Moradkhani et al. 2005)

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \mathbf{w}_{k-1} \qquad (5)$$

where $\mathbf{w}_{k-1}$ is a zero-mean Gaussian random variable with a very low variance $\mathbf{Q}$. A similar additive Brownian random noise is usually assumed for the state transition function, to account for modeling errors also (Burgers et al. 1998, Evensen 2009). However, in this work we assume that the modeling error is captured through the additive error in the parameter vector. Based on the augmented state vector $\mathbf{z}_k$, with the artificial parameter dynamics defined in Eq. (5), the EnKF algorithm will be explained.

As mentioned earlier, the EnKF will use an ensemble of Monte Carlo samples to represent the uncertain variables. Assume that at time $k = 0$, we have a multivariate probability distribution $p(\theta)$ which represents the prior knowledge about the parameter vector. In reservoir engineering, this distribution will be defined by the geostatistical models which define the geological uncertainty, i.e.,

$$p(\boldsymbol{\theta}) \equiv \text{Prior Geological Uncertainty}$$

The first step in the EnKF is to generate N_e samples from the prior probability distribution $p(\theta)$. Let $\boldsymbol{\theta}_{0|0} = \{\theta^1_{0|0}, \theta^2_{0|0}, \ldots\ldots\ldots\ldots \theta^{Ne}_{0|0}\}$ represent the set of N_e samples of the parameters. Here we assume that the initial conditions of the reservoir is accurately known and given by $\mathbf{x}_0$. If the initial conditions are also uncertain, then a *set* of initial conditions representing this uncertainty should also be generated. Let the production data from the field at the date instance k be represented as y_k^{obs}. Then, the EnKF algorithm sequentially updates the reservoir as follows:

1. Prediction step:

The prediction step is a forward simulation of the reservoir using each of the N_e parameter samples. We represent the predicted ensemble as $\hat{Z}_{k|k-1} \equiv \{z^1_{k|k-1}, z^2_{k|k-1}, \ldots\ldots\ldots\ldots, z^{N_e}_{k|k-1}\}$, $z^i_{k|k-1}$ is generated using Eq. (4), i.e., the reservoir simulator $\mathbf{f}$ with the parameter value defined by the values in $\theta^i_{k-1|k-1}$. Thus, we get

$$\mathbf{z}^i_{k|k-1} = \begin{bmatrix} \mathbf{f}(\mathbf{x}^i_{k-1|k-1}, \boldsymbol{\theta}^i_{k-1|k-1}) \\ \boldsymbol{\theta}^i_{k-1|k-1} + \mathbf{w}^i_{k-1} \end{bmatrix} \qquad (6)$$

Let,

$$\boldsymbol{\mu}^z_{k|k-1} = \frac{1}{N_e} \sum_{i=1}^{N_e} z^i_{k|k-1} \qquad (7)$$

be the mean of the predicted ensemble. The notation 'k|k-1' is used to indicate that we have used the information from the previous time step (k-1) to obtain the current value of the variable. In the EnKF algorithm, the error/uncertainty in the predicted states of the reservoir is defined around the predicted state ensemble mean as

$$e^i_{k|k-1} = \mathbf{z}^i_{k|k-1} - \boldsymbol{\mu}^z_{k|k-1} \qquad (8)$$

Note that for each $\mathbf{z}^i_{k|k-1}$, we will simultaneously obtain an ensemble of *simulated production data* in any commercial reservoir simulator. Let $\hat{Y}_{k|k-1} \equiv \{y^1_{k|k-1}, y^2_{k|k-1}, \ldots\ldots\ldots\ldots, y^{N_e}_{k|k-1}\}$ be the ensemble of predicted production data. Similar to Eq. (8), the error/uncertainty in the predicted measurements is defined around the predicted measurement ensemble mean as

$$\varepsilon^i_{k|k-1} = y^i_{k|k-1} - \boldsymbol{\mu}^y_{k|k-1} \qquad (9)$$

where,

$$\boldsymbol{\mu}^{y}_{k|k-1} = \frac{1}{N_e}\sum_{i=1}^{N_e} y^{i}_{k|k-1} \tag{10}$$

The uncertainty in the predicted states and measurements are calculated as sample covariance of the error ensemble generated in equations (8) and (9). Two covariance matrices are calculated:

$$\hat{\mathbf{P}}^{\varepsilon,\varepsilon}_{k|k-1} = \frac{1}{N_e - 1}\sum_{i=1}^{N_e} (\varepsilon^{i}_{k|k-1})(\varepsilon^{i}_{k|k-1})^{T} \tag{11}$$

$$\hat{\mathbf{P}}^{e,\varepsilon}_{k|k-1} = \frac{1}{N_e - 1}\sum_{i=1}^{N_e} (e^{i}_{k|k-1})(\varepsilon^{i}_{k|k-1})^{T} \tag{12}$$

The covariance $\hat{\mathbf{P}}^{e,\varepsilon}_{k|k-1}$ is the cross-covariance between the state prediction ensemble and measurement model ensemble. In other words, it is measure of how strongly correlated is the production data with respect to the uncertainty in the reservoir states/parameters. The main assumption in the EnKF is that the distribution of the state prediction error and measurement error is zero-mean multivariate Gaussian.

2. Correction step:

The correction step is the sequential history matching step in the EnKF algorithm. In the correction step, each of the samples in the predicted ensemble is updated using the production measurement y^{obs}_{k} from the actual reservoir. While using the production data, it is required to treat them as random variable contaminated by noise (Burgers et al. 1998). Hence, a set of N_e perturbed measurements around y^{obs}_{k} are generated using random samples from the measurement noise distribution defined in Eq. (3) ($\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{R})$) as follows

$$\{ y^{i,obs}_{k} = y^{obs}_{k} + v^{i}_{k} \}_{\text{for } i=1 \text{to } N_e} \tag{13}$$

The perturbed observations are used to update the reservoir states and the unknown parameters using the well-known Kalman update equation

$$\{\ \mathbf{z}^{i}_{k|k} = \mathbf{z}^{i}_{k|k-1} + \mathbf{K}_{gain}(\mathbf{y}^{i,obs}_{k} - \mathbf{y}^{i}_{k|k-1})\ \}_{\text{for } i=1 \text{to } N_e} \tag{14}$$

Note that we use the subscript notation 'k|k' to indicate that the information from observations at the current time instant k is assimilated into the updated variable. The Kalman gain $\mathbf{K}_{gain}$ is calculated using the sample covariance matrices in Eqs. (11) and (12) as follows

$$\mathbf{K}_{gain} = \hat{\mathbf{P}}^{e,\varepsilon}_{k|k-1}(\hat{\mathbf{P}}^{\varepsilon,\varepsilon}_{k|k-1} + \mathbf{R})^{-1} \tag{15}$$

The Kalman update is applied to each of the predicted state ensemble to obtain the *corrected ensemble*, i.e., the history matched ensemble of reservoir states and parameters represented as $\hat{Z}_{k|k} \equiv \{\mathbf{z}^{1}_{k|k}, \mathbf{z}^{2}_{k|k}, ..., \mathbf{z}^{N_e}_{k|k}\}$. The history matched ensemble so obtained will be propagated forward through the prediction step until the next history matching date, then the update procedure is repeated, and so on. This sequential prediction and correction procedure is continued and stopped when the end of the period selected for history matching is reached. To our knowledge, selection of the optimal stopping point for the sequential history matching problem is not investigated in the current literature. A heuristic method, though computationally expensive, would be to evaluate the predictive accuracy of the reservoir model by restarting the simulation from $k = 0$ using the most currently updated parameter ensemble.

Note that in the update step, the EnKF adjusts the predicted ensemble by a value proportional to the error in the predicted measurement $(\mathbf{y}^{i,obs}_{k} - \mathbf{y}^{i}_{k|k-1})$, i.e.,

$$(\mathbf{z}^{i}_{k|k} - \mathbf{z}^{i}_{k|k-1}) = \mathbf{K}_{gain}(\mathbf{y}^{i,obs}_{k} - \mathbf{y}^{i}_{k|k-1}) \tag{16}$$

The Kalman gain applies a weighted correction where the weighting is roughly given by the ratio of $\hat{\mathbf{P}}^{e,\varepsilon}_{k|k-1}$ to $(\hat{\mathbf{P}}^{\varepsilon,\varepsilon}_{k|k-1} + \mathbf{R})$. If $\hat{\mathbf{P}}^{e,\varepsilon}_{k|k-1}$ is relatively high compared to $(\hat{\mathbf{P}}^{\varepsilon,\varepsilon}_{k|k-1} + \mathbf{R})$, it would signify that (a) there is a high correlation between the measurement error and the estimation error, and (b) the measurement uncertainty is low. In that case, the Kalman gain will be high and the

correction made to the reservoir states/parameters will be very sensitive to $(\mathbf{y}_k^{i,obs} - \mathbf{y}_{k|k-1}^{i})$. Thus, the Kalman gain in the EnKF can be interpreted similar to that of the classical Kalman filter for linear systems (Kalman 1960). The prediction and correction step is graphically illustrated in Fig. 1.

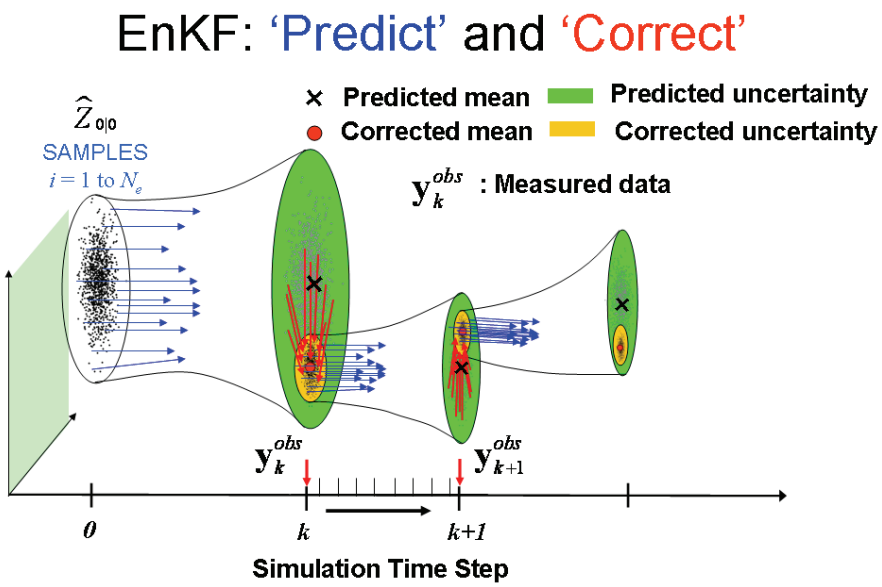


Fig. 1. Illustration of 'Prediction' and 'Correction' step in EnKF

2.1 EnKF with Confirming Option

One of the drawbacks of the conventional EnKF algorithm when applied to history matching is that the update step in Eq. (14) can cause the reservoir states to be unrealistic after the history matching. This is because the reservoir simulation equations are not explicitly used in any of the calculations of the update step. For example, the overall mass balance of the reservoir can be totally violated if the reservoir states are directly updated using the Kalman update rule. Moreover, due to the nonlinear nature of the combined state and parameter estimation problem, the updated reservoir states may not have a consistent relation to the updated parameters, i.e., if we re-simulate the reservoir with the updated parameters and the initial conditions, we may not get back the same updated states. In order to overcome this issue, it was suggested to use the Kalman update only on the parameters and indirectly update the states by re-simulating the reservoir starting from the previous history matched date (Wen and Chen 2006). Following this suggestion, we use the EnKF algorithm with the confirming option in this paper. With the confirming option, we separate out the parameter vector from the augmented state to calculate the state prediction error and the Kalman gain, i.e.,

$$e_{k|k-1}^{i} = \boldsymbol{\theta}_{k|k-1}^{i} - \boldsymbol{\mu}_{k|k-1}^{\theta} \qquad (17)$$

The update step can be divided into 2 sub-calculations:

2(a) Parameter update: Update the parameters using the Kalman update

$$\{\, \boldsymbol{\theta}_{k|k}^{i} = \boldsymbol{\theta}_{k|k-1}^{i} + \mathbf{K}_{gain}(\mathbf{y}_k^{i,obs} - \mathbf{y}_{k|k-1}^{i}) \,\}_{\text{for } i=1 \text{to } N_e} \qquad (18)$$

2(b) Confirming step (State update): Re-simulate the reservoir starting from the previous history matched date but using the newly updated parameters

$$\{\mathbf{x}_{k|k}^{i} = \mathbf{f}(\mathbf{x}_{k-1|k-1}^{i}, \boldsymbol{\theta}_{k|k}^{i})\}_{\text{for } i=1 \text{to } N_e} \qquad (19)$$

By performing the state update using the reservoir simulator model **f**, we can ensure that all the reservoir states are updated consistent with the updated model parameters and also stay within realistic domains. The assumption is that the uncertainty in the reservoir model is solely manifested through the parameter uncertainty. The EnKF algorithm is essentially used as a parameter estimation algorithm which is the main goal in history matching.

2.2 Localized EnKF

In large scale systems, the observations are scattered at various locations on the entire grid volume. For example, the production wells in a reservoir will be separated from each other by several grid blocks. We can assume that the heterogeneous geological properties at the grid block locations will only influence the production data of nearby wells. In that case, we can perform the update step in a piecewise fashion, where we will use only the measurements from near-by wells to update a particular

grid block property. This is called localization technique in the EnKF literature. Later in this article, we demonstrate the application of the localized EnKF on a highly heterogeneous 2-D reservoir. In order to do this we have to first define the region of influence of a particular well a priori. One of the advantages of this localized analysis scheme is that it will reduce the impact of spurious correlations from measurements that are insensitive to a particular parameter. In addition, it is also expected that a lesser number of ensembles will be required than would be required with a global EnKF, especially for large scale systems. The localized EnKF we described here is one of the several variants available in literature. Several types of the localization schemes and their respective advantages have been reviewed by Evensen (2009).

2.3 Geostatistical Prior

Several authors have used geostatistical models of permeability and porosity to generate the initial ensembles (Naevdal et al. 2005, Wen and Chen 2007, Haugen et al. 2008, Zhang and Oliver 2009). We consider this as a very important step for history matching using the EnKF algorithm. In all Bayesian estimation techniques, it is well known that the prior information is very valuable to inference the posterior. In the EnKF algorithm, the initial ensembles represent the prior information we have about the distribution of the static variables. The petroleum engineering literature is rich with techniques such as Kriging algorithms which can estimate the static variables at unsampled locations given hard data measured at well locations. Also, there are sophisticated techniques such as the sequential Gaussian simulation (Deutsch and Journel 1998) which can perform conditional simulation to generate multiple plausible realizations of the static parameters. The geostatistical realizations obtained from such simulations can be conditioned to honor the histogram and spatial correlations. However, the uncertainty in the predictions from these realizations can be very high if directly used for simulating the reservoir. This is evident from the fact that the production data is not used to condition these simulations. The EnKF is a natural way forward to assimilate the production data from the wells into the geostatistical realizations. We expect that by using geologically realistic covariance models to generate the initial ensemble, the EnKF algorithm will yield history matching results which are more realistic with respect to the geological heterogeneity of the reservoir.

2.4 Choice of observation variables

One of the critical factors which can affect the quality of any parameter estimation algorithm is the amount of information about the parameters that the data (vector $\mathbf{y}_k^{obs}$) carries. Hence, an important pre-processing step is to choose the production variables which contain meaningful information about the geostatistical properties. Usually production variables such as monthly production rate of oil/gas/water, steam-oil/gas-oil/gas-water ratios etc. are assumed to be sensitive to the reservoir permeabilities. However, the degree of sensitivity can vary from case-to-case. For example, the Bottom-Hole-Pressures may be remaining constant and unaffected by the permeability in a reservoir simulation model. Hence, a careful choice of production variables based on a sensitivity analysis can be expected to complement the history matching results. Sensitivity can be analyzed directly through Monte Carlo simulations of the reservoir model. The idea is to use random realizations of the static parameters from the Geostatistical prior and check for variability in the simulated production variables. Higher variability in the simulations can be interpreted as higher sensitivity.

Another important guideline is to avoid redundancy in the data used for identifying the parameters. Note that the EnKF algorithm will fuse the information from different variables in a multivariate fashion through the correction step in Eq. (18). If two variables in $\mathbf{y}_k^{obs}$ contain redundant information, then it can cause spurious update of the parameters in the correction step. For example, the Monthly Production of Oil (bbl) will be highly correlated to the Monthly Oil rate (bbl/day). If both these variables are part of the vector $\mathbf{y}_k^{obs}$, then their respective prediction errors will cause redundant update of the parameters. Hence, only one of these variables should be chosen in the objective for history matching. On the other hand, variables such as production rate of oil and that of gas will usually contain diverse information about the static properties and hence both these variables can be part of the vector $\mathbf{y}_k^{obs}$.

2.5 Choice of Q and R

For history matching applications, the favorable values for parameter noise variance **Q** usually fall in the range of 1e-4 to 1e-6. The low noise variance signifies that the parameters are static properties such as *ln*(permeability) and porosity. A low variance will also ensure that the parameters are not perturbed too far away from the span of the updated ensembles from the previous time instant. The noise variance can be chosen to be spatially varying to represent the varying level of uncertainty at different spatial locations (E.g.: variance resulting from Kriging estimation). The matrix **R** should be used as a tuning parameter in the EnKF algorithm and their favorable values differ from case-to-case. Note that the Kalman gain has an inverse relationship with respect to the measurement noise matrix **R.** Hence, the EnKF update can be expected to be more sensitive to those measurement variables for

which the measurement noise is lower. Hence, a good choice for the measurement noise will require some experience of applying the EnKF on the particular reservoir model.

2.6 Quality of History Match

In order to measure the quality of the ensembles and rank them after history matching, we define a weighted mean square error for the i^{th} realization in the ensemble as

$$WMSE^i = \frac{1}{N}\sum_{k=t_1}^{t_N}(\mathbf{y}_k^{obs} - \hat{\mathbf{y}}_k^i)^T \mathbf{R}^{-1}(\mathbf{y}_k^{obs} - \hat{\mathbf{y}}_k^i) \qquad (20)$$

where, the variables $\mathbf{y}_k^{obs}, \hat{\mathbf{y}}_k^i$ are assumed to be column vectors containing the production variables at time instant k. The vector $\hat{\mathbf{y}}_k^i$ denotes the predictions from the i^{th} model in the history matched ensemble of models. The measurement noise variance matrix applies a weighting that is inversely proportional to the amount of noise in the production variables. Since we use the EnKF as a parameter estimation algorithm, all predictions are based on simulations starting from the initial conditions. However, there are several options to choose the time period for measuring the history match performance (t_1 need not necessarily be at the initial conditions). The history match quality can be evaluated on data within the period of history matching. If the predictive capability of the model has to be evaluated, then a more reliable model validation procedure is to cross-validate against data falling outside the period of history matching. For the latter approach, it is required to divide the past production data it into *estimation* (i.e. history matching) and *validation* data set (Ljung 1999). The disadvantage would be that we cannot use all the available historical data for the model building exercise.

In order to evaluate the overall quality of history match, we define a normalized measure of fit which can be expressed in percentage scale. Note that the *WMSE* quality measure defined earlier will be dependent on the units of the production variables and the magnitude of measurement noise covariance matrix. A scale-independent measure for evaluating the overall quality of the history matching will be the commonly used *R-square* value of fit (Ljung 1999) which is defined as

$$R_{i,j}^2 = 1 - \frac{\sum_{k=t_1}^{t_N}(y_k^{obs_j} - \hat{y}_k^i)^2}{\sum_{k=t_1}^{t_N}(y_k^{obs_j} - \bar{y}^{obs_j})^2} \qquad (21)$$

In Eq. (21), we have defined the R^2 value for the i^{th} realization in the history matched ensemble and j^{th} production variable. The scalar $\bar{y}^{obs_j}$ in the denominator refers to the time average of the j^{th} production variable, averaged over the time interval $k = t_1$ to t_N. The aggregate quality of i^{tch} realization can be defined by averaging $R_{i,j}^2$ over all the j production variables, i.e.,

$$R_i^2 = \frac{1}{N_{prod}}\sum_{j=1}^{N_{prod}} R_{i,j}^2 \qquad (22)$$

where, N_{prod} is the number of production variables selected for history matching. A value of R^2 =1 should be interpreted as perfect fit between the observed data and the model predictions; a value close to 0 would mean that the predictions are just as good as the historical mean of production data (Note: Negative values would mean that the predictions are worse than forecasting just the average of the past data). The advantage of the R^2 value is that it is always guaranteed to be less than 1 and can be used as a scale-independent overall measure of history match quality. The percentage quality can be obtained by multiplying Eq. (22) by 100 %.

3. Case study on Heterogeneous Reservoirs

We apply the methodology described in the previous section to characterize and history match two synthetic heterogeneous reservoirs. The entire EnKF workflow was developed in the MATLAB computational environment by interfacing it with a commercial reservoir simulator. In both case studies, the reservoir model grid blocks were first populated with some known permeability values which are assumed as the truth case. The production data simulated from the truth case is treated as data available from the field for history matching. The truth case permeability is completely hidden from the reservoir model used in the EnKF workflow, except for the grid blocks with core hole measurements (hard data). In the case studies presented in this work, we used the algorithms available in the MATLAB geostatistical toolbox 'mGstat' (Hansen 2009) to generate the initial ensembles conditioned on the hard data. Specifically, we utilized the mGstat toolbox interfaces to the geostatistical simulation packages VISIM (Hansen and Mosegaard 2008) and SGeMS (Remy et al. 2009). The objective was to estimate all the grid block permeabilities using the production data from the truth case. Since *ln*(permeability) is usually assumed to follow a Gaussian

distribution, all the EnKF calculations are based on the log-transformed values of the permeability, i.e., in Eqs. (17) and (18) we have

$$\boldsymbol{\theta}_{j}^{i} = ln(\text{permeability})$$

For the reservoir simulator **f** in Eq. (19), the log-permeabilities are transformed back into the actual permeability values. Also, it is assumed that prior information about the upper and lower limit of the grid block permeability is available. The limits are used to constrain the permeability values at all times during the history matching process.

3.1 Heterogeneous Black-oil reservoir (2-D case study)

The first case study was performed on a benchmark data set from the 2001 SPE Comparative solution project (Christie and Blunt 2001). The model has a simple 2D vertical cross sectional geometry containing oil and gas (2-phase), modeled using a black-oil formulation. The fine-scale grid of 100x1x20 was used. Initially the reservoir was assumed to be fully saturated with oil. In the original model, there was one injector (injecting gas) and one producer well. In this case study, this was modified with one injector well at the center of the reservoir with two producer wells (Pro-1 & Pro-2) placed symmetrically on either side of the injector. The grid block permeability values in the I direction (2000 values) in the original benchmark data set are shown in Fig. 2(a). The J and K direction permeabilties were assumed to be equal to PERM-I. From the figure, this reservoir model can be observed to be highly heterogeneous with high permeability grids surrounded by extremely low permeability layers. The objective was to estimate all the 2000 grid block permeabilities using the well production data from the truth case. Figure 2(a) also shows the modified well placements. In addition to this, we also assume that two vertical core holes are drilled symmetrically in between the producer and injector wells (at i = 25, 75) to obtain hard data along these grid blocks. Figure 2(b) highlights the core-hole locations C-1 and C-2. We assume that the grid block permeabilities along Pro-1, Pro-2, C-1, and C-2 are initially known (highlighted by the black lines in Fig. 2(b)).

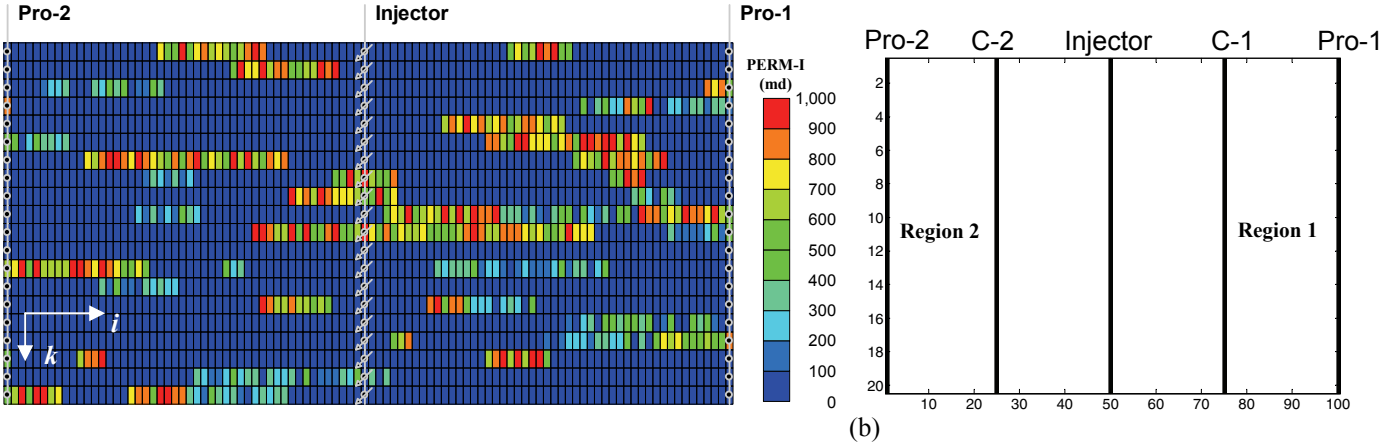


Fig. 2. (a) Heterogeneous Black-oil reservoir (2-D) with one injector and two producers, (b) Permeability hard data locations

The regions 1 and 2 shown in Fig. 2(b) show the localization regions. The region 1 permeability was updated using the production data from well Pro-1 only and the region 2 was updated using Pro-2 only. The remaining part of the reservoir (region between C-1 and C-2) was updated using both Pro-1 and Pro-2. The production data used for history matching consisted of the Monthly Gas and Oil productions (ft^3) for the two wells. Thus, the vector $\mathbf{y}_k^{obs}$ will be given by

$$\mathbf{y}_k^{obs} = \begin{bmatrix} \text{Gas Production of Pro-1} \\ \text{Gas Production of Pro-2} \\ \text{Oil Production of Pro-1} \\ \text{Oil Production of Pro-2} \end{bmatrix}_{k^{th}\, month}$$

The measurement noise variance for Gas and Oil productions were chosen as 5 ft^3 and 10 ft^3 respectively. The parameter noise variance was chosen as 1e-4 at all grid block locations. These values were arrived through a tuning exercise to obtain a good history matching result. The fit of the history matched ensemble predictions to the production data was visually analyzed to tune these values.

The first step in the EnKF workflow is to generate the initial ensemble (ensemble size = 30 realizations) using prior knowledge of the Geostatistics. Thirty realizations representing the initial ensemble were generated using a direct sequential simulation in the VISIM package, the simulation being conditioned on the hard data at the locations shown in Fig. 2(b). An anisotropic spherical variogram function with a maximum correlation range of 75 grid blocks in the east-west direction was used in the conditional simulations. Figure 3(b) shows the E-type mean of the initial ensemble. As can be observed from the figure, the initial ensemble mean is very homogenous because very few measurements of the static data is available. The localized EnKF algorithm was implemented to history match the production data falling in the period of August, 2001- March, 2013. The update step was carried out sequentially at 8 consecutive time points in this period. Counting the days from the beginning of simulation, the update was done sequentially after 250 (August, 2001), 400, 500, 1000, 1750, 2500, 3500, and 4500 (March, 2013) days respectively.

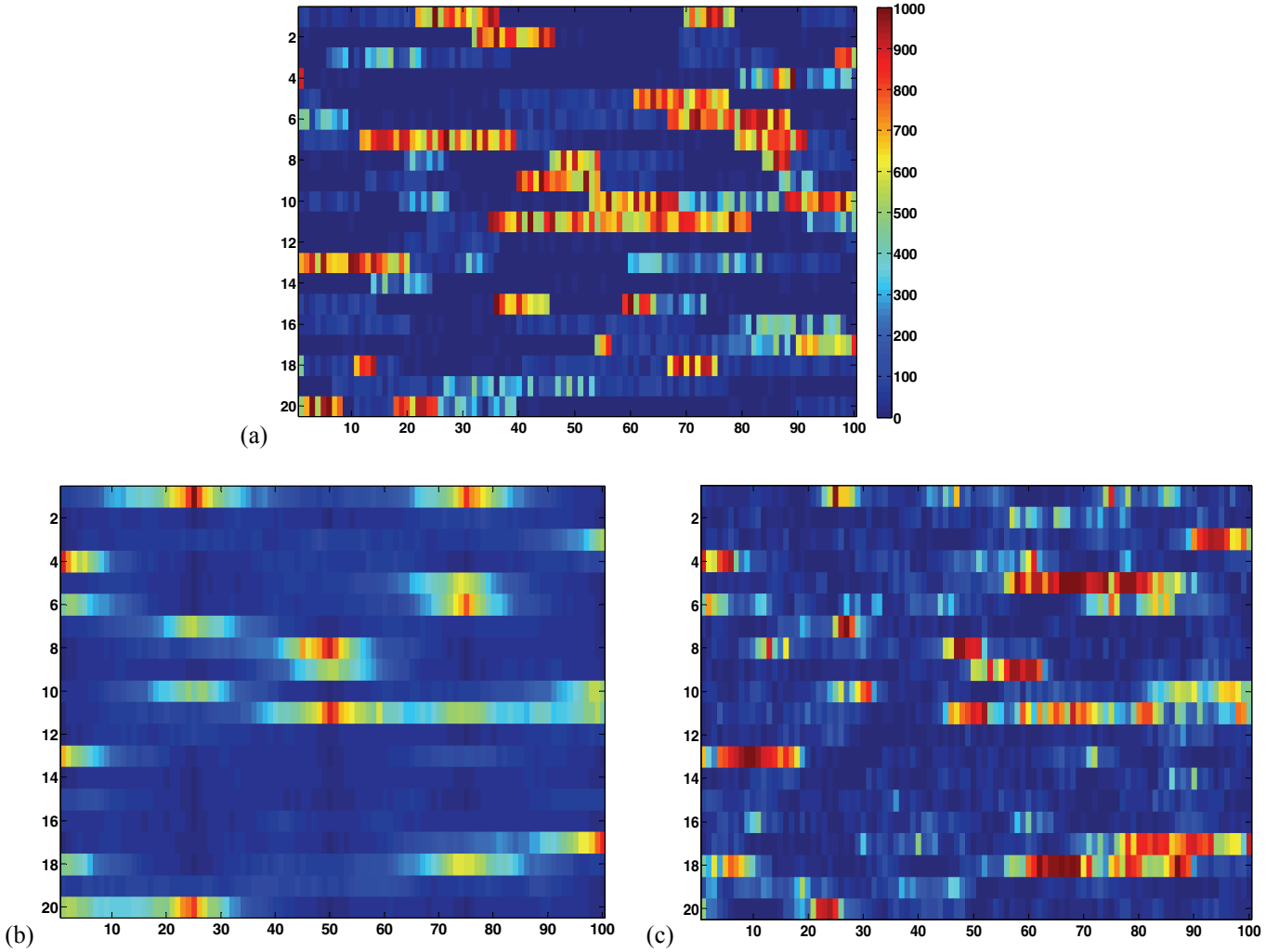
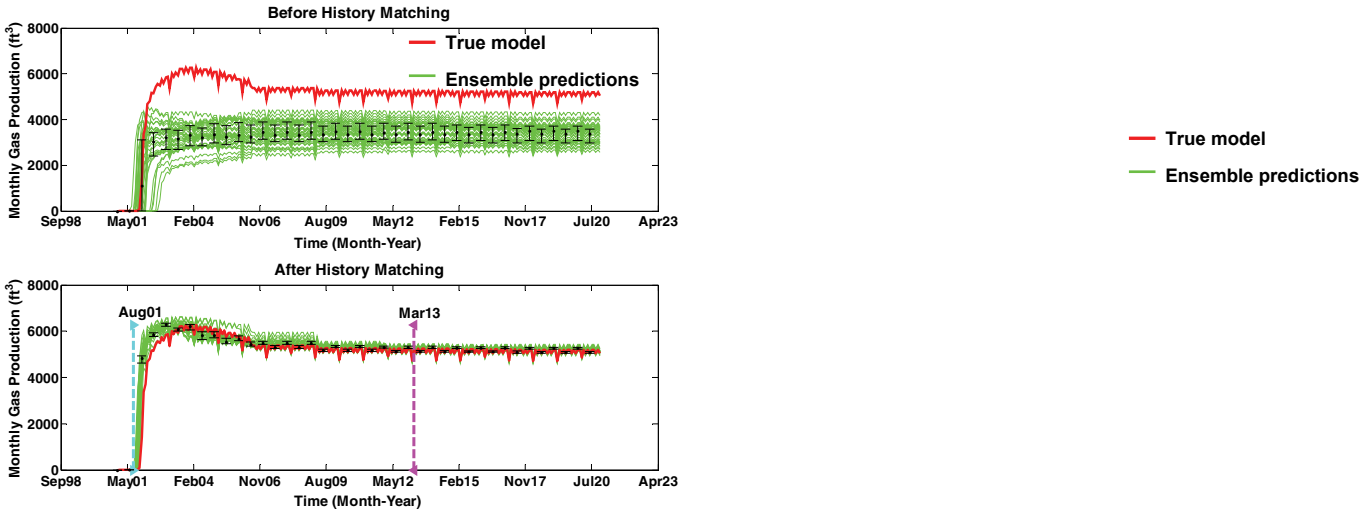


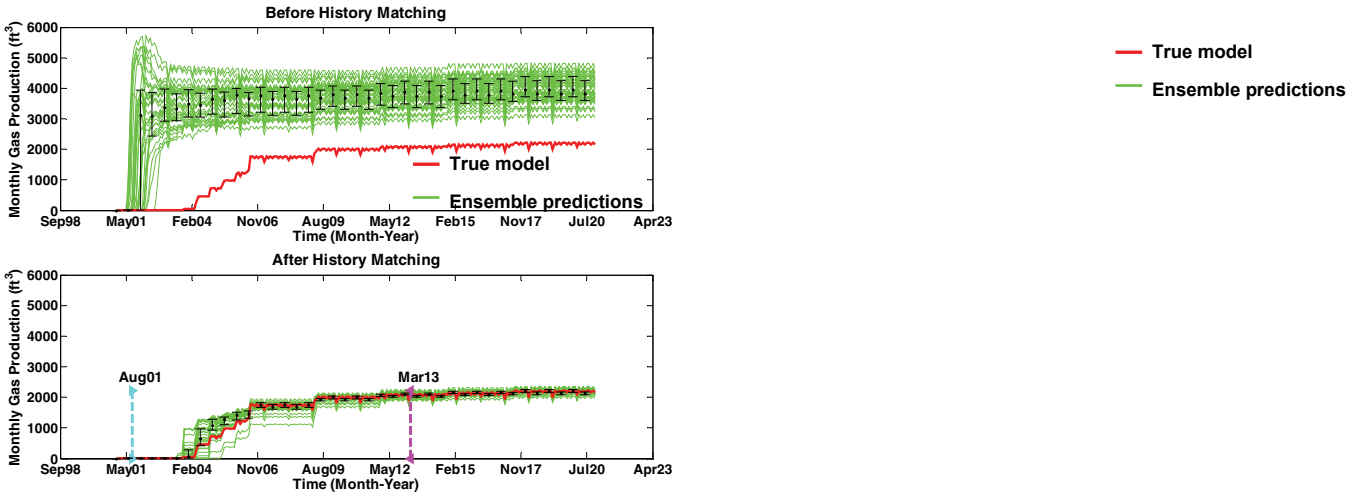
Fig. 3. Grid block permeabilities of the 2-D heterogeneous reservoir showing (a) Truth case (SPE benchmark data set), (b) E-type mean of initial ensemble realizations (30) generated using a geostatistical model, and (c) E-type mean of history matched ensemble resulting from the localized EnKF algorithm

The history matched ensemble mean is shown in Fig. 3(c). The mean permeability of the history matched ensemble show a closer match to the true reservoir heterogeneity when compared to the initial ensemble. Figures 4(a)–(d) show the comparison of ensemble predictions of the production data before and after history matching. All the predictions (before and after) shown in the figure were generated through simulations starting from the initial conditions of the reservoir. The predictions *after* history matching were obtained through simulations initialized with the ensemble of permeabilities resulting from the last update of the EnKF algorithm (March, 2013). Clearly, the EnKF algorithm is able to estimate permeabilities that result in predictions matching the reservoir history more closely compared to the initial ensemble realizations. Note that the initial/prior ensembles were based on the prior geostatistical model only. However, the history matched ensembles encompass the information assimilated through the

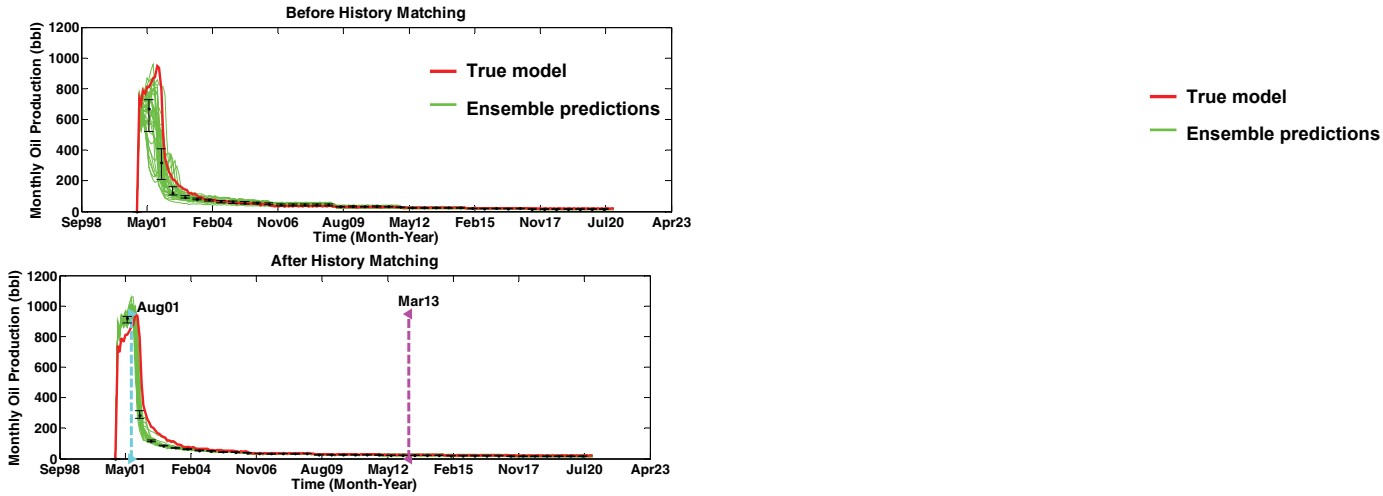
EnKF update step and hence yield predictions that closely track the true reservoir production. The error bars clearly indicate the reduction in uncertainty in terms of the smaller inter-quartile range of the realizations. There is also a significant reduction in the variance of the model predictions after the history matching. This is a result of the EnKF algorithm correction step, where each ensemble is updated by taking into account the error in the production data predictions.



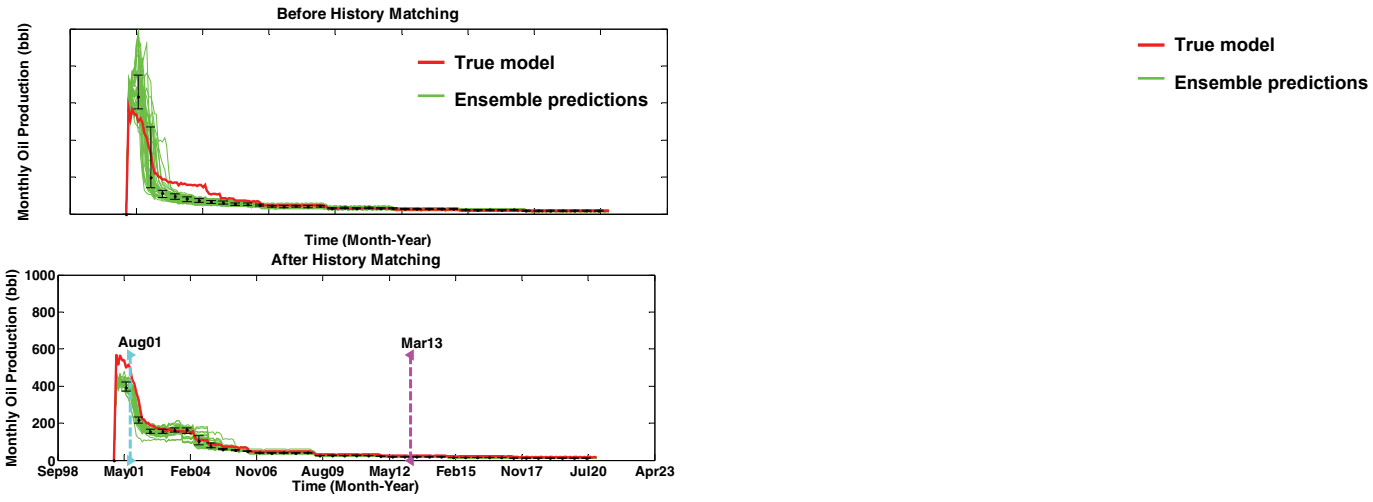
(a)



(b)



(c)

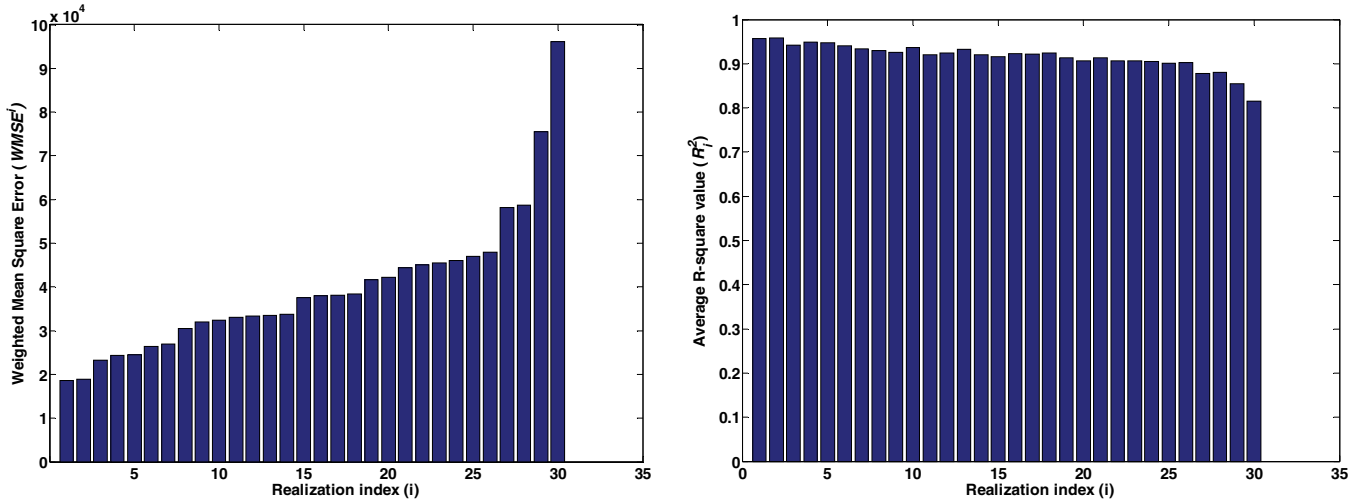


(d)

Fig. 4. Comparison of ensemble predictions (30 realizations) *before* and *after* localized EnKF history matching for (a) Well Pro-1 gas production, (b) Well Pro-2 gas production, (c) Well Pro-1 oil production, and (d) Well Pro-2 oil production. The starting and ending period of data used for history matching are flagged using the cyan and magenta lines. Error bars indicate the inter-quartile range of the ensemble predictions (1st and 3rd quartile of the ensemble).

The quality of the ensemble realizations in terms of their WMSE and R-square values are shown in Fig. 5. The realizations are sorted in the increasing order of their respective WMSE. From Fig. 5(a), we can clearly see that there is a sharp decline in quality after the 26th ensemble realization. The high sensitivity of the WMSE plot makes it a good tool for ranking the ensemble realizations in terms of their prediction accuracy. On the other hand the R-square error is less sensitive to the errors in the individual realizations and is a better indicator of the overall quality of the history match run. As shown in Fig. 5(b), the R square value of most of the realizations is close to 0.9 indicating 90% fit between the true reservoir data and the ensemble predictions. In order to evaluate the advantage of localization in this case study, we also implemented a global EnKF algorithm where all the well measurements were used to update all the grid block permeability, irrespective of the grid block locations. Though there was no significant improvement in the quality of production data match, we observed that the localized scheme yielded permeability values that had lower error compared to the truth case. This is depicted in Fig. 6 where the root mean square error (RMSE) in the estimated permeability for the global and localized EnKF is shown for the 30 realizations in the ensemble. Note that the true reservoir permeability is known a priori for this synthetic example. From the figure we can observe that the RMSE in the estimated permeability of the localized EnKF is consistently lower than that of the global EnKF algorithm.

Figure 7 shows the evolution of the ensemble mean and standard deviation as the EnKF progressively updates the reservoir. Note that there is a significant change in the mean permeability of the ensembles from k = 1 to k = 4. This is an indication of significant correction applied to the permeability ensembles by the Kalman update step to compensate for the prediction errors when compared to the true reservoir production data. We can also see that the standard deviation progressively decreases with each update step, which is manifestation of the reduction in the uncertainty of the Kalman update technique (variance minimization). Beyond k = 5, the variance remains at the low range; this is an indication that the EnKF algorithm has converged and further updates will not bring significant decrease in the variance. The zero-variance along core hole locations signify that no update in performed at these locations because the true permeability is measured/available in the form of hard data along these grid blocks. Figure 8 shows the evolution of the aggregate mean of the E-type standard deviation, the averaging being done over the entire reservoir at each update time step. The reduction in the overall variance of the permeability ensemble with the progression of the Kalman updates is reiterated in this figure.



(a) (b)

Fig. 5. Quality of the ensembles (30 realizations) *after* localized EnKF history matching in terms of (a) WMSE: The ensembles are sorted in the ascending order of the WMSE and those with relatively poor WMSE can be discarded if necessary, and (b) R-square value: For a reliable history match run, all the ensembles will show high R^2 (maximum possible value is 1). This will indicate the overall quality of the EnKF history match.

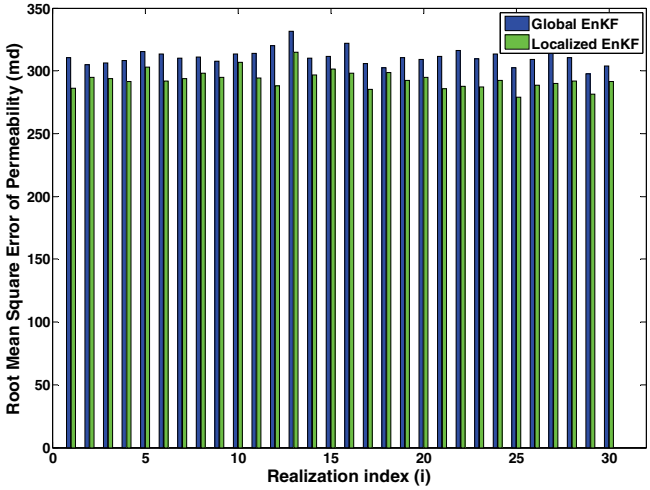


Fig. 6. Comparison of the Root Mean Square Error of estimated permeability of *localized* and *global* EnKF algorithms

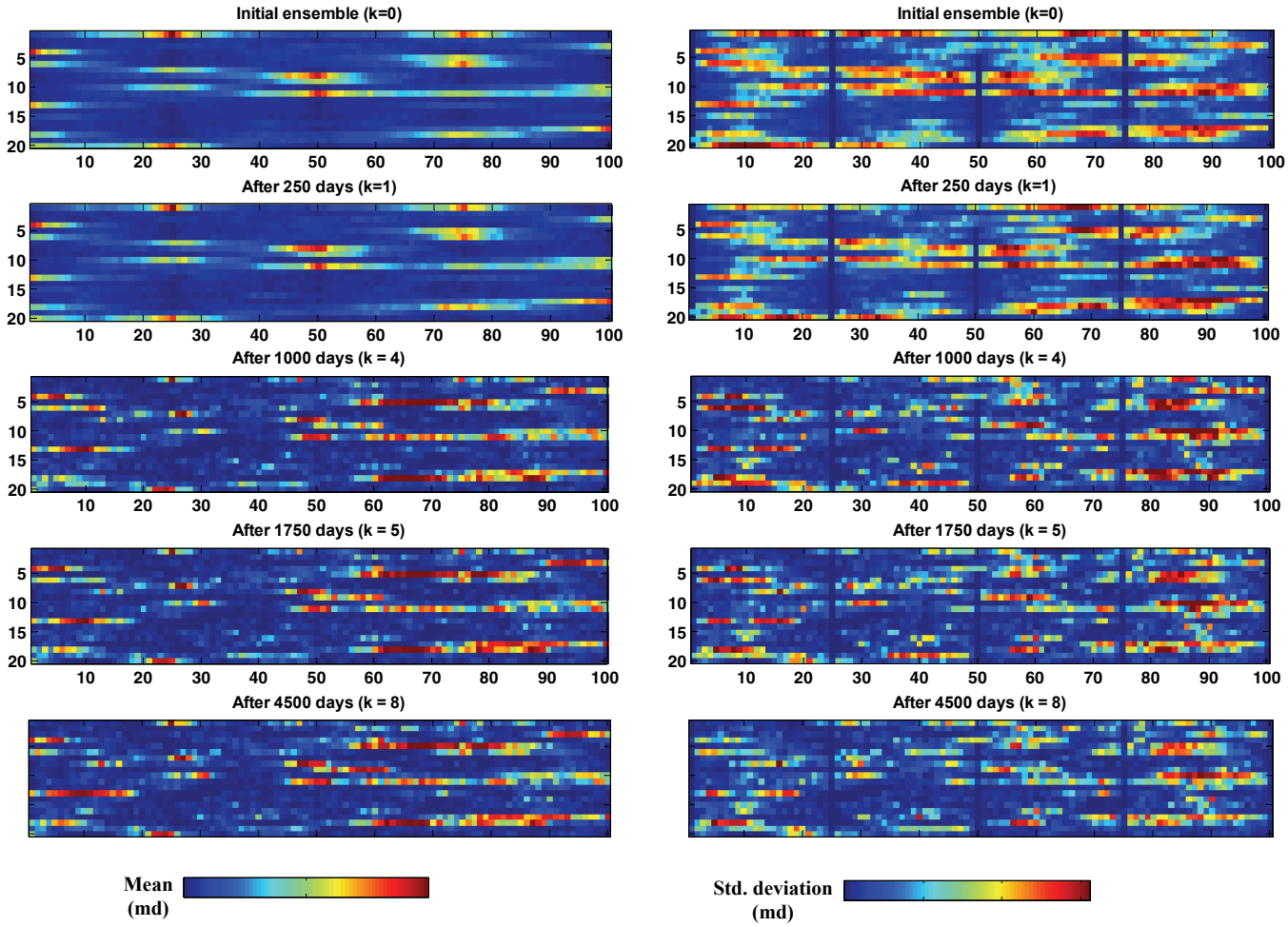


Fig. 7. Evolution of E-type mean (*on the left*) and standard deviation (*on the right*) of the permeability ensembles with the progression of localized EnKF update steps

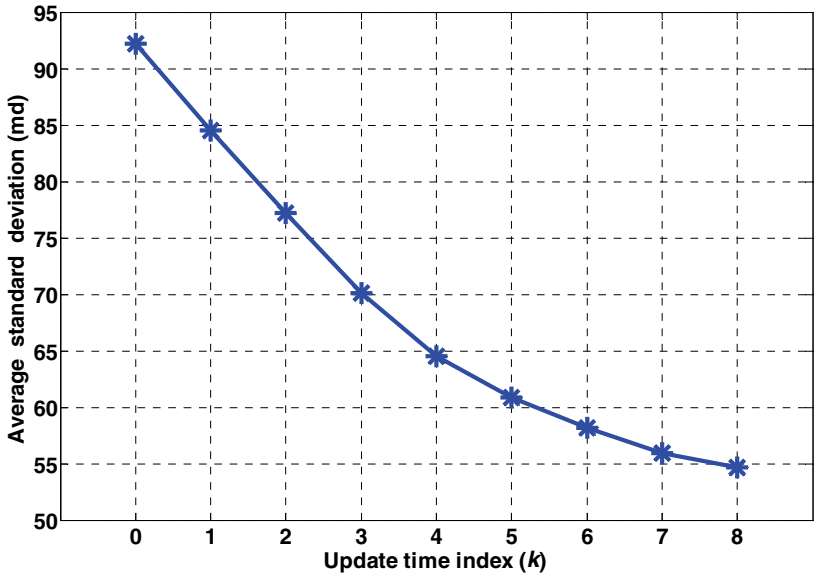
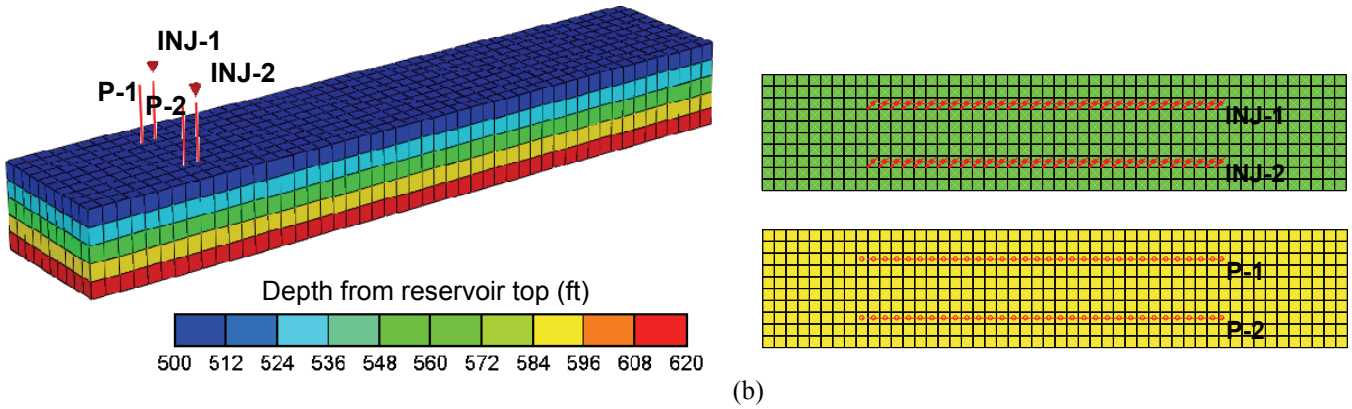


Fig. 8. Evolution of the aggregate mean of permeability E-type standard deviation with the progression of localized EnKF update steps

3.2 SAGD reservoir (3-D case study)

The second case study was performed on a synthetic 3-dimensional Steam Assisted Gravity Drainage (SAGD) reservoir with 80% initial heavy oil saturation. The dimensions of the reservoir were 5000x1000x150 ft, modeled using a coarse scale grid of size 100x100x30 ft in the respective dimensions (Number of grids = 50x10x5). Two pairs of Injector-Producer wells of length 30 grid blocks (length = 3000 ft) were assumed as shown in Fig. 9(a). The horizontal trajectory of the injector wells (INJ-1, INJ-2) in layer 3, and that of the producer wells (P-1, P-1) in layer 4 are also shown in Fig. 9(b). Figure 10(a) shows the spatial permeability distribution of the truth case, generated through a geostatistical simulation using the SGeMS package by assuming a geostatistical variogram model. The sequential Gaussian simulation (sgsim) technique was chosen as the simulation method and *ln*(permeability) was the simulated property. The overall variance of the random field (variogram sill value) was set at 1. An anisotropic spherical variogram model with maximum correlation range of 15 grid blocks was used. Steam at 650^0F was injected through the injection wells and 10 years of the reservoir production was simulated to generate the truth case production history. The producer wells were initially shut-in, for a period of 1 year of steam injection, and opened for production from the 2nd year onwards. The simulated data was used for estimating the grid block permeabilities using the EnKF, by assuming that the true grid-block permeability is not known.



(a) (b)

Fig. 9. (a) Synthetic SAGD reservoir of size 50x10x5 containing two Injector-Producer well pairs, (b) Locations of the Injector wells (INJ-1, INJ-2) and Producer wells (P-1, P-2) in the 3rd and 4th layer of the reservoir respectively

PERM-I
(md)

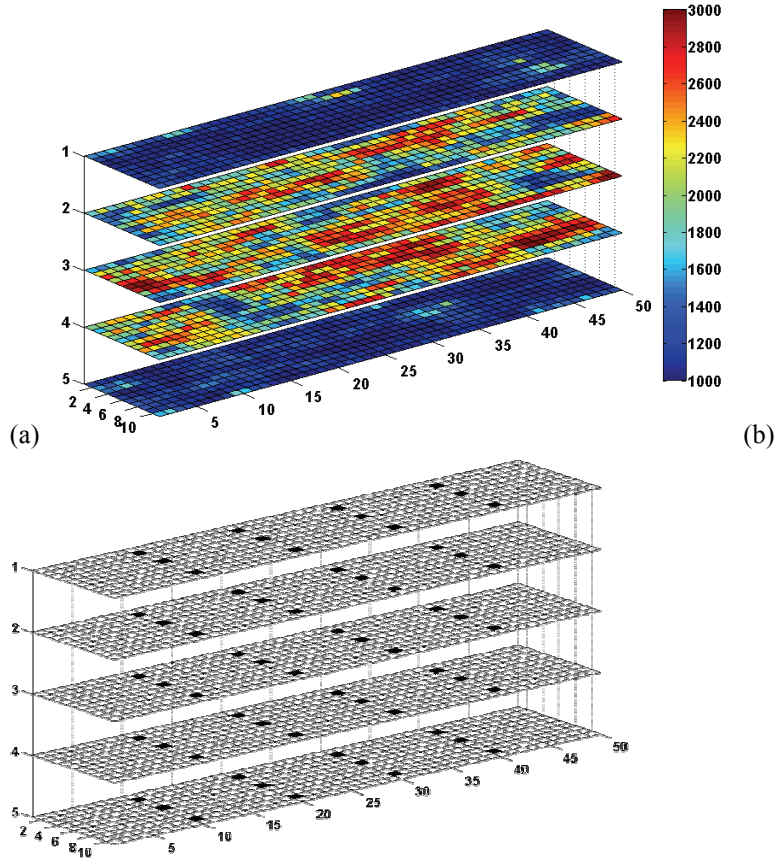


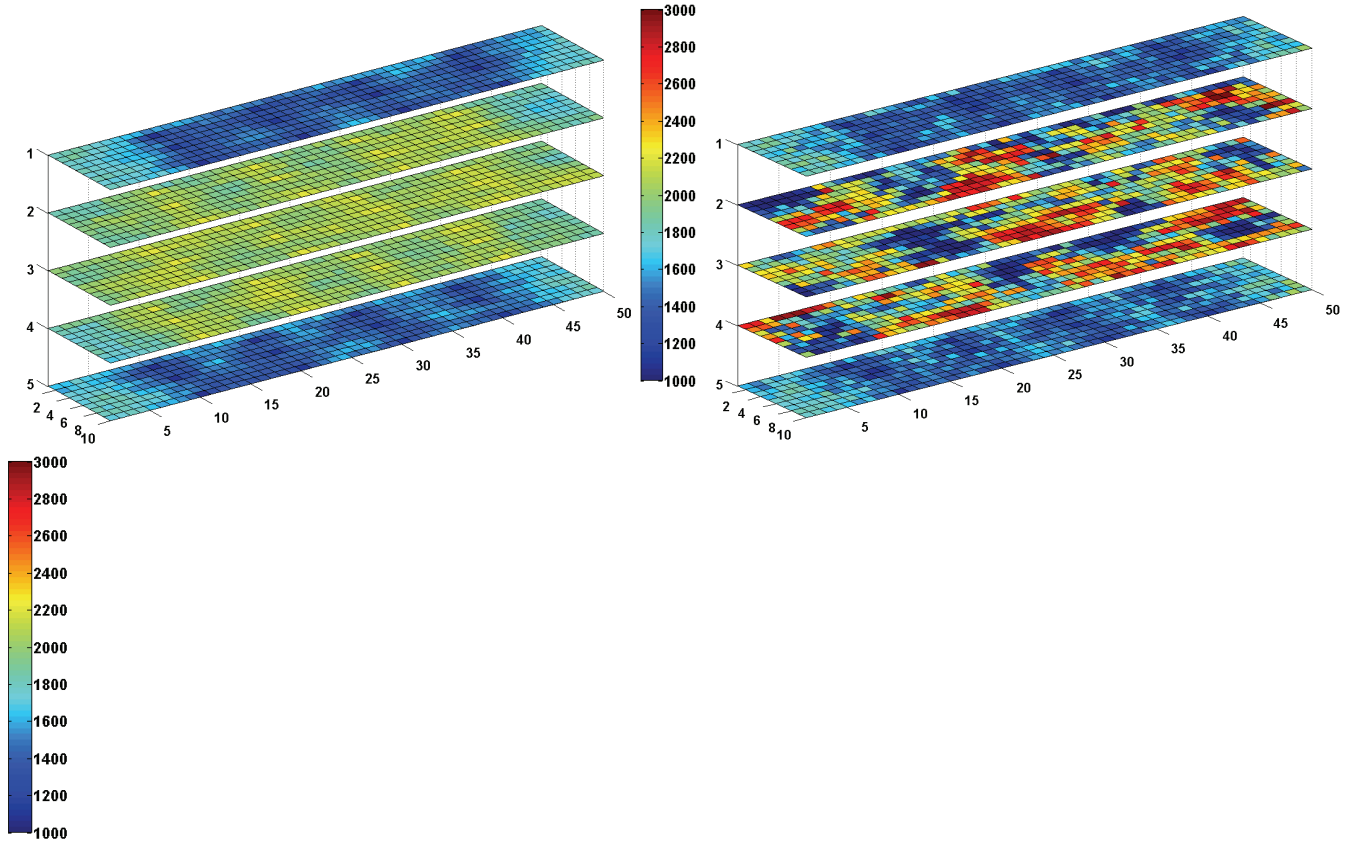
Fig. 10. (a) Permeability of the true reservoir generated using a sequential Gaussian simulation, (b) Locations of 12 vertical core holes at which hard data of the true reservoir permeability were assumed to be available. The hard data at core hole used as conditioning data for generating the initial ensembles and bootstrap the EnKF algorithm

The production performance of the SAGD process is evaluated using two production variables which are the oil production rate (OP_{rate}) and the cumulative steam-to-oil ratio (SOR). The geological heterogeneity will affect these production variables McLennan and Deutsch 2010 and hence we have used these variables, evaluated for the total field, as the criteria for history matching in this case study. Thus, the measurement vector in the EnKF algorithm consists of the combined total OP_{rate} and SOR for the entire field and is given by

$$\mathbf{y}_k^{obs} = \begin{bmatrix} \text{Cumulative SOR of entire field} \\ \text{OP}_{\text{rate}} \text{ of entire field} \end{bmatrix}_{k^{th}\, month}$$

A measurement noise variance of 1e-4 and 10 were assigned for SOR and OP_{rate} respectively in the EnKF algorithm. The permeability perturbation noise variance was chosen as 1e-4 at all grid block locations. Similar to the previous case study, we assume that hard data samples of the true permeability is initially available at a few core-hole locations. Vertical core holes traversing the five layers were assumed and their spatial locations are as shown in Fig. 10(b). A conditional simulation, using the hard data as conditioning data and sgsim as the simulation algorithm, was performed to generate the initial ensemble realizations (ensemble size = 30). Figure 11(a) shows the E-type mean of the initial ensemble realizations. The variogram function of the *ln*(permeability) random field used for the initial ensemble generation was intentionally made different from the truth case. A variance of 2 and correlation range of 10 grid blocks were used, instead of the corresponding values of 1 and 15 in the true reservoir. The error in these variogram function parameters is a manifestation of the uncertainty/lack of exact knowledge about the reservoir geology which is the main motivation for characterizing the reservoir by assimilating the production data using the EnKF algorithm.

PERM-I (md) **PERM-I (md)**



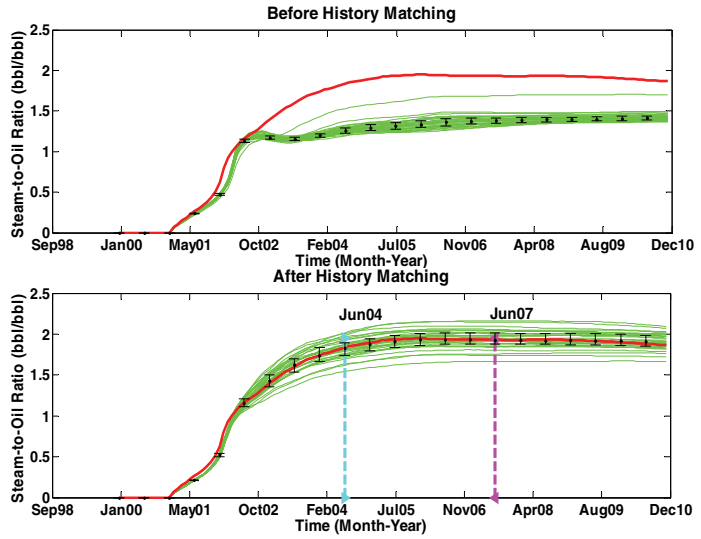
(a) (b)

Fig. 11. E-type mean of the ensemble realizations (30) showing permeability (a) *before* history matching, and (b) a*fter* history matching using the EnKF algorithm

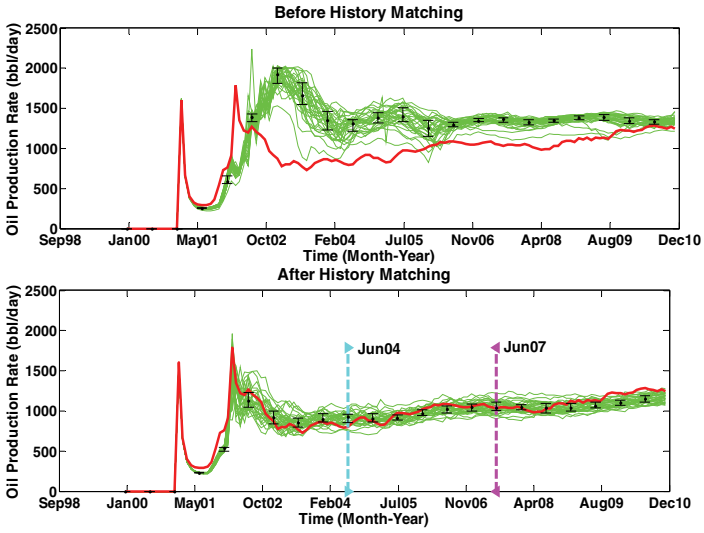
In this case study, an additional pre-processing of the initial ensembles was performed before executing the EnKF algorithm. This was necessary to account for the fact that the SAGD production variables are more sensitive to layers close to the well pairs (layers 2, 3 and 4). The variance of the permeability of the grid blocks in layer 1 and layer 5 after the sequential Gaussian simulation was extremely high (around 600 md). The pre-processing was done in a localized manner in layer 1 and layer 5 only in order to reduce their variance among the ensembles (E-type variance). The E-type variance of the layer 1 and layer 5 was reduced by replacing the geostatistical simulations with the E-type mean of these two layers, with the addition of a Gaussian random noise of standard deviation 50 md. Note that the Kalman update is more sensitive to those regions for which the cross-covariance $\hat{\mathbf{P}}_{k|k-1}^{e,\varepsilon}$ is high. Without this pre-processing step, the high variance of layer 1 and 5 permeabilities was found to induce spurious correlations between the ensemble production data and these layers, which eventually resulted in false update of the permeabilities in these layers.

The history matching was performed for a period of 3 years and the production data in the $\mathbf{y}_k^{obs}$ vector was assimilated every 6 months, starting from June, 2004. Figure 11(b) shows the E-type mean of the history matched ensemble of grid block permeabilities. Comparing Fig. 11(b) and the true reservoir permeability in Fig. 10(a), we can observe that the EnKF updated ensembles contain more plausible information about the spatial heterogeneity. Figure 12 shows the comparison of ensemble predictions of the production data before and after history matching. Similar to the previous case study, the ensemble of history matched grid block permeabilities at the end of the last update step was used to simulate the '*After history matching'* case starting from the initial conditions. The predictions so obtained were compared to the simulated predictions of the true reservoir. To compare the effectiveness of history matching, the initial ensembles were used in a separate simulation to produce the *'Before history matching'* case. From the comparison of the predictions, we can clearly observe that the realizations after the EnkF based history matching are much more plausible when compared to the initial ensembles. The quality of the ensembles in terms of the

WMSE and R-square value is shown in Fig. 13. The R-square value suggests more than 90% fit between the predictions and the observed production data for all the ensembles, this showing that the history match is reliable.

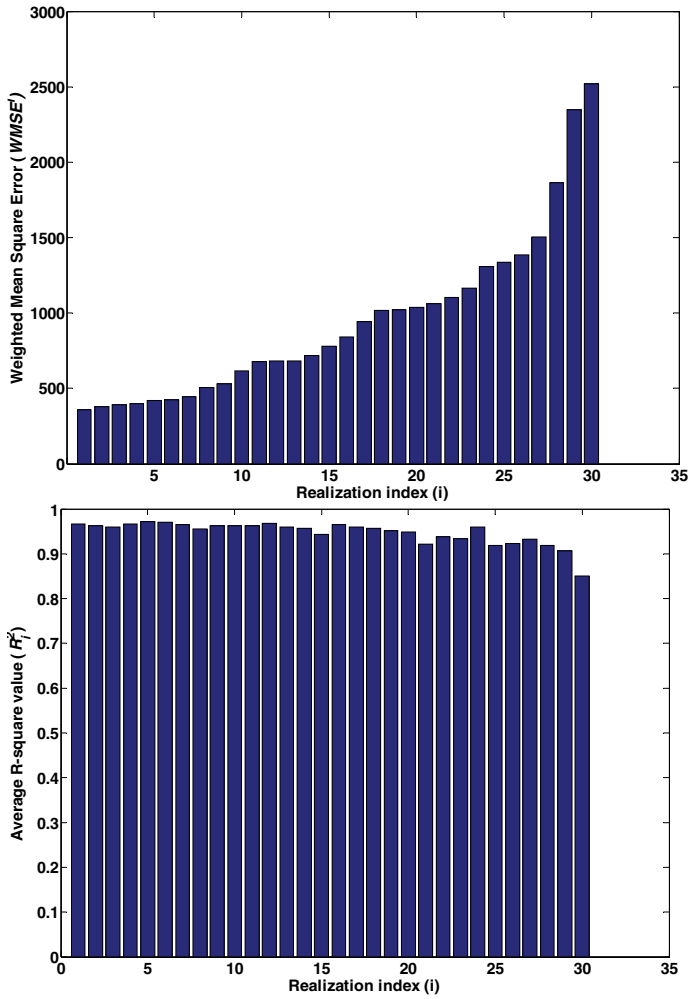


(a)



(b)

Fig. 12. Comparison of ensemble predictions (30 realizations) *before* and *after* EnKF history matching for (a) Cumulative Steam-to-Oil Ratio (SOR), and (b) Oil Production Rate (OP_{rate}) for the entire field



(a) (b)

Fig. 13. Quality of the ensembles (30 realizations) *after* EnKF history matching of the SAGD reservoir in terms of (a) WMSE: The realizations after i = 25 can be considered to have poor accuracy relative to the others, (b) R-square value: The high R-square value of all the ensembles indicate that the history match is reliable.

Figure 14 shows how the E-type mean and standard deviation of the ensemble permeability evolves as the EnKF algorithm sequentially updates the reservoir. It can be observed that a significant shift in the mean of the ensembles happens at the first update instance. This is due to large errors in the initial ensemble predictions which calls for a higher error compensation in the Kalman update step. There is also a significant reduction in the ensemble variance that occurs at the first update step. After the third update onwards (June, 2005), there model prediction errors are significantly lower and hence only minute adjustments are required. The EnKF algorithm automatically infers this vital information purely based on the sample statistics (covariance and cross-covariance) of the ensembles and takes account of this information when making any parameter updates through the Kalman update technique. The reduction in the overall variance of the permeability ensembles can be clearly observed in Fig. 15. In Fig. 16, we compare the steam chamber growth predictions after history matching. For this comparison, we used the realization of the history matched ensemble that yielded the lowest WMSE, i.e. ranked best in terms of the production data history match. Overall, we can observe a good match between the truth case and the predicted spatial distribution of the oil saturation.

Mean (md)

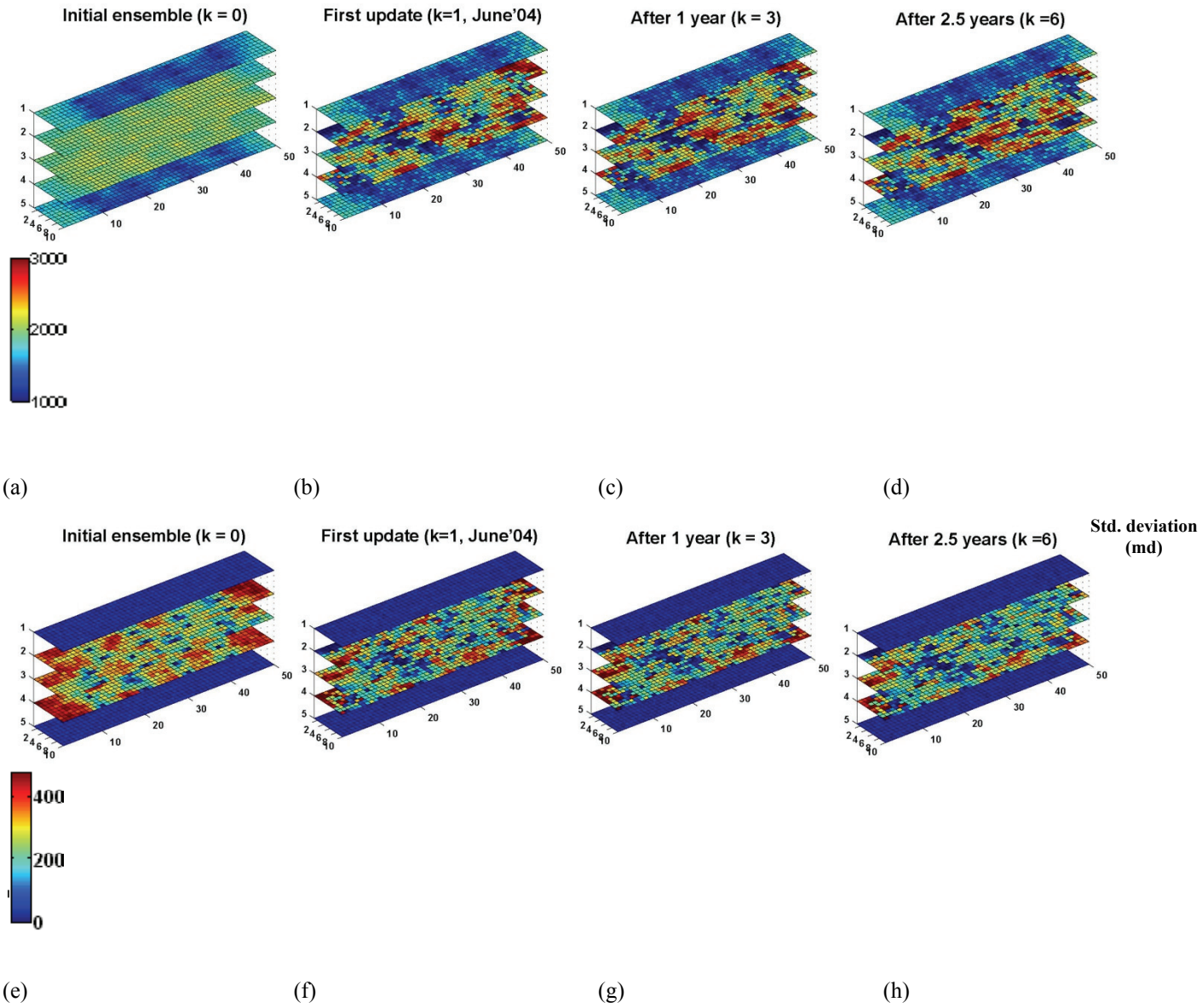


Fig. 14. Evolution of E-type mean (*top row*) and standard deviation (*bottom row*) of the permeability ensembles with the progression of EnKF updates, (a, e): Initial ensemble generated using geostatistical simulation, (b, f): After first update, (c, g): After 1 year from first update (3rd update), and (d, h): After 2.5 year from first update (6th update). The ensemble std. deviation significantly reduces after the first update and remains low thereafter showing significant convergence of the EnKF algorithm.

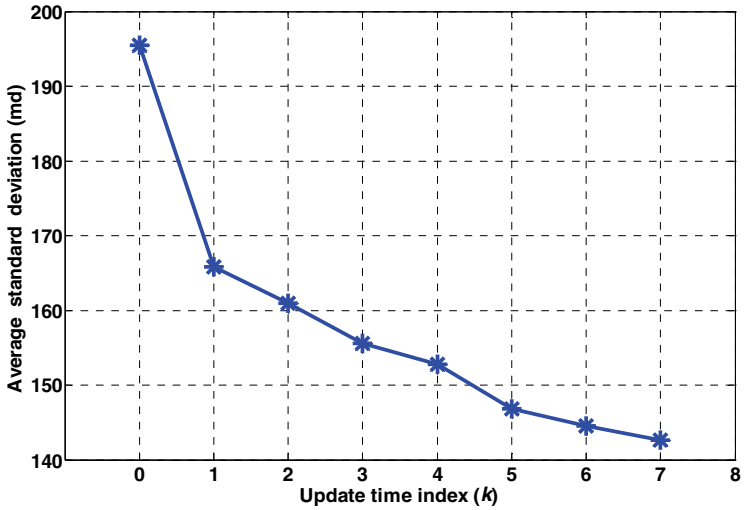
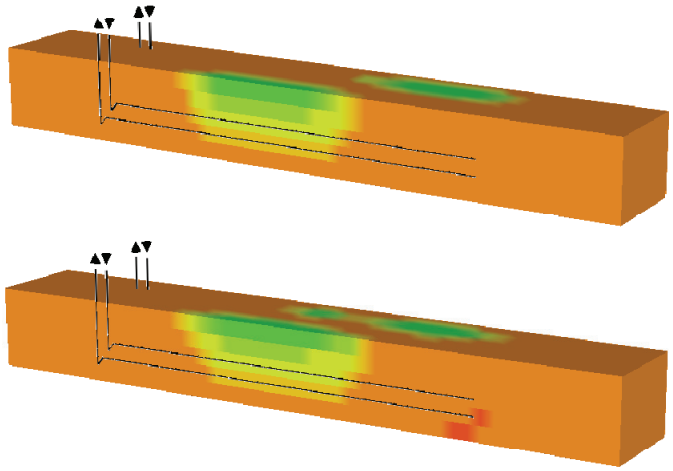


Fig. 15. Evolution of the aggregate mean of permeability E-type standard deviation with the progression of EnKF updates

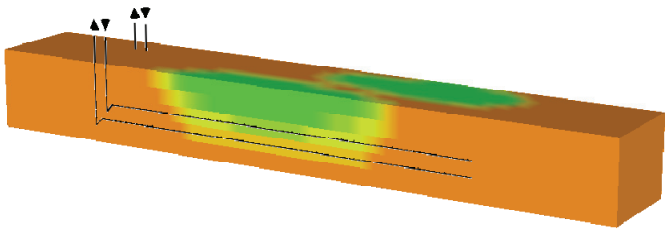
(a) June, 2005

Benchmark (Truth case) EnKF Realization (Best case)



(b) June, 2008

Benchmark (Truth case) EnKF Realization (Best case)



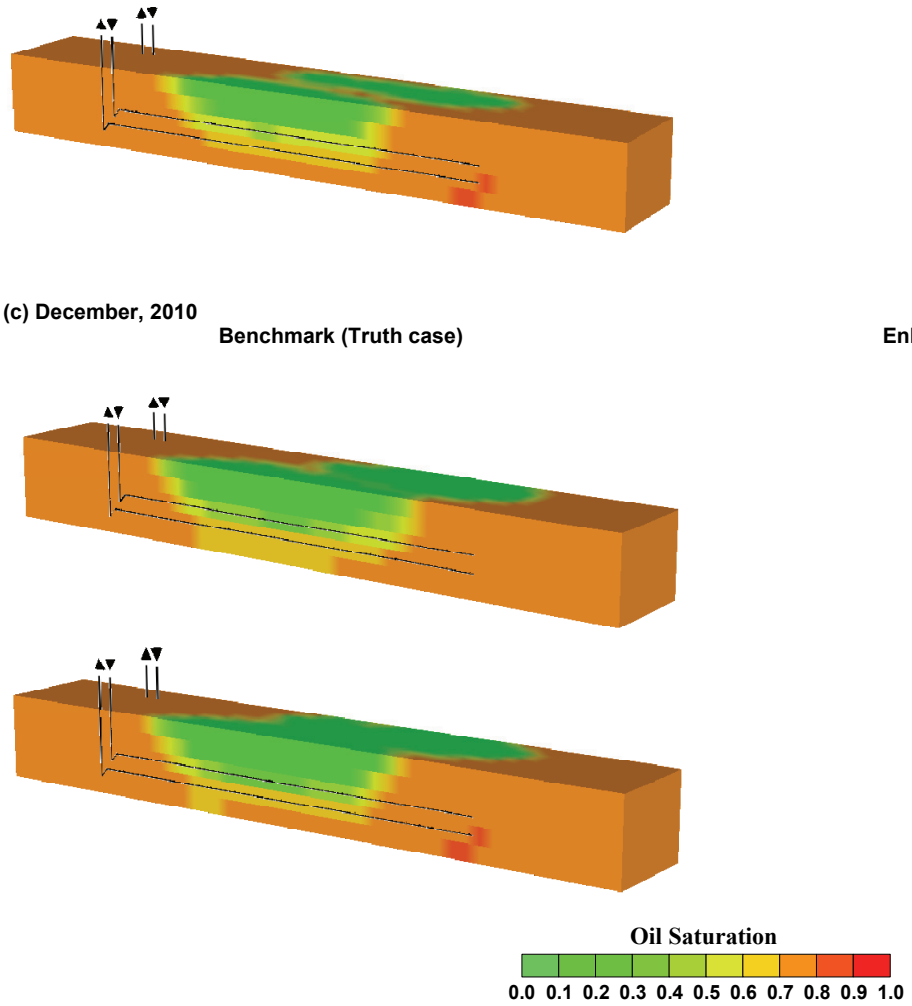


Fig. 16. Comparison of the steam chamber growth of the Truth case (*on the left*) and the EnKF realization ranked highest after history matching (*on the right*), for the month of (a) June'05, (b) June'08, and (c) December'10. The figures show the oil saturation in a reservoir section cut out along the vertical plane containing one of the well pairs.

4. Conclusion

In this work, we have demonstrated the efficacy of the EnKF algorithm for history matching of heterogeneous reservoirs by taking into account the prior knowledge about the geological heterogeneity. Considering two synthetic case studies, a 2-dimensional heterogeneous black oil reservoir and a 3-dimensional SAGD reservoir, we have shown the application of this technique for sequentially updating the heterogeneous permeabilities at the grid block locations. Two quality measures for evaluating the efficacy of history matching was proposed, which can summarize the overall quality as well as the quality of individual ensembles at the end of history matching in the EnKF algorithm. The prior knowledge about the geological heterogeneity can be easily integrated through geostatistical simulation techniques which modern day reservoir engineers are accustomed to doing as part of any reservoir uncertainty evaluation project. The entire algorithm is non-iterative in nature and does not suffer from the limitations of expensive numerical optimization routines. One of the main advantages of the EnKF algorithm is that it can account for the uncertainty in the unknown parameters such as permeability and porosity using a few Monte Carlo realizations ($O(50)$). The algorithm allows for each of the Monte Carlo realizations to be simulated on any commercial reservoir simulator and the computations does not involve the partial differential equations of the reservoir simulator in any form. Moreover, the technique inherently allows each of the simulations to be carried out independently from the others, in a parallel fashion. Significant reduction of overall time required for history matching is possible by implementing it on modern day multi-processor computing architectures.

Acknowledgements

The authors would like to express sincere gratitude to Ali Gul, graduate student in the Dept. of Civil and Environmental Engineering at the University of Alberta, for providing the reservoir simulation models used in this work. Financial support through the NSERC-MATRIKON-SUNCORE-ICORE Industrial Research Chair program in Computer Process Control at the University of Alberta is gratefully acknowledged. Also, the authors would like to acknowledge Computer Modeling Group Ltd., Calgary for providing academic licenses of the reservoir simulation package used in this work at the University of Alberta.

Nomenclature

$e^i_{k|k-1}$ = *error in i^{th} predicted state ensemble at time instant k (state prediction error ensemble)*

$\mathbf{f}(.)$ = *state transition function (dynamic model of reservoir states)*

$\mathbf{g}(.)$ = *measurement function*

i = *ensemble member/realization index*

k = *time instant index*

$\mathbf{K}_{gain}$ = *Kalman gain*

N = *number of time instants in a time interval*

N_e = *number of ensemble members (ensemble size)*

$\hat{\mathbf{P}}^{\varepsilon,\varepsilon}_{k|k-1}$ = *sample covariance of measurement error ensemble*

$\hat{\mathbf{P}}^{e,\varepsilon}_{k|k-1}$ = *sample cross-covariance between state prediction error and measurement error*

$\mathbf{Q}$ = *parameter perturbation noise covariance matrix*

$\mathbf{R}$ = *measurement noise covariance matrix*

$R^2_{i,j}$ = *R-square value/goodness of fit of i^{th} ensemble member for j^{th} production variable*

R^2_i = *Average R-square value of i^{th} ensemble member (quality of history match)*

$\mathbf{v}_k$ = *measurement noise vector*

$\mathbf{w}_k$ = *Gaussian noise vector for parameter perturbation (captures modeling errors)*

$WMSE^i$ = *weighted mean square error of i^{th} ensemble member after history matching*

$\mathbf{x}_k$ = *reservoir state vector*

$\mathbf{y}_k$ = *measurement vector (dynamic production variables)*

$\hat{\mathbf{y}}^i_k$ = *predicted production data vector of i^{th} ensemble at time instant k (predicted measurements)*

$\mathbf{y}^{obs}_k$ = *observed production data vector at time instant k (field production measurements)*

$\mathbf{z}_k \equiv \begin{bmatrix} \mathbf{x}_k \\ \boldsymbol{\theta}_k \end{bmatrix}$ = *augmented state vector for combined state and parameter estimation*

$z^i_{k|k-1}$ =*predicted augmented state vector (i^{th} ensemble member)*

$\boldsymbol{\theta}_k$ = *model parameter vector (Permeability, porosity etc.)*

$\boldsymbol{\mu}^z_{k|k-1}$ = *predicted state ensemble mean*

$\boldsymbol{\mu}^y_{k|k-1}$ = *predicted measurement ensemble mean*

$\varepsilon^i_{k|k-1}$ = *error in i^{th} predicted measurement ensemble at time instant k (measurement error ensemble)*

References

1. Aanonsen, S.I., G. Naevdal, D.S. Oliver, A.C. Reynolds, and B. Valles, *The Ensemble Kalman Filter in Reservoir Engineering - a Review.* SPE Journal, 2009. **14**(3): p. 393-412.
2. Burgers, G., P.J. van Leeuwen, and G. Evensen, *Analysis scheme in the ensemble Kalman filter.* Monthly Weather Review, 1998. **126**: p. 1719-1724.
3. Christie, M. and M. Blunt. *SPE Comparative Solution Project* 2001; First data set (Model-1). Available from: http://www.spe.org/web/csp//datasets/set01.htm.
4. Deutsch, C.V. and A.G. Journel, *GSLIB: Geostatistical Software Library and User's Guide*, ed. A.G. Journel. 1998: Oxford University Press, New York.
5. 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**: p. 10143-10162.
6. Evensen, G., *The Ensemble Kalman Filer: theoretical formulation and practical implementation.* Ocean Dynamics, 2003. **53**: p. 343-367.
7. Evensen, G., *Data Assimilation: The Ensemble Kalman Filter*. 2007: Springer Berlin.
8. Evensen, G., *The ensemble Kalman filter for combined state and parameter estimation.* IEEE Control Systems Magazine, 2009. **29**(3): p. 82-104.
9. Evensen, G., J. Hove, H.C. Meisingset, E. Reiso, K.S. Seim, and O. Espelid, *Using the EnKF for Assisted History Matching of a North Sea Reservoir Model*, in *SPE Reservoir Simulation Symposium, Houston, TX*. 2007, Society of Petroleum Engineers, Inc.
10. Hansen, T.M. *mGstat : A Geostatistical Matlab toolbox*. 2009; Version 0.95. Available from: http://mgstat.sourceforge.net.
11. Hansen, T.M. and K. Mosegaard, *VISIM : Sequential simulation for linear inverse problems.* Computers and Geosciences, 2008. **34**(1): p. 53-76.
12. Haugen, V., L.J. Natvik, G. Evensen, A. Berg, K. Flornes, and G. Naevdal, *History Matching Using the Ensemble Kalman Filter on a North Sea Field Case.* SPE Journal, 2008. **13**(4): p. 382-391.
13. Kalman, R.E., *A New Approach to Linear Filtering and Prediction Problems.* Transactions of the ASME--Journal of Basic Engineering, 1960. **82**(Series D): p. 35-45.
14. Ljung, L., *Comparing model structures* (chap. 16, sec. 16.4), in *System Identification:Theory for the User*. 1999, Prentice Hall PTR, New Jersey, USA.
15. Lorentzen, R.J., K.K. Fjelde, J. Froyen, A.C.V.M. Lage, G. Naevdal, and E.H. Vefring, *Underbalanced and Low-head Drilling Operations: Real Time Interpretation of Measured Data and Operational Support*, in *SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana*. 2001, Society of Petroleum Engineers, Inc.
16. McLennan, J.A. and C.V. Deutsch, *Best Practice Reservoir Characterization for the Alberta Oilsands*, in *7th Canadian International Petroleum Conference, Calgary, AB, Canada*. 2006.
17. McLennan, J.A. and C.V. Deutsch, *SAGD Reservoir Characterization Using Geostatistics: Application to the Athabasca Oil Sands, Alberta, Canada*, in *CHOA Handbook*. 2010, Canadian Heavy Oil Association. p. 321-340.
18. Moradkhani, H., S. Sorooshian, H.V. Gupta, and P.R. Houser, *Dual state–parameter estimation of hydrological models using ensemble Kalman filter.* Advances in Water Resources, 2005. **28**: p. 135-147.
19. Naevdal, G., L.M. Johnsen, S.I. Aanonsen, and E.H. Vefring, *Reservoir Monitoring and Continuous Model Updating Using Ensemble Kalman Filter.* SPE Journal, 2005. **10**(1): p. 66-74.
20. Oliver, D.S., A.C. Reynolds, and N. Liu, *Inverse Theory for Petroleum Reservoir Characterization and History Matching*. 2008: Cambridge University Press, New York.
21. Remy, N., A. Boucher, and J. Wu, *Applied Geostatistics with SGeMS: A User's Guide*. 2009: Cambridge University Press.
22. Romeu, R.K., *History Matching and Forecasting.* Journal of Petroleum Technology, 2010. **April**: p. 80.
23. Schulze-Riegert, R. and S. Ghedan, *Modern Techniques for History Matching*, in *9th International Forum on Reservoir Simulation, Abu Dhabi, United Arab Emirates*. 2007.
24. Schulze-Riegert, R.W., J.K. Axmann, O. Haase, D.T. Rian, and Y.L. You, *Optimization Methods for History Matching of Complex Reservoirs*, in *SPE Reservoir Simulation Symposium, Houston, Texas*. 2001, Society of Petroleum Engineers, Inc.
25. Seiler, A., G. Evensen, J.-A. Skjervheim, J. Hove, and J.G. Vabo, *Advanced Reservoir Management workflow using an EnKF based assisted history matching method*, in *SPE Reservoir Simulation Symposium, Woodlands, TX*. 2009, Society of Petroleum Engineers, Inc.
26. Wen, X.-H. and W.H. Chen, *Real-Time Reservoir Model Updating Using Ensemble Kalman Filter With Confirming Option.* SPE Journal, 2006. **11**(4): p. 431-442.
27. Wen, X.-H. and W.H. Chen, *Some Practical Issues on Real-Time Reservoir Model Updating Using Ensemble Kalman Filter.* SPE Journal, 2007. **12**(2): p. 156-166.

28. Zhang, Y. and D.S. Oliver, *History Matching Using a Hierarchical Stochastic Model With the Ensemble Kalman Filter: A Field Case Study*, in *SPE Reservoir Simulation Symposium, Woodlands, TX*. 2009, Society of Petroleum Engineers, Inc.