The Ensemble Kalman Filter for Continuous Updating of Reservoir Simulation Models

Yaqing Gu
Mewbourne School of Petroleum and Geological Eng., The University of Oklahoma, 100 East Boyd Street, SEC T301, Norman, OK 73019
e-mail: yaqing.gu-1@ou.edu

Dean S. Oliver
Mewbourne School of Petroleum and Geological Eng., The University of Oklahoma, 100 East Boyd Street, SEC T301, Norman, OK 73019
e-mail: dsoliver@ou.edu

This paper reports the use of ensemble Kalman filter (EnKF) for automatic history matching. EnKF is a Monte Carlo method, in which an ensemble of reservoir state variables are generated and kept up-to-date as data are assimilated sequentially. The uncertainty of reservoir state variables is estimated from the ensemble at any time step. Two synthetic problems are selected to investigate two primary concerns with the application of the EnKF. The first concern is whether it is possible to use a Kalman filter to make corrections to state variables in a problem for which the covariance matrix almost certainly provides a poor representation of the distribution of variables. It is tested with a one-dimensional, two-phase waterflood problem. The water saturation takes large values behind the flood front, and small values ahead of the front. The saturation distribution is bimodal and is not well modeled by the mean and variance. The second concern is the representation of the covariance via a relatively small ensemble of state vectors may be inadequate. It is tested by a two-dimensional, two-phase problem. The number of ensemble members is kept the same as for the one-dimensional problem. Hence the number of ensemble members used to create the covariance matrix is far less than the number of state variables. We conclude that EnKF can provide satisfactory history matching results while requiring less computation work than traditional history matching methods. [DOI: 10.1115/1.2134735]

Introduction

History matching is the process of adjusting the variables in a reservoir simulation model so that the computed values of observables such as rates, or pressures, or saturations, at individual wells are in reasonable agreement with actual measurements of those quantities. To do this, one typically attempts to minimize the square of the mismatch of all measurements and computed values. Great progress has been made towards the automation of the adjustment procedure, and it is now possible to perform a history match of multiphase flow data at the cost of approximately 100 reservoir simulation runs.

The increase in deployment of permanent sensors for monitoring pressure, temperature, resistivity, or flow rate has added impetus to the related problem of continuous model updating. Instead of simultaneously using all recorded data to generate an appropriate reservoir flow model, it has become important to incorporate the data as soon as they are obtained so that the reservoir model is always up-to-date.

The Kalman filter has historically been the most widely applied method for assimilating new measurements to continuously update the estimate of state variables. Kalman filters have occasionally been applied to the problem of estimating values of petroleum model variables [1,2], but they are most appropriate when the problems are characterized by relatively small numbers of variables and when the variables to be estimated are linearly related to the observations. Most data assimilation problems in petroleum reservoir engineering are highly non-linear and are characterized by many variables, often two or more variables per simulator gridblock.

The problem of weather forecasting is in many respects similar to the problem of predicting future petroleum reservoir performance. The economic impact of inaccurate predictions is substantial in both cases, as is the difficulty of assimilating very large data sets and updating very large numerical models. One method that has been recently developed for assimilating data in weather forecasting is ensemble Kalman filtering [3–8]. The method is now beginning to be applied for data assimilation in groundwater hydrology [9] and in petroleum engineering [10–16].

Background on Kalman Filters

The Kalman filter is usually used for state estimation in systems that evolve with time. The methodology consists of a forecast step (stepping the system forward in time) and an assimilation step in which variables describing the state of the system are corrected to honor the observations. Because this paper concerns the application of the Kalman filter to the problem of reservoir model updating, instead of using the traditional notation of Kalman filters, notation from Bayesian history matching will be used to explain the assimilation step. The forecast step in the Kalman filter for a dynamical system is unimportant for our application because that computation is performed by a numerical reservoir simulator.

In our applications, the reservoir state vector consists of all the reservoir variables that are uncertain, and that need to be specified in order to run the reservoir flow simulator. We denote the state vector as y. At any time t_k, y consists of two parts: model variables (porosities, permeabilities, saturations, and pressures) and theoretical data (well production rates, bottom-hole pressures, water oil ratios, etc.). In the notation, we neglect the time dependence of the variables and write the state vector for the reservoir model as

$$y = [m^T d^T]^T. \tag{1}$$

Note that in most history-matching applications, the collection of variables to be estimated consists only of the variables that do not change with time, e.g., porosity and permeability. Pressure and saturation would be determined from the knowledge of the static properties by solving the reservoir flow equations. Because saturations, pressures, porosities, and permeabilities, are simultaneously updated in the assimilation step, there is a potential for inconsistency, which we will discuss later.

Contributed by the Petroleum Division of ASME for publication in the JOURNAL OF ENERGY RESOURCES AND TECHNOLOGY. Manuscript received January 7, 2005; final manuscript received July 27, 2005. Assoc. Editor: Andrew K. Wojtanowicz.

For simplicity of exposition, we will temporarily assume that the relationship between the model variables and the theoretical data is linear, in which case the relationship between the observation vector and the reservoir state vector can be written as

$$d_{\text{obs}} = Gm^{\text{true}} + \epsilon = Hy^{\text{true}} + \epsilon, \tag{2}$$

where ε is a vector of (unknown) measurement errors, G is the sensitivity matrix, and $E[\varepsilon\varepsilon^T]=C_D$, which is assumed to be known. $H \in R^{N_d \times N_y}$ is the measurement operator that relates the state vector to theoretical observations. Because the theoretical observations are part of the state vector y, H is a trivial matrix with only 0 and 1 as its components. We can always arrange H as

$$H = [0|I], \tag{3}$$

where 0 is a $N_d \times (N_y - N_d)$ matrix with all 0s as entries, I is a $N_d \times N_d$ identity matrix, N_d is the number of measurements and N_y is the number of variables in the state vector, y.

The "best" estimate (the one that has maximum probability density) of y is obtained by minimizing the objective function

$$S(y) = \frac{1}{2}(Hy - d_{\text{obs}})^T C_D^{-1}(Hy - d_{\text{obs}}) + \frac{1}{2}(y - y^p)^T C_Y^{-1}(y - y^p), \tag{4}$$

where C_Y is the covariance matrix for the state vector y and y^p is the prior estimate of the variable y. Because we have assumed for this section that the relationship between data and model variables is linear (i.e., $d=Gm$), it is easy to see that

$$C_Y = \begin{bmatrix} C_M & C_M G^T \\ GC_M & GC_M G^T \end{bmatrix}, \tag{5}$$

where C_M is the covariance matrix for the vector of model variables m.

The estimate of y obtained by minimizing the objective function [Eq. (4)] is

$$\langle y \rangle = y^p + C_Y H^T (HC_Y H^T + C_D)^{-1}(d_{\text{obs}} - Hy^p), \tag{6}$$

(see, e.g., [17], page 70), or, in terms of C_M and G,

$$\langle m \rangle = m^p + C_M G^T (GC_M G^T + C_D)^{-1}(d_{\text{obs}} - Hy^p), \tag{7}$$

because

$$C_Y H^T = \begin{bmatrix} C_M G^T \\ GC_M G^T \end{bmatrix}$$

and $HC_Y H^T = GC_M G^T$. In traditional history matching, all available data are included in the data vector, d_{obs}. If the errors in the data are independent, there is nothing to prevent one from assimilating one piece of data at a time, however [17, pages 294], Oliver [18] showed that the maximum a posteriori estimates of permeability values obtained by using all pressure data simultaneously were similar to the results obtained by using one data at a time. In order to use a method that assimilates one data at a time, it is necessary to update the covariance matrix after each assimilation

$$C_{M'} = (C_M^{-1} + G^T C_D^{-1} G)^{-1} = C_M - C_M G^T (C_D + GC_M G^T)^{-1} GC_M. \tag{8}$$

There are two obvious difficulties in the application of this method to the problem of reservoir data assimilation. One is that the size of the covariance matrix for the model variables is typically very large, and the second is that the computation of the sensitivity matrix G is expensive for non-linear models and would need to be repeated for the assimilation of each datum. Also note that for non-linear models, the model covariance matrix computed from linearization of the flow equation is only approximate. This step is unnecessary when using the ensemble Kalman filter because the information needed to estimate the covariance matrix is in the ensemble.

The Ensemble Kalman Filter. The ensemble Kalman filter takes a different approach. A large number of initial state vectors (more accurately, only porosity and logarithm permeability for our case) are generated by sampling from the probability density function for the state vectors before the assimilation of any data. The jth reservoir state vector, y_j^p, is updated using the ensemble approximation to the Kalman gain. The formulas used to update the state vector are similar to the formulas used in standard parameter estimation [Eqs. (7) and (8)], as described for example in [17]. The primary difference is that the dynamic state variables such as pressure and saturation are updated as well as the static state variables such as porosity and permeability. The second difference is that the estimate of the Kalman gain matrix (or, equivalently, the sensitivity matrix and model covariance matrix) is estimated directly from the ensemble of states

$$y_j^u = y_j^p + K_e(d_j - Hy_j^p) \quad (j = 1,2, \ldots, N_e), \tag{9}$$

where K_e is Kalman gain and computed as:

$$K_e = C_{Y,e}^p H^T (HC_{Y,e}^p H^T + C_D)^{-1}. \tag{10}$$

In Eqs. (9) and (10), the superscript p denotes *prior*, meaning that the values are output from the simulator before update; u represents the values after the update. The subscript e denotes a value that is computed from the ensemble of state vectors; j counts from 1 to the number of ensemble member, N_e. Note that while Eq. (9) is essentially equivalent to Eq. (6), the EnKF approach avoids the difficulty of computing the sensitivity matrix G and the covariance matrix C_M, and instead approximates the product $C_M G^T$ directly from the ensemble.

The covariance matrix for the state variables at any time could be estimated from the ensemble using the standard statistical formula:

$$C_{Y,e}^p = \frac{1}{N_e - 1}(Y^p - \bar{y}^p)(Y^p - \bar{y}^p)^T, \tag{11}$$

where Y denotes the collection of all ensemble state vectors and is a matrix with dimension $N_y \times N_e$; $\bar{y}$ denotes the mean of state variables calculated across the ensemble members and is a vector with dimension N_y; N_y is the dimension of the state vectors.

$$Y = [y_1, y_2, \ldots, y_j, \ldots, y_{N_e}]. \tag{12}$$

In more detail, any element $c_{m,l}$ in the covariance matrix $C_{Y,e}^p$ can be computed as the following

$$c_{m,l} = \frac{1}{N_e - 1}\sum_{j=1}^{N_e}(x_{m,j} - \bar{x}_m)(x_{l,j} - \bar{x}_l) \quad (m,l = 1,2, \ldots, N_y), \tag{13}$$

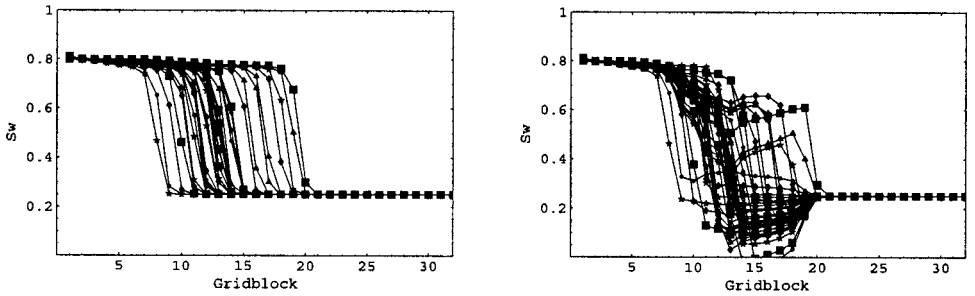
where $x_{m,j}$ and $x_{l,j}$ are the mth and lth variables, respectively, in the state vectors for the jth ensemble member; $\bar{x}_m$ and $\bar{x}_l$ are the means of the mth and lth variables, respectively, in the state vector, and are calculated across the ensemble members; $c_{m,l}$ is the covariance between the mth and lth variables in the state vector.

We can compute the covariance matrix at any assimilation time step only from that time instant and do not need to advance the covariance matrix as in a standard Kalman filter shown in Eq. (8). In practice, it is not necessary to compute an approximation of the covariance matrix, because we only require the product $HC_{Y,e}^p$.

Application of the Ensemble Kalman Filter

The following outline is intended to provide a general idea of the basic computations and the reasons for the computational efficiency.

1. Advance the state vectors using the numerical reservoir simulator to the time for data assimilation.
2. Assemble the state vectors for each ensemble member, y_j $(j=1, \ldots N_e)$.



(a) Before Kalman correction (b) After Kalman correction

Fig. 1 Water saturation profiles before and after first application of the Kalman correction at 110 days

$$y_j = [\phi_{j,1}, \ldots, \phi_{j,N}, \ln K_{j,1}, \ldots, \ln K_{j,N}, p_{j,1}, \ldots, p_{j,N}, S_{j,1}, \ldots, S_{j,N}, d_{j,1} \ldots, d_{j,N_d}]^T \quad (14)$$

where N is the number of gridblocks.

3. Compute the vector of mean values of state variables.

$$\bar{y} = \frac{1}{N_e} \sum_{j=1}^{N_e} y_j \quad (15)$$

4. Compute, ΔY, the matrix of deviations of state variables from the mean. The jth column of ΔY is

$$\Delta y_j = y_j - \bar{y} \quad (16)$$

5. Compute the product of the transpose of the matrix of deviations, ΔY^T, with the transpose of the matrix, H^T. In fact, this requires only the selection of a few columns of ΔY^T.

$$A = \Delta Y^T H^T \quad (17)$$

6. Compute the Kalman gain matrix.

$$K_e = \frac{1}{N_e - 1} \Delta Y A \left(\frac{1}{N_e - 1} A^T A + C_D \right)^{-1} \quad (18)$$

7. Update the state vectors.

$$y_j^u = y_j^p + K_e (d_{\text{obs}} - H y_j^p) \ (j = 1, 2, \ldots, N_e) \quad (19)$$

8. If there are additional data, return to step 1.

Examples

There are at least two primary concerns with the application of the ensemble Kalman filter to the problem of updating reservoir flow models. One is that whether it is possible to use a Kalman filter to make corrections to state variables in a problem for which the covariance almost certainly provides a poor representation of the distribution of variables. The second concern should be the representation of the covariance via a relatively small ensemble of state vectors may be inadequate.

Two synthetic problems are selected to investigate the potential difficulties. The first one is a one-dimensional, two-phase waterflood problem with the first data assimilation time occurring at a fairly late time. The water saturations take large values behind the water flood front, and small values ahead of the front. The saturation distribution is bimodal in this case and is not well modeled by the mean and variance. The key issue in this case is the ability to update the water saturations realistically. In the second problem, the reservoir model is two-dimensional so that the number of state variables is increased substantially. The number of ensemble members is kept the same as for the one-dimensional problem, and the main concern is that the number of ensemble members used to create the covariance matrix is far less than the number of state variables.

In both cases, the wells are constrained by bottom-hole pressure. The producers are produced at a constant pressure of 1900 psig (1 psi=6,894.76 pascal) and injectors are injected at a constant pressure of 4500 psig. The reservoir gridblocks are uniform. The dimension of a gridblock is 60×60 *feet*^{2}.

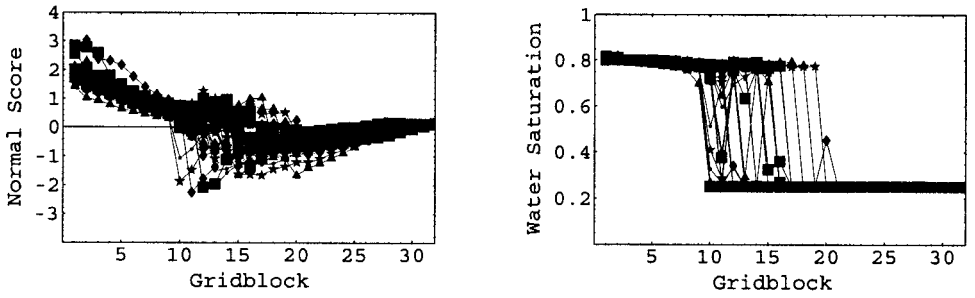
Example 1: One-Dimensional Waterflood. The reservoir model in this test case is 32 gridblocks in length. Water is injected in gridblock 1 and fluids are produced from gridblock 32. Forty nonuniform reservoir models are generated from a multi-normal distribution. The porosity is assumed to have a mean value of 0.2 and standard deviation of 0.04. The permeability is log-normally distributed. $\ln K$ has a mean of 5.5 (when K is in md. 1 md $=0.9896233 \times 10^{-12}$ m^2) and standard deviation of 0.5. The two variables are correlated with cross-correlation coefficient of 0.5 and the distributions of both variables are described by the exponential covariance model with range of 18 gridblocks.

The data to be assimilated in this problem are the water injection rate, the water and oil production rates, and the water saturation in gridblock 21. The data are obtained every 10 days starting at day 110. Measurement errors in the rates are assumed to have a standard deviation of 1 barrel per day (bpd) and measurement error in the saturation is assumed to have a standard deviation of 0.01.

From Fig. 1, we can see that prior to the first application of Kalman correction, there is considerable variation in the location of the saturation fronts. After the correction, two obvious problems with the methodology are apparent. One is that some water saturations obtained nonphysical values, for example $S_w < 0$ for some gridblocks after the water front. The other is the water saturation values do not always decrease monotonically from high at the injector to low at the producer.

Three methods of resolving the problems with water saturations are considered in the paper: (1) transformation of the saturation to a variable whose univariate distribution is normal, (2) use of the location of water shock front, instead of saturations as a state variable, and (3) iterating the update.

Normal Score Transform. One obvious way to avoid the problem with nonphysical values of saturation is to use the normal score transform values [19] of the saturations instead of the saturations themselves as state variables. The non-parametric transformation is constructed from the empirical cdf of gridblock saturations from the reservoir simulator. The variable after transformation is normally distributed. Figure 2 shows the results of applying the ensemble Kalman filter to the transformed variable. From Fig. 2(*b*), we can see that after applying the normal score transform, saturations do not go outside of reasonable bounds. Unfortunately, the saturations still oscillate spatially between high and low values. The oscillations are not very obvious in Fig. 2(*b*), but become more pronounced at later times.



(a) The transformed variable. (b) Saturation after back transform.

Fig. 2 The transformed variable and water saturation profiles after first application of the Kalman correction at 110 days

Saturation Front Location as a State Variable. Instead of applying an explicit transformation to the water saturations, we replaced the saturations in the state vector [Eq. (14)] with a variable which locates the shock front. The idea is to improve the linearity of the relationship between variables. The jth state vector now becomes

$$y_j = [\phi_{j,1}, \ldots, \phi_{j,N}, \ln K_{j,1}, \ldots, \ln K_{j,N}, p_{j,1}, \ldots, p_{j,N}, x_{\mathrm{Front},j}, d_{j,1} \ldots, d_{j,N_d}]^T. \quad (20)$$

The Kalman update formula is then used to adjust the location. Saturation values at gridblock locations are computed by interpolation based on a normalized table of saturation versus distance, computed for each reservoir realization. Figure 3(*a*) shows the water saturation profiles after the first Kalman correction. The saturation values are within reasonable bounds and change monotonically from high at the injector to low at the producer.

Iterating the Update. The front location method is applicable to one-dimensional problem. However, for two- and three-dimensional problems, it is not straightforward to describe the shock front locations in the state vector. We attempted to use a more general approach. Wen and Chen [15] used a confirmation check in their application to assure the consistency between the updated static and dynamic variables. Using a similar idea, whenever the updated saturations are detected out of physical bound (in this paper, we used $0.0 < S_w < 1.0$), we rerun the simulator from previous time step to recompute the dynamic variables (pressure and saturation) using the updated static variables (porosity and permeability). The difference of the new computational data and the observation is then used to update the state variables. The extra iteration step is repeated until the corrected saturation profiles satisfy the physical bound, or the iteration exceeds a maximum number (we used twice). We noticed that, in this problem, the extra iterations are required when the first data were assimilated (at 110 days), when the water front reached the observation well in gridblock 21 (at 340 days), and some times following the water break-through at the observation well. Figure 3(*b*) displays the water saturation profiles at 110 days after correction with 1 extra iteration step. Comparing with Fig. 1(*b*), the saturation profiles are improved.

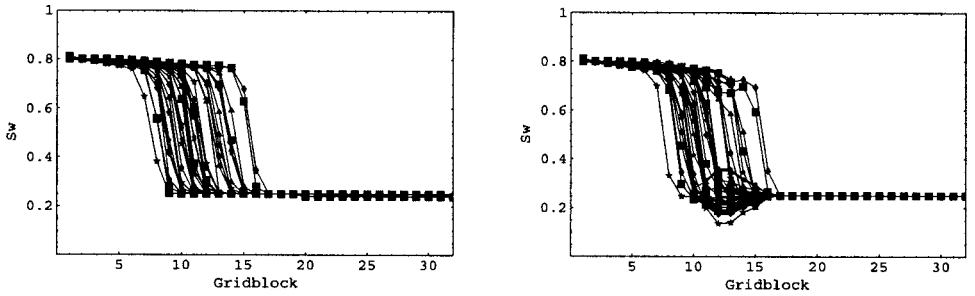
Figures 4 and 5 compare the computed water and oil production rates (darker color) with the corresponding observed values (lighter color), without and with iteration, respectively. From Fig. 4, we can see that the timing of water break-through is not captured without iteration. However, with iteration steps, shown in Fig. 5, not only do the saturation profiles get more realistic [see Fig. 3(*b*)], but also the data are better honored.

Figure 6 plots the root mean square (rms) errors of the estimate of porosity and natural logarithm permeability. It is computed as

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\hat{x}_i - x_i^{\mathrm{true}})^2}. \quad (21)$$

where $\hat{x}$ can be the mean estimate of porosity or natural logarithm permeability; x^{true} is the corresponding true value of the estimate; i is the index for the gridblock; N is the total number of the gridblocks. rms measures the deviation of the estimate from the truth at each gridblock in an average sense. It can reflect the improvement of the estimate with time when decreasing.

Because of the boundary conditions on flow (fixed pressure), and the type of data (rates), the observations do not inform about porosity until the rate of water advance can be observed. As the permeability is adjusted to honor the rate measurements, the porosity changes slightly because of its correlation with permeability. Unfortunately, the correlation appears to be moving the porosity away from the truth in this example. Figure 6 demonstrates the



(a) With shock front in state vector. (b) With one extra iteration update.

Fig. 3 110 days, saturation profiles after application of the Kalman correction

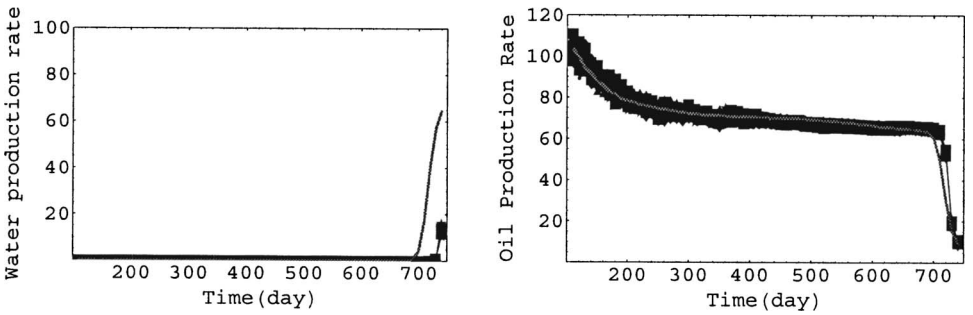


Fig. 4 The observed and computed water and oil production rates from all ensemble members after Kalman correction, without iteration

failure of the assimilation at the water break-through time at the producer when no iteration is applied. With iteration, the estimate of permeability improves steadily as more data are assimilated, while the error in the porosity estimate decreases at early time, then begins to increase after water arrival at observation well and the saturation datum is assimilated (at 340 days).

Example 2: Two-Dimensional Problem. A larger test is created to investigate the performance of the ensemble Kalman filter on a problem for which the number of state variables is far greater than the number of ensemble members. The grid for this test problem is 16×16, so the number of elements in the state vector is 4×256 plus the number of data. We again use 40 ensemble members to represent the covariance between variables.

The true permeability and porosity fields are shown in Fig. 7. Both fields are generated using an exponential variogram model with a range of 30 gridblocks in the x-direction and 15 gridblocks in the y-direction. The mean values of the natural logarithm of permeability (md) and porosity are 5.5 and 0.2 respectively, and the standard deviations are 0.5 and 0.02, respectively. The correlation coefficient between log-permeability and porosity is 0.5.

An injector is located approximately in the center of the reservoir, and four producers are located at the four corners. The data consist of oil and water production rates at the producers and the water injection rate at the injector and are acquired every 10 days beginning at day 10. The standard deviation of the measurement errors in rates is assumed to be 1 bpd.

Figure 8 shows comparisons of the oil rates at the four production wells used for assimilation (lighter color), and the predicted rates from each of the ensemble members (darker color). The oil rate at producer 1 drops fairly rapidly after 170 days because water breaks through to the well at that time. The agreement between observations and predictions appears to be good, and the spread of the realizations is reasonable.

In the Kalman filter application, the permeability and porosity fields are continuously updated as data are assimilated. It is important that the corrections to the permeability and porosity fields be done in such a way that the plausibility of the fields, as indicated by the spatial autocorrelation, is maintained. Figure 9 shows

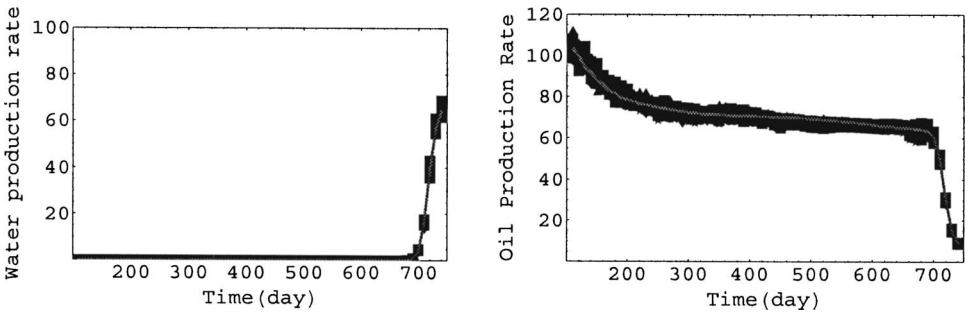


Fig. 5 The observed and computed water and oil production rates from all ensemble members after Kalman correction, with iteration

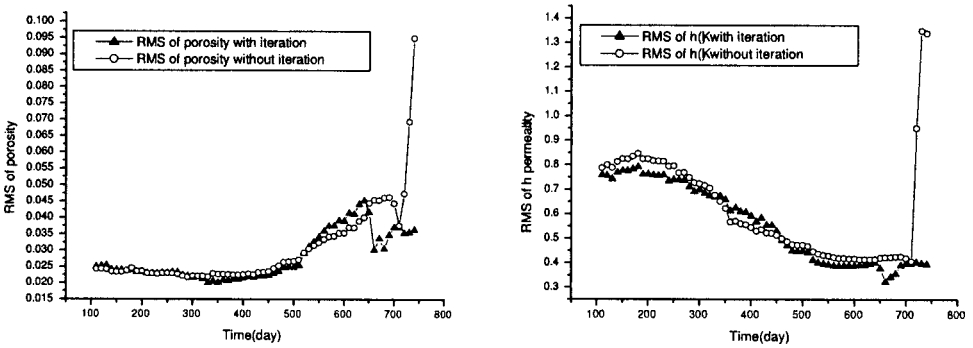


Fig. 6 RMS error of porosity (left) and natural logarithm permeability (right) with and without iteration

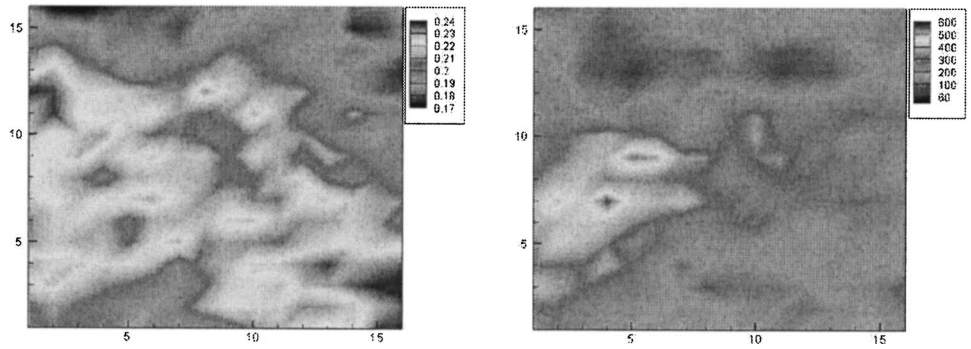


Fig. 7 The true porosity (left) and permeability (right) fields

both the initial permeability field for model 1 (the first of the 40 ensemble members), and the permeability field after the first data assimilation. Figure 10 shows the porosity and permeability fields for model 1 after assimilation of 200 days of production data. These fields should be compared with the true fields in Fig. 7. Note in particular, the general reduction in permeability in model 1 that has occurred in the upper left quadrant. While the fields are not identical, it appears that the final property fields are not unrealistic, and that the changes are largely reasonable.

In the two-dimensional test, we include saturations in state vectors, even though they causes problems in the one-dimensional test problem. Figure 11 shows the saturation profile on a cross-section through the injector at 10 days. The saturations after the update are mostly reasonable—we do not see profiles that are not decreasing away from the injector. The only indication of potential problem is the saturation on one profile that is slightly below

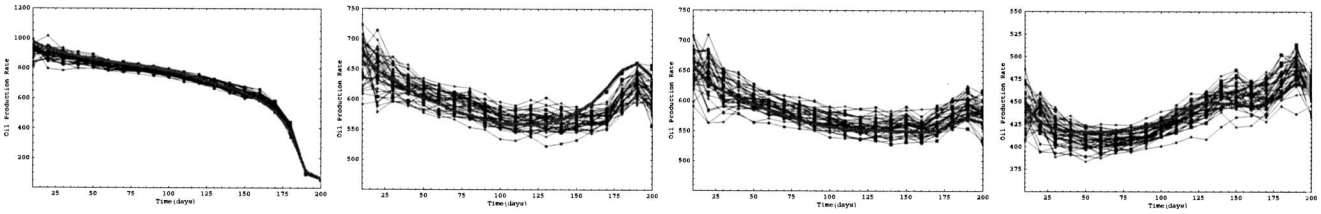


Fig. 8 Oil production rates at the 4 producers after Kalman correction

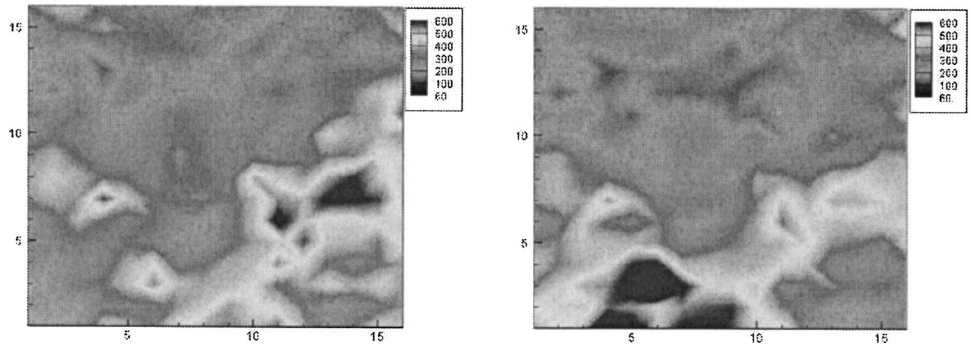


Fig. 9 The permeability field for model 1 before (left) and after (right) kalman correction at 10 days

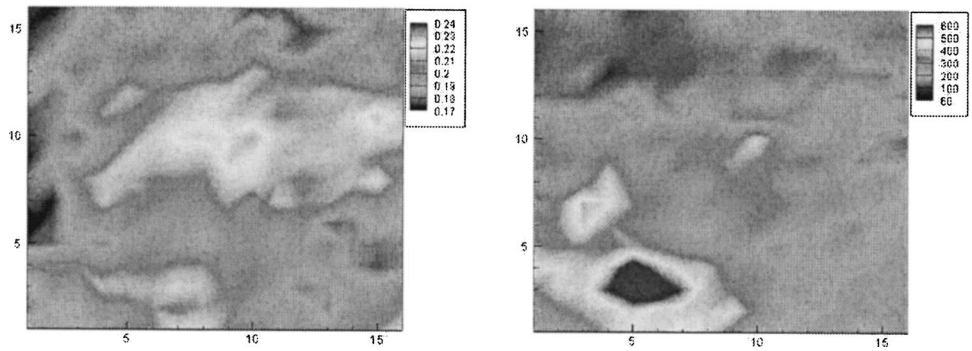


Fig. 10 The final porosity (left) and permeability (right) fields for model 1 after 200 days of Kalman correction

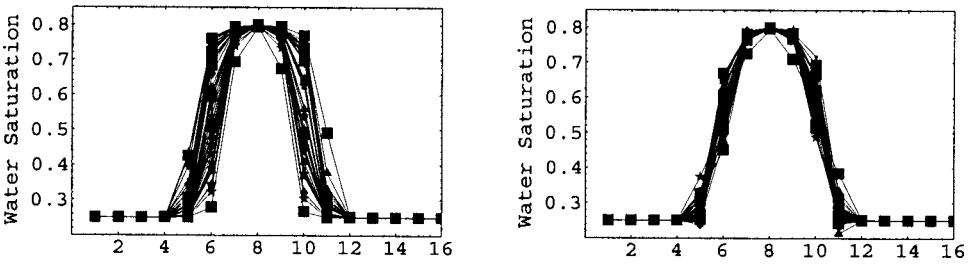


Fig. 11 10 days, water saturation at the middle column before (left) and after (right) Kalman correction

the initial water saturation. In this case, it does not appear to be necessary to change variables or iterate. We believe that this will generally be the case when changes to the state variables in the update step are relatively small.

Material Balance Check. In traditional history matching, only the permeability and porosity fields would be updated using an optimizer, then the pressure and saturation fields that are consistent with the permeability and porosity fields would be computed by running the reservoir simulator. In the Kalman filter method, the permeability, porosity, pressure, and saturation are all updated simultaneously. In essence a linearized approximation to the simulator is used to make a prediction of the saturation changes and pressure changes that would result from the porosity and permeability changes.

It is not at all clear, however, that material balance is honored during the Kalman correction steps. In order to check the global material balance error, we compare the total water-in-place at the end of the 200 days for reservoir models that have been continuously updated, to reservoir models that use the final permeability and porosity fields, but use the reservoir simulator to compute water-in-place. Figure 12(*a*) shows a histogram of the magnitude of the error for each ensemble member. Most errors are less than 0.8%. For comparison, the distribution of water-in-place in million barrel (1 MB=158,987.3 m^3) for the 40 ensemble members is shown in Fig. 12(*b*). Of course the errors in individual gridblocks can be fairly large.

One feature of the Kalman filter that can be disconcerting when used for reservoir model updating, is that the permeability and porosity fields are changing with time as additional data are assimilated. It seems quite possible that the final permeability and porosity fields that result from 200 days of data assimilation, might no longer honor the data from the previous observations. We have already seen that the rock property fields are much different after 200 days than they are after 10 days (compare Figure 10 to Figure 9). In Fig. 13, we see a comparison of the actual data that are used in the assimilation, with the predicted values of the measurements from each of the ensemble permeability and porosity fields at 200 days. In this case, the data seem to be honored at all times.

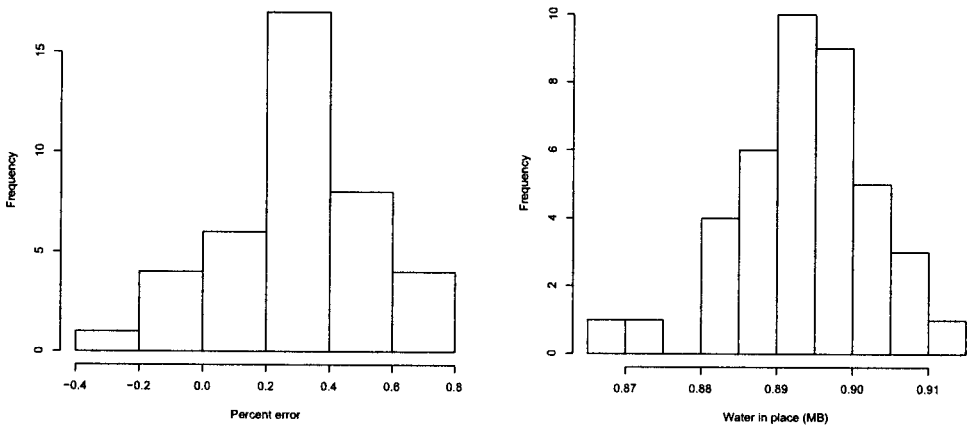
Conclusions

When the distribution of one of the state variables (saturation) is very far from normal, we find that it is possible to generate non-physical values during the Kalman update step. If necessary, it could be resolved by a change of variables, or iteration. It is not necessary to make a change of variables for the two-dimensional problem, presumably because the changes in model variables in any update step are relatively small.

We also find that the final model, at the end of data assimilation, while much different from the earlier models, is still consistent with the early data. The material balance errors are relatively small, at least compared to the uncertainty in the actual values. This is encouraging for the use of the method in reservoir history matching because we would want a model that honored all data.

Because of the apparent need for small corrections at any timestep, it seems that there will be some reservoir data assimilation problems that this method will not be suitable for. It does, however, seem to be ideally suited for the assimilation of data from time series, for example, data from permanent sensors.

The results with a relatively small number of ensemble members are remarkably good. The computational effort needed to obtain 40 "history-matched" models is approximately the cost of 40 simulation runs plus the overhead associated with the cost of computing the Kalman gain. Although the simulator is called at



(a) Frequency of errors in water-in-place (b) Frequency of water-in-palce

Fig. 12 Material balance checking after 200 days of data assimilation

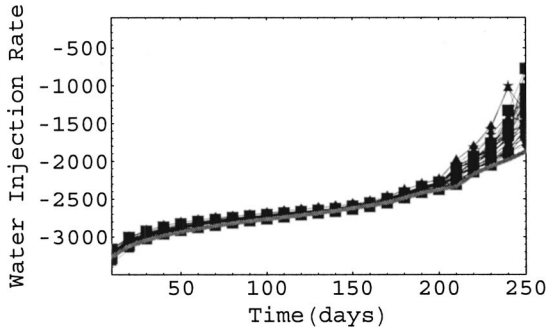


Fig. 13 Comparison of the water injection rate in each model with the data

the time of every update, it is only used to advance the simulator from the time of the last data assimilation to the time of the next data assimilation. It is not necessary to rerun the simulation from the initial time. It seems likely that larger ensembles will be required for problems with larger amounts of data to be assimilated. At this time, we do not know if the results might begin to deteriorate if the assimilation period is much longer or if the models are much larger.

Nomenclature

$c_{m,l}$ = Covariance between the mth and lth variables in the state vector
C_D = Covariance matrix of the data noises
C_M = Covariance matrix of the reservoir model variables
$C_{M'}$ = Posterior covariance matrix of the reservoir model variables
C_Y = Covariance matrix of the state vector
d = The simulated/computational data
d_{obs} = Observation data
G = Sensitivity matrix
G_m = Sensitivity matrix evaluated at m
H = Operator matrix that relates the state vector to simulated/computational data
m = Model variables
$S(y)$ = Objective function
t_k = Time step
y = The state vector
Y = Collection of the state vectors
ε = Data noise

Superscripts

$^{-1}$ = The inverse
^{p} = Values computed from the reservoir simulator before Kalman correction
$^{\text{true}}$ = The true value
^{T} = The transpose
^{u} = The values after Kalman correction

References

[1] Eisenmann, P., Gounot, M.-T., Juchereau, B., and Whittaker, S. J., 1994, "Improved Rxo Measurements Through Semi-Active Focusing, SPE-28437," in Proceedings of the SPE 69th Annual Technical Conference and Exhibition.
[2] Corser, G. P., Harmse, J. E., Corser, B. A., Weiss, M. W., and Whitflow, G. L., 2000, "Field Test Results for a Real-Time Intelligent Drilling Monitor, SPE-59227," in Proceedings of the 2000 IADC/SPE Drilling Conference.
[3] Evensen, G., 1994, "Sequential Data Assimilation With a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics," J. Geophys. Res., **99**(C5), pp. 10143–10162.
[4] Houtekamer, P. L., and Mitchell, H. L., 1998, "Data Assimilation Using an Ensemble Kalman Filter Technique," Mon. Weather Rev., **126**(3), pp. 796–811.
[5] Anderson, J. L., and Anderson, S. L., 1999, "A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts," Mon. Weather Rev., **127**(12), pp. 2741–2758.
[6] Hamill, T. M., Snyder, C., Baumhefner, D. P., Toth, Z., and Mullen, S. L., 2000, "Ensemble Forecasting in the Short to Medium Range: Report From a Workshop," Bull. Am. Meteorol. Soc. (to be published).
[7] Houtekamer, P. L., and Mitchell, H. L., 2001, "A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation," Mon. Weather Rev., **129**(1), pp. 123–137.
[8] Evensen, G., 2003, "The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation," Ocean Dyn., **53**(4), pp. 343–367.
[9] Reichle, R. H., McLaughlin, D. B., and Entekhabi, D., 2002, "Hydrologic Data Assimilation With the Ensemble Kalman Filter," Mon. Weather Rev., **130**(1), pp. 103–114.
[10] Nævdal, G., Mannseth, T., and Vefring, E. H., 2002, "Near-Well Reservoir Monitoring Through Ensemble Kalman Filter: SPE-75235," in Proceedings of SPE/DOE Improved Oil Recovery Symposium.
[11] Nævdal, G., Johnsen, L. M., Aanonsen, S. I., and Vefring, E. H., 2003, "Reservoir Monitoring and Continuous Model Updating Using Ensemble Kalman Filter," *SPE* 84372.
[12] Gu, Y., and Oliver, D. S., 2005, "History Matching of the PUNQ-S3 Reservoir Model Using the Ensemble Kalman Filter," SPE J., **10**(2), pp. 217–224.
[13] Liu, N., and Oliver, D. S., 2005, "Critical Evaluation of the Ensemble Kalman Filter on History Matching of Geologic Facies, SPE 92867," in 2005 SPE Reservoir Simulation Symposium.
[14] Gao, G., Zafari, M., and Reynolds, A. C., 2005, "Quantifying Uncertainty for the PUNQ-S3 Problem in a Bayesian Setting With RML and Enkf, SPE 93324," in 2005 SPE Reservoir Simulation Symposium.
[15] Wen, X.-H., and Chen, W. H., 2005, "Real-Time Reservoir Model Updating Using Ensemble Kalman Filter, SPE-92991," in Proceedings of the SPE Reservoir Simulation Symposium.
[16] Dong, Y., 2005, *Integration of Time-Lapse Seismic Data Into Automatic History Matching*, Ph.D. thesis, The University of Oklahoma.
[17] Tarantola, A., 1987, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, Amsterdam, The Netherlands.
[18] Oliver, D. S., 1994, "Incorporation of Transient Pressure Data Into Reservoir Characterization," In Situ, **18**(3).
[19] Goovaerts, P., 1997, *Geostatistics for Natural Resources Evaluation*, Oxford Univ. Press, New York.

Yaqing Gu *is a research assistant in the University of Oklahoma, Norman, OK. She has working experience with Chevron. Her research interests include automatic history matching, optimal control theory, and geostatistics. She holds MS degree in petroleum geology, BS degree in applied geophysics, and is currently working on her PhD degree in petroleum engineering.*

Dean Oliver *is the Eberly Family Chair Professor and the Director of the Mewbourne School of Petroleum and Geological Engineering at the University of Oklahoma. He holds a BS degree in physics and a PhD in Geophysics. Prior to joining The University of Oklahoma, he was a professor in the Petroleum Engineering department at The University of Tulsa for six years. He worked seventeen years for Chevron as a research geophysicist, a staff reservoir engineer, a research scientist in reservoir characterization. His research interests are in inverse theory, reservoir characterization and optimization.*