SPE 109808

An Iterative Ensemble Kalman Filter for Data Assimilation

Gaoming Li, SPE, U. of Tulsa and A. C. Reynolds, SPE, U. of Tulsa

Copyright 2007, Society of Petroleum Engineers

This paper was prepared for presentation at the 2007 SPE Annual Technical Conference and Exhibition held in Anaheim, California, U.S.A., 11–14 November 2007.

This paper was selected for presentation by an SPE Program Committee following review of information contained in an abstract submitted by the author(s). Contents of the paper, as presented, have not been reviewed by the Society of Petroleum Engineers and are subject to correction by the author(s). The material, as presented, does not necessarily reflect any position of the Society of Petroleum Engineers, its officers, or members. Papers presented at SPE meetings are subject to publication review by Editorial Committees of the Society of Petroleum Engineers. Electronic reproduction, distribution, or storage of any part of this paper for commercial purposes without the written consent of the Society of Petroleum Engineers is prohibited. Permission to reproduce in print is restricted to an abstract of not more than 300 words; illustrations may not be copied. The abstract must contain conspicuous acknowledgment of where and by whom the paper was presented. Write Librarian, SPE, P.O. Box 833836, Richardson, Texas 75083-3836 U.S.A., fax 01-972-952-9435.

Abstract

The ensemble Kalman filter (EnKF) is a subject of intensive investigation for use as a reservoir management tool. For strongly nonlinear problems, however, EnKF can fail to achieve an acceptable data match at certain times in the assimilation process. Here, we provide iterative EnKF procedures to remedy this deficiency and explore the validity of these iterative methods compared to standard EnKF by considering two examples, one of which is pertains to a simple problem where the posterior probability density function has two modes. In both examples, we are able to obtain better data matches using iterative methods than with standard EnKF.

In Appendix A, we enumerate the assumptions that must hold in order to show that EnKF provides a correct sampling of the probability distribution for the random variables. This derivation calls into question the common derivation in which one adds the data to the original combined state vector of model parameters and dynamical variables. In fact, it appears that there is no assurance that this trick for turning a nonlinear problem into a linear problem results in a correct sampling of the pdf one wishes to sample. However, we show that augmenting the state vector with the data results in a correct procedure for sampling the pdf if at every data assimilation step, the predicted data vector is a linear function of the combined (unaugmented) state vector and the average predicted data vector is equal to the predicted data evaluated at the average of the predicted combined state vector. Without these assumptions, we know of no way to show EnKF samples correctly. For completeness, in Appendix C, we show that each ensemble member of model parameters obtained at each step of EnKF is a linear combination of the initial

ensemble, which emphasizes the importance of obtaining a sufficiently large initial ensemble.

Introduction

The ensemble Kalman filter (EnKF) was introduced by Evensen (1994) in the context of ocean dynamics literature as a Monte Carlo approximation of the extended Kalman filter and has been extensively discussed in the weather prediction literature. EnKF was recently introduced into the petroleum engineering literature (Naevdal et al., 2002, 2003) and adapted to the problem of estimating reservoir variables or parameters (permeability and porosity fields).

Since its introduction into the petroleum engineering literature, EnKF has been investigated by a variety of researchers including Gu and Oliver (2004); Skjervheim et al. (2005); Gao et al. (2005); Liu and Oliver (2005); Wen and Chen (2005); Zafari and Reynolds (2005a); Skjervheim et al. (2006); Thulin et al. (2007) in a reservoir characterization setting. The method has also recently been applied successfully to a true field case (Evensen et al. (2007)). As shown in Gao et al. (2006), EnKF and the more computationally intense randomized maximum likelihood (RML) method gave a similar model estimate and a similar characterization of uncertainty in reservoir performance predictions for the well known PUNQ-S3 problem. For the most part, EnKF has performed well for reservoir characterization examples. However, it is relatively easy to generate toy problems with multimodal conditional pdf's for which EnKF samples very poorly and hence provides a poor assessment of uncertainty Zafari (2005); Zafari and Reynolds (2005b); Reynolds et al. (2006). Reynolds et al. (2006) also showed a small, but representative reservoir problem where EnKF has difficulty correctly assimilating water-cut data, and because of this, they designed an iterative process that combines features of randomized maximum likelihood (Oliver et al., 1996; Zhang and Reynolds, 2002; Zhang et al., 2005). Reynolds et al. (2006) also showed the EnKF update (analysis) equation is the same equation as one obtains by using RML with one Gauss-Newton iteration with a full step using the EnKF forecast (prediction) as the initial guess. Because of this result, it not surprising that it may be necessary to use an iterative procedure to obtain an acceptable match of data for highly nonlinear problems. Here, we present a detailed derivation of our current version of the Reynolds et al. (2006) algorithm and refer to this algorithm as IEnKF(1). In IEnKF(1), we simply match data sequentially in time as

in the normal EnKF procedure. However, we iterate using a gradient based algorithm to obtain a better match of data than can be obtained by EnKF. IEnKF(2) simply refers to converting from EnKF to randomized maximum likelihood when the EnKF fails to provide a good match of data and will not be discussed here. To iterate with either of these two methods, we compute at each iteration the gradient of an objective function using the adjoint procedure (Li et al., 2003; Zhang and Reynolds, 2002) which requires one forward run from time zero to the current data assimilation time and one solution of the adjoint system backward from the current time to time zero. To avoid the forward and adjoint solution from time zero, we formulate an iterative scheme, IENKF(3), which at each iteration, simply requires a forward run from the previous data assimilation time to the current data assimilation time and an adjoint solution from the current data assimilation time back to the previous data assimilation time. While this last iterative scheme is far more efficient than the first two, it requires iteratively updating the primary variables of the reservoir simulator at the previous data assimilation time. It is conceivable that this could lead to highly non-physical values which could introduce errors into our estimates. Our limited experiments show that this highly efficient method gives reasonable results, but far more testing is needed. Like EnKF and RML, one of the assumptions necessary to prove the three iterative EnKF methods sample correctly is that there is a linear relation between data and vector of model parameters.

Historically, in the atmospheric literature, predictions were made forward in time based on the ensemble of states obtained at the most recent data assimilation step. In reservoir simulation applications, it has not been clearly established whether predictions should be made forward in time using the reservoir simulator from the last updated (analyzed) ensemble of reservoir parameters and simulation primary variables or instead, should be made from time zero using the ensemble of reservoir parameters obtained at the last data assimilation step. The first procedure has the advantage of computational efficiency, but the second has the advantage that we avoid nonphysical values of pressure and saturation at all time steps and maintain material balances. The hope is that the two methods will give equivalent or at least very similar results; in some cases this is true, but in other the results can be radically different (Zhao et al., 2007). For a linear problem with no model error, a Gaussian prior model and fixed known initial conditions, Li and Reynolds (2007) have shown that running from time zero with the final ensemble of model parameters gives the same predictions of states (primary variables) as are obtained by running forward from the last data assimilation step. A refined version of this result is given in Thulin et al. (2007). Similarly, Zafari and Reynolds (2005a) have shown that for the a linear problem with a Gaussian prior, no model error and fixed initial conditions, EnKF becomes equivalent to randomized maximum

likelihood as the number of ensembles goes to infinity. Thus, in this situation, EnKF samples the correct pdf at least asymptotically. For the same situation, the same methodology can be used to show the iterative methods given here sample the pdf (Eq. 3) correctly as the ensemble size goes to infinity. While this is comforting, there is no guarantee the methods sample the pdf correctly for nonlinear, non-Gaussian problems.

Conditional PDF

The N_m -dimensional column vectors of model parameters is denoted by m . Model parameters can include reservoir gridblock permeabilities and porosities, fault transmissibilities, fluid contacts, initial fluid distributions or parameters describing relative permeabilities, but in the examples presented here only porosities and log-permeabilities are included as model parameters. We let t_i , $i = 1, 2, \dots$, denote the simulation time steps with $t_0 = 0$ and let the random N_p -dimensional column vector p^i denote the vector of dynamical variables, i.e., the primary variables of the reservoir simulation equations at time t_i , where p^0 denotes the random vector which represents the initial conditions. For a black oil system, p^i includes pressures, saturations and dissolved GOR ratios. Note that the random vector p^i depends on time but the reservoir model m , i.e., the N_m -dimensional random column vector of model parameters, m , does not. However, the joint conditional pdf for these random vectors evolves in time as more data are assimilated. In the specific examples considered here, p^0 is fixed and known; an example where this is not the case is given in Thulin et al. (2007). We also neglect model error in our examples. Boundary conditions are also assumed known. Thus, there is a deterministic relationship between the reservoir simulator and the model parameters and the primary variable in the simulator are random variables through their relation to the random vector m . We nevertheless use EnKF to sequentially update both m and the primary variables as this conceptually avoid the necessity to rerun the reservoir simulator from time zero with each updated ensemble member.

The vector d^n represents the N_n -dimensional random column vector of predicted data vector at time t_{i_n} , $n = 1, 2, \dots$, where these times denote the times at which we wish to assimilate data, i.e., condition m and the p^{i_n} 's, to observations, d_{obs}^n .

It is conceptually possible to sample the pdf $f(p^{i_n}, p^{i_n-1}, \dots, p^1, p^0, m | d_{\text{obs}}^n, d_{\text{obs}}^{n-1}, \dots, d_{\text{obs}}^1)$ using a sequential Bayesian updating scheme, and the ensemble Kalman smoother (EnKS) is applicable for such a problem (Evensen, 2006). However, our focus on the ensemble Kalman filter (EnKF) which was designed to sample and estimate the mean of the pdf $f(p^{i_n}, m | d_{\text{obs}}^n, d_{\text{obs}}^{n-1}, \dots, d_{\text{obs}}^1)$ and to make predictions forward in time from the last data assimilation time. The reservoir simulator relation between the N_n -dimensional column vector of data at the

n th data assimilation step (time t_{i_n}) is given by

$$d^n = g_n(m, p^{i_n}) = g_n(y^{i_n}), \quad (1)$$

the N_y -dimensional column vector, y^{i_n} , is given by

$$y^{i_n} = \begin{bmatrix} m \\ p^{i_n} \end{bmatrix} = \begin{bmatrix} m^n \\ p^{i_n} \end{bmatrix}, \quad (2)$$

where $N_y = N_m + N_p$. In the last expression, we have put a superscript n on m simply to denote that we are focusing on conditioning m to data at the n th data assimilation step which corresponds to time, t_{i_n} . The prior marginal pdf for m marginal pdf at t_{i_n} is represented by the ensemble of updated (analyzed) realizations of m conditioned to data at all previous data assimilation steps. Under standard assumptions, which are discussed in Appendix A, the conditional pdf we wish to sample is given by

$$\begin{aligned} f(p^{i_n}, m | d_{\text{obs}}^n, \dots, d_{\text{obs}}^1) &= \\ a \exp \left(-\frac{1}{2} \left[\frac{m - \overline{m^{n,p}}}{p^{i_n} - \overline{p^{i_n,p}}} \right]^T C_{Y^{i_n,p}}^{-1} \left[\frac{m - \overline{m^{n,p}}}{p^{i_n} - \overline{p^{i_n,p}}} \right] - \right. \\ &\quad \left. \frac{1}{2} (d^n - d_{\text{obs}}^n)^T C_{D^n}^{-1} (d^n - d_{\text{obs}}^n) \right) \\ &= a \exp \left(-\frac{1}{2} (y^{i_n} - \overline{y^{i_n,p}})^T C_{Y^{i_n,p}}^{-1} (y^{i_n} - \overline{y^{i_n,p}}) - \right. \\ &\quad \left. \frac{1}{2} (d^n - d_{\text{obs}}^n)^T C_{D^n}^{-1} (d^n - d_{\text{obs}}^n) \right). \quad (3) \end{aligned}$$

It is critical to keep in mind throughout that we use C_{D^n} to denote the measurement error covariance matrix for observed data, whereas, as discussed in detail later, we use C_{D^n, D^n} to denote the autocovariance between predicted data.

Sampling the PDF with IEnKF(1) As IEnKF(1) is a modification of randomized maximum likelihood (RML) obtained by using a finite ensemble and EnKF methodology to approximate the Hessian in the Gauss-Newton optimization algorithm, we note that the RML method (Oliver et al., 1996; Zhang and Reynolds, 2002) for generating N_e samples of the conditional pdf of Eq. 3 is given by the following procedure:

1. Sample the prior Gaussian distribution for y^{i_n} which has mean $\overline{y^{i_n,p}}$ and covariance $C_{Y^{i_n,p}}$ to obtain a set of realizations, which are denoted here by $y_j^{i_n,p}$, $j = 1, 2, \dots, N_e$.
2. Add noise to the observed data, i.e., sample the Gaussian for the data which has mean d_{obs}^n and covariance C_{D^n} to obtain a set of vectors of perturbed observations, which are denoted here by $d_{uc,j}^n$, $j = 1, 2, \dots, N_e$.

3. Apply an optimization algorithm to minimize the set of objective functions

$$O_j = \frac{1}{2} \left\{ \left[y^{i_n} - y_j^{i_n,p} \right]^T C_{Y^{i_n}}^{-1} \left[y^{i_n} - y_j^{i_n,p} \right] + (d^n - d_{uc,j}^n)^T C_{D^n}^{-1} (d^n - d_{uc,j}^n) \right\}, \quad (4)$$

for $j = 1, 2, \dots, N_e$. Here, these set of realizations are denoted by $y_j^{i_n,u}$ for $j = 1, 2, \dots, N_e$ and ideally represent samples from the pdf of Eq. 3. Although RML can be proved to sample correctly only for the case where the relation between data and model is linear, there is considerable evidence that the realizations obtained often give a reasonable approximation of the pdf of Eq. 3 even for a nongaussian multi-modal distribution.

As noted in Appendix A, the terms that involve y^{i_n} in Eq. 4 arise from the assumption that the prior model for y^{i_n} is Gaussian and represents the distribution for y^{i_n} obtained by assimilating all observed data prior to time t_{i_n} . The upper left $N_m \times N_m$ block of the matrix $C_{Y^{i_n}}$, is the covariance matrix for marginal pdf of m conditional to all data prior to t_{i_n} and is denoted by C_{M^n} . At this point, we have not yet assumed that we can approximate these covariances using a finite ensemble.

To minimize the objective function of Eq. 4 with the Gauss-Newton algorithm to obtain $y_j^{i_n,u}$, we iterate according to the following scheme:

$$\begin{aligned} &\left(C_{Y^{i_n,p}}^{-1} + [\nabla_y(d^n)^T]_\ell C_{D^n}^{-1} [\nabla_y(d^n)^T]_\ell^T \right)^{-1} \delta y_j^{i_n,\ell+1} = \\ &- [C_{Y^{i_n,p}}^{-1} (y_j^{i_n,\ell} - \overline{y^{i_n,p}}) + [\nabla_y(d^n)^T]_\ell C_{D^n}^{-1} (d^n - d_{uc,j}^n)] \end{aligned} \quad (5)$$

$$y_j^{i_n,\ell+1} = y_j^{i_n,\ell} + \alpha_j^{\ell+1} \delta y_j^{i_n,\ell+1}, \quad (6)$$

where the subscript or superscript ℓ denotes the iteration index, $\alpha_j^{\ell+1}$ is the stepsize in the search direction, $\delta y_j^{i_n,\ell+1}$, and is determined by a line search algorithm (Nocedal and Wright, 1999). Upon convergence, we set $y_j^{i_n,u}$ equal to the vector $y_j^{i_n}$ obtained at convergence.

The term $[\nabla_y(d^n)^T]_\ell$ denotes the gradient of d^n with respect to y^{i_n} evaluated at $y_j^{i_n,\ell}$ and will be denoted by the matrix $S_{n,\ell}^T$, i.e.,

$$S_{n,\ell} = [\nabla_y(d^n)^T]_\ell^T \quad (7)$$

is the $N_n \times N_y$ sensitivity matrix. A standard matrix inversion lemma (Tarantola, 1987; Chu et al., 1995), which is often known as the Sherman-Morrison-Woodbury formula (Golub and van Loan, 1989), gives

$$\begin{aligned} &\left(C_{Y^{i_n,p}}^{-1} + S_{n,\ell}^T C_{D^n}^{-1} S_{n,\ell} \right)^{-1} = C_{Y^{i_n,p}} + \\ &C_{Y^{i_n,p}} S_{n,\ell}^T (C_{D^n} + S_{n,\ell} C_{Y^{i_n,p}} S_{n,\ell}^T)^{-1} S_{n,\ell} C_{Y^{i_n,p}}. \quad (8) \end{aligned}$$

Using Eq. 8, we can rewrite Eq. 5 as

$$\begin{aligned} \delta y_j^{i_n, \ell+1} = & \\ & - \left(C_{Y^{i_n, p}} + C_{Y^{i_n, p}} S_{n, \ell}^T (C_{D^n} + S_{n, \ell} C_{Y^{i_n, p}} S_{n, \ell}^T)^{-1} S_{n, \ell} C_{Y^{i_n, p}} \right) \\ & \{ C_{Y^{i_n, p}}^{-1} (y_j^{i_n, \ell} - y_j^{n, p}) + S_{n, \ell}^T C_{D^n}^{-1} (d^n(y_j^{i_n, \ell}) - d_{uc, j}^n) \}. \end{aligned} \quad (9)$$

If we assume that the determination of p^{i_n} is deterministic given m and p^{i_n-1} , then similar to Eq. 9, we have the iterative scheme

$$\begin{aligned} \delta m_j^{n, \ell+1} = & \\ & - \left(C_{M^n} + C_{M^n} G_{n, \ell}^T (C_{D^n} + G_{n, \ell} C_{M^n} G_{n, \ell}^T)^{-1} G_{n, \ell} C_{M^n} \right) \\ & \{ C_{M^n}^{-1} (m_j^{n, \ell} - m_j^{n, p}) + G_{n, \ell}^T C_{D^n}^{-1} (d^n(y_j^{i_n, \ell}) - d_{uc, j}^n) \}, \end{aligned} \quad (10)$$

for assimilating data at the n th assimilation time, which represents RML for generating samples of the conditional pdf for m conditional to d_{obs}^n . Here, $G_{n, \ell}^T$ denotes the gradient of $(d^n)^T$ with respect to m evaluated at $m_j^{n, \ell}$ and $m_j^{n, p}$ denotes the j realization obtained by conditioning to data up to and including the $(n-1)$ st data assimilation time. As in Eq. 6,

$$m_j^{n, \ell+1} = m_j^{n, \ell} + \alpha_j^{\ell+1} \delta m_j^{n, \ell+1}. \quad (11)$$

In IEnKF(1), at a fixed data assimilation time, we use the same inverse Hessian at each iteration for all realizations, where the inverse Gauss-Newton Hessian is defined by the first equality of the following equation:

$$\begin{aligned} \tilde{H} &\equiv \left(C_{M^n} + C_{M^n} G_{n, \ell}^T (C_{D^n} + G_{n, \ell} C_{M^n} G_{n, \ell}^T)^{-1} G_{n, \ell} C_{M^n} \right) \\ &= \left(C_{M^n} + C_{M^n, D^n} (C_{D^n} + C_{D^n, D^n})^{-1} C_{D^n, M^n} \right). \end{aligned} \quad (12)$$

The second equality of Eq. 12 is an approximation and used to indicate that we will approximate the inverse Hessian using the EnKF methodology with a finite number of ensemble members. Note this result is motivated by the fact established in Reynolds et al. (2006) that $C_{M^n} G_{n, \ell}^T$, $C_{D^n} + G_{n, \ell} C_{M^n} G_{n, \ell}^T$ and $G_{n, \ell} C_{M^n}$, respectively, can be approximated by C_{M^n, D^n} , C_{D^n, D^n} and C_{D^n, M^n} . These last covariance matrices are discussed in detail when we provide the implementation of IEnKF(3). It is important to note that even though m^n changes from iteration to iteration, C_{M^n} is held fixed because the prior model is known, i.e., similar to the normal EnKF procedure, C_{M^n} is represented by the ensemble of models obtained by conditioning to all data prior to time t_{i_n} . The terms in braces in Eq. 9 represents the gradient of the objective function with respect to y^{i_n} . The term in braces in Eq. 10 represents the gradient of the objection function we minimize when we apply RML to generate samples of the pdf for m for the case where there is a deterministic relation between dynamic variables and m , no model error and known initial

conditions, which is the case considered in this work. In IEnKF(1), the term

$$G_{n, \ell}^T C_{D^n}^{-1} (d^n(y_j^{i_n, \ell}) - d_{uc, j}^n) \quad (13)$$

is calculated by using the adjoint solution from time t_{i_n} all the way back to time zero using the our standard implementation of the adjoint method (Li, 2001; Zhang and Reynolds, 2002; Gao and Reynolds, 2006). We use this procedure, however, only when the standard EnKF failed to give an “acceptable data match.”

As C_{M^n} is estimated from the ensemble, it will usually be singular, but the term $x_n \equiv C_{M^n}^{-1} (m_j^{n, \ell} - m_j^{n, p})$ can be calculated by using singular value decomposition (SVD) to solve

$$C_{M^n} x_n = (m_j^{n, \ell} - m_j^{n, p}). \quad (14)$$

In generating a solution of Eq. 14 by SVD, we do not need to form C_{M^n} explicitly; a SVD of the matrix with j th column given by m_j is sufficient to generate a solution of Eq. 14; see Appendix B for details. Once the adjoint gradient (Eq. 13) is calculated, the application of Eq. 14 requires only the inversion of the same matrix that is inverted in EnKF and the multiplication of different vectors by the matrices C_{M^n, D^n} , C_{D^n, M^n} and C_{M^n} . These matrix multiplications do not require explicit formulation of these covariance matrices and is shown in our discussion of IEnkf(3).

After assimilating data at time t_{i_n} , the updated $m_j^{n, u}$ s represent the updated covariance matrix C_{M^n} although, as in EnKF, we never explicitly compute C_{M^n} . This updated covariance provides the prior covariance for the next data assimilation. This means, that like EnKF, we are approximating the prior covariance of m by a Gaussian, but since we match data, similar to RML, we can still approximate a non-Gaussian distribution.

Sampling the PDF with IEnKF(2) In our implementation, IEnKF(1) is identical to RML using LBFGS to minimize the appropriate objective functions, (Zhang and Reynolds, 2002; Gao and Reynolds, 2006). The difference between IEnKF(1) and IEnKF(2) is that in IEnKF(2), we match all data up to the current assimilation time so the prior model is the prior model at time zero (initial ensemble), whereas in IEnKF(1), at each data assimilation step we match only observed data corresponding to the same time, and the prior model for m is the represented by the ensemble generated at the previous data assimilation step. In our implementation of the LBFGS algorithm, the prior covariance matrix represents the initial approximation to the inverse Hessian. Although LBFGS can formulated in a way that eliminates the need to solve matrix problems involving the prior covariance matrix, it would also be possible to use a finite ensemble size to represent the prior covariance matrix and further reduce the computational effort.

Sampling the PDF with IEnKF(3) In the third iterative ensemble Kalman filter (IEnKF(3)), we use the anal-

ysis step at t_{i_n} to update m and $p^{i_{n-1}}$, instead of m and p^{i_n} . As a pure EnKF method, this method has no advantage; in fact, we could use the ensemble Kalman smoother to update m , $p^{i_{n-1}}$ and p^{i_n} using observed data, d_{obs}^n , at t_{i_n} . However, by formulating the pdf for m and $p^{i_{n-1}}$ conditional to d_{obs}^n , we can derive a RML/Gauss-Newton method which computes the necessary sensitivities by applying the adjoint method over only a single time-step instead of all the way back to time zero. Thus, one iteration for assimilating data at t_{i_n} require only a forward run of the reservoir simulator from $t_{i_{n-1}}$ to t_{i_n} and one adjoint solution from time t_{i_n} backwards to time $t_{i_{n-1}}$.

The reservoir simulator equations are such that it is reasonable to assume that p^{i_n} can be considered as a function of $p^{i_{n-1}}$ so that we can replace Eq. 1

$$d^{n+1} = g_n(m, p^{i_n}) = g_n(y^{i_n}). \quad (15)$$

With the same assumptions that led to Eq. A-5, it can be shown that

$$f(p^{i_n}, m | d_{obs}^{n+1}, d_{obs}^n, \dots, d_{obs}^1) = a f(d_{obs}^{n+1} | p^{i_n}, m) f(p^{i_n}, m | d_{obs}^n, d_{obs}^{n-1}, \dots, d_{obs}^1). \quad (16)$$

If we use the same type of approach and assumptions that led to Eqs. A-23 through A-25, we obtain

$$y^{i_n, u} = y^{i_n, p} + \begin{bmatrix} C_{M^{n,p} D^{n+1}} \\ C_{P^{i_n} D^{n+1}} \\ C_{D^{n+1} D^{n+1}} \end{bmatrix} \times (C_{D^{n+1}} + C_{D^{n+1} D^{n+1}})^{-1} (d_{uc}^{n+1} - d^{n+1, p}), \quad (17)$$

$$m^{n, u} = m^{n, p} + C_{M^{n,p} D^{n+1}} \times (C_{D^{n+1}} + C_{D^{n+1} D^{n+1}})^{-1} (d_{uc}^{n+1} - d^{n+1, p}), \quad (18)$$

and

$$p^{i_n, u} = p^{i_n, p} + C_{P^{i_n} D^{n+1}} \times (C_{D^{n+1}} + C_{D^{n+1} D^{n+1}})^{-1} (d_{uc}^{n+1} - d^{n+1, p}), \quad (19)$$

As in Appendix A, we assume the likelihood part of the pdf is Gaussian and the prior model part is also Gaussian. Thus, when Eq. 15 applies, the analogue of Eq. A-29 is given by

$$f(p^{i_n}, m | d_{obs}^{n+1}, \dots, d_{obs}^1) = a \exp \left(-\frac{1}{2} \begin{bmatrix} m - \overline{m^{n,p}} \\ p^{i_n} - \overline{p^{i_n,p}} \end{bmatrix}^T C_{Y^{i_n,p}}^{-1} \begin{bmatrix} m - \overline{m^{n,p}} \\ p^{i_n} - \overline{p^{i_n,p}} \end{bmatrix} - \frac{1}{2} (d^{n+1} - d_{obs}^{n+1})^T C_{D^{n+1}}^{-1} (d^{n+1} - d_{obs}^{n+1}) \right). \quad (20)$$

Throughout, the superscript p is used to indicate the prior model. Thus, $y^{i_n, p}$ ($p^{i_n, p}$ and $m^{n, p}$) represent samples from

the pdf for y^{i_n} conditional to data d_{obj}^j , $j = n, n-1, \dots, 1$. In the computational implementation of IEnKF(3), a realization of $y^{i_n, p}$ is obtained by running the simulation from time $t_{i_{n-1}}$ to t_{i_n} using the corresponding realization of $y^{i_{n-1}}$ (denoted by $m^{n-1, a} = m^{n, p}$ and $p^{i_{n-1}, a}$) obtained after assimilating data up to d_{obs}^n . In Eq. 20,

$$C_{Y^{i_n,p}} = \begin{bmatrix} C_{M^{n,p}} & C_{M^{n,p} P^{i_n,p}} \\ C_{P^{i_n,p} M^{n,p}} & C_{P^{i_n,p}} \end{bmatrix}. \quad (21)$$

Here $C_{M^{n,p}}$ and $C_{P^{i_n,p}}$, respectively, denote the auto-covariances of $m^{n,p}$ and $p^{i_n,p}$ where $(m^{n,p}, p^{i_n,p})$ represent a sample from the prior pdf $f(p^{i_n}, m | d_{obs}^n, \dots, d_{obs}^1)$. Note similar notation applies for the other submatrices in $C_{Y^{i_n,p}}$. Note $C_{Y^{i_n,p}}$ is the prior autocovariance matrix for

$$y^{i_n} = \begin{bmatrix} m^{n,p} \\ p^{i_n,p} \end{bmatrix}. \quad (22)$$

The pdf of Eq. 20 can be written as

$$f(y^{i_n} | d_{obs}^{n+1}, \dots, d_{obs}^1) = a \exp \left(-\frac{1}{2} [y^{i_n} - \overline{y^{n,p}}]^T C_{Y^{i_n,p}}^{-1} [y^{i_n} - \overline{y^{n,p}}] - \frac{1}{2} (d^{n+1} - d_{obs}^{n+1})^T C_{D^{n+1}}^{-1} (d^{n+1} - d_{obs}^{n+1}) \right). \quad (23)$$

We let S_{n+1} be the sensitivity matrix defined by

$$S_{n+1} = [\nabla_{y^{i_n}} (d^{n+1})^T]^T = [G_{n+1} \quad B_{n+1}], \quad (24)$$

where G_{n+1} and B_{n+1} are given, respectively, by

$$G_{n+1} = [\nabla_{m^n} (d^{n+1})^T]^T, \quad (25)$$

and

$$B_{n+1} = [\nabla_{p^{i_n}} (d^{n+1})^T]^T. \quad (26)$$

To generate N_e samples of the pdf of Eq. 23 by RML, we minimize the set of objective functions

$$O_j(y^{i_n}) = \frac{1}{2} \left\{ [y^{i_n} - y_j^{i_n,p}]^T C_{Y^{i_n,p}}^{-1} [y^{i_n} - y_j^{i_n,p}] + (d^{n+1} - d_{uc,j}^{n+1})^T C_{D^{n+1}}^{-1} (d^{n+1} - d_{uc,j}^{n+1}) \right\}, \quad (27)$$

for $j = 1, 2, \dots, N_e$, where $y_j^{i_n,p}$, $j = 1, 2, \dots, N_e$ denote samples from the prior pdf for y^{i_n} .

Using ℓ as the iteration index, the Gauss-Newton method for calculating the search direction at the $(\ell+1)$ st iteration, $\delta y_j^{n,\ell+1}$, can be written as

$$\delta y_j^{n,\ell+1} = -\{C_{Y^{i_n,p}} + C_{Y^{i_n,p}} S_{n+1,j}^T (C_{D^{n+1}} + S_{n+1,j} C_{Y^{i_n,p}} S_{n+1,j}^T)^{-1} S_{n+1,j} C_{Y^{i_n,p}} \} [C_{Y^{i_n,p}}^{-1} (y_j^{n,\ell} - y_j^{n,p}) + S_{n+1,j}^T C_{D^{n+1}}^{-1} (d^{n+1} (y_j^{i_n,\ell}) - d_{uc,j}^{n+1})]. \quad (28)$$

where $S_{n+1,j}$ denotes the sensitivity evaluated at $y_j^{i_n,\ell}$.

Computational Implementation Similar to the discussion on first iterative method (IEnKF(1)), the gradient of the objective function represented by $S_{n+1,j}^T C_{D^{n+1}}^{-1} (d^{n+1}(y_j^{n,\ell}) - d_{uc,j}^{n+1})$ is calculated by the adjoint procedure. But unlike in the first iterative method, the adjoint is only apply backward from time t_{in} to t_{in-1} instead of all the way back to time zero. Moreover, instead of computing a gradient of the data mismatch part of the objective function with respect to y^{in} and instead of adjusting only m by Gauss-Newton iteration, we adjust m and p^{in} . The calculation of the adjoint is fairly straightforward and similar to the implementation given in Eydinov et al. (2007) where it was necessary to compute the gradient of the initial saturation distribution with respect to model parameters.

Here we give the computational steps that are done at every iteration of EnKF(3) for each $j = 1, 2, \dots, N_e$. We assume $y_j^{in,\ell}$ has been computed and we wish to compute $\delta y_j^{in,\ell+1}$ so that we can compute the next iterate $y_j^{in,\ell+1}$.

Step 1. Run the simulator from time t_{in-1} to time t_{in} to obtain the prediction $y_j^{in,p}$ and run the adjoint solution backward in time from t_{in} to t_{in-1} to compute $S_{n+1,j}^T C_{D^{n+1}}^{-1} (d^{n+1}(y_j^{in}) - d_{uc,j}^{n+1})$ evaluated at $y_j^{in} = y_j^{in,\ell}$. Note the prior ensemble ($y_j^{in,p}$, $j = 1, 2, \dots, N_e$) does not change during the Gauss-Newton iteration.

Step 2. Compute $C_{Y^{in,p}}^{-1} (y_j^{in,\ell} - y_j^{in,p})$ using the singular value decomposition procedure given in Appendix B. In this procedure, similar to the classical EnKF method, the covariance matrix $C_{Y^{in,p}}$ is represented by the prior ensemble but is never explicitly computed. By adding the results of steps 2 and 3, we obtain the gradients of the objective functions, i.e., the term in square brackets on the right side of Eq. 28. Here, we denote the gradient associated with ensemble j by gradient by g_j^ℓ .

Step 3. For each j , compute $u_j^\ell \equiv S_{n+1} C_{Y^{in,p}} g_j^\ell$. In this step, we use the ideas from the standard EnKF procedure, i.e., as in Reynolds et al. (2006), for each j , we approximate $S_{n+1,j}$ by the sensitivity matrix evaluated at $y_j^{in,p}$. We denote this last sensitivity by $\bar{S}_{n+1}$. That is, we replace the above definition of u_j^ℓ by the approximation

$$u_j^\ell = \bar{S}_{n+1} C_{Y^{in,p}} g_j^\ell. \quad (29)$$

In addition, we apply the Taylor series approximation

$$d^{n+1}(y_j^{in,p}) - d^{n+1}(\bar{y}^{in,p}) = \bar{S}_{n+1} (y_j^{in,p} - \bar{y}^{in,p}). \quad (30)$$

Using Eq. 30 and the standard estimator of $C_{Y^{in,p}}$ based on the prior ensemble, Eq. 29 gives

$$u_j^\ell = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left[\bar{S}_{n+1} (y_j^{in,p} - \bar{y}^{in,p}) (y_j^{in,p} - \bar{y}^{in,p})^T g_j^\ell \right] = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left[(d^{n+1}(y_j^{in,p}) - d^{n+1}(\bar{y}^{in,p})) (y_j^{in,p} - \bar{y}^{in,p})^T g_j^\ell \right]. \quad (31)$$

In our implementation, as in regular EnKF, we replace $d^{n+1}(\bar{y}^{in,p})$ by the average of the set of data predicted from the prior ensemble, i.e., the average of the set $d^{n+1}(y_j^{in,p})$, $j = 1, 2, \dots, N_e$ given by

$$\overline{d^{n+1,p}} = \frac{1}{N_e} \sum_{j=1}^{N_e} d^{n+1}(y_j^{in,p}). \quad (32)$$

Note that the only part of u_j^ℓ that changes from iteration to iteration is the gradient g_j^ℓ . Also note that to calculate u_j^ℓ simply requires computing the numbers $\alpha_k = (y_k^{in,p} - \bar{y}^{in,p})^T g_j^\ell$ for each k and then computing the vector sum,

$$u_j^\ell = \left[\frac{1}{N_e - 1} \sum_{k=1}^{N_e} \left[\alpha_k (d^{n+1}(y_k^{in,p}) - \overline{d^{n+1,p}}) \right] \right]. \quad (33)$$

Step 4. For all j , replace $S_{n+1,j} C_{Y^{in,p}}^T S_{n+1,j}^T$ by $\bar{S}_{n+1} C_{Y^{in,p}}^T \bar{S}_{n+1}^T$. Using the same procedure used above to approximate $\bar{S}_{n+1} C_{Y^{in,p}}$, we can approximate the term $S_{n+1,j} C_{Y^{in,p}} S_{n+1,j}^T$ in Eq. 28 by

$$S_{n+1,j} C_{Y^{in,p}} S_{n+1,j}^T \approx \bar{S}_{n+1} C_{Y^{in,p}} \bar{S}_{n+1}^T \approx C_{D^{n+1}, D^{n+1}}. \quad (34)$$

where $C_{D^{n+1}, D^{n+1}}$ is the autocovariance of d^{n+1} estimated as

$$C_{D^{n+1}, D^{n+1}} = \frac{1}{N_e - 1} \sum_{k=1}^{N_e} (d_k^{n+1,p} - \overline{d^{n+1,p}}) (d_k^{n+1,p} - \overline{d^{n+1,p}})^T. \quad (35)$$

At this point, the inverse matrix Eq. 28 is approximated by $(C_{D^{n+1}} + C_{D^{n+1}, D^{n+1}})^{-1}$ and we compute

$$v_j^\ell = (C_{D^{n+1}} + C_{D^{n+1}, D^{n+1}})^{-1} u_j^\ell, \quad (36)$$

by solving the matrix problem

$$(C_{D^{n+1}} + C_{D^{n+1}, D^{n+1}}) v_j^\ell = u_j^\ell, \quad (37)$$

for $j = 1, 2, \dots, N_e$.

Step 5. Again using the same average gradient for each $S_{n+1,j}$, and using the calculations of the preceding steps, Eq. 28 is approximated by

$$\delta y_j^{n,\ell+1} = -\{C_{Y^{in,p}} g_j^\ell + C_{Y^{in,p}} \bar{S}_{n+1}^T v_j^\ell\}. \quad (38)$$

Similar to Eq. 33, we approximate $C_{Y^{in,p}} \bar{S}_{n+1}^T v_j^\ell$ and $C_{Y^{in,p}} g_j^\ell$, respectively, as

$$C_{Y^{in,p}} \bar{S}_{n+1}^T v_j^\ell = \frac{1}{N_e - 1} \sum_{k=1}^{N_e} \left[(y_k^{in,p} - \bar{y}^{in,p}) (d^{n+1}(y_k^{in,p}) - \overline{d^{n+1,p}})^T v_j^\ell \right], \quad (39)$$

and

$$C_{Y^{i_n,p} g_j^\ell} = \frac{1}{N_e - 1} \sum_{k=1}^{N_e} \left[(y_k^{i_n,p} - \overline{y^{i_n,p}}) (y_k^{i_n,p} - \overline{y^{i_n,p}})^T g_j^\ell \right]. \quad (40)$$

Similar to Eq. 33, calculation of the right sides of Eqs. 39 and 40 requires only the calculation of vector inner products and vector sums.

Comments on Iteration

Finally, we should note that we recommend standard EnKF without iteration as long as data can be properly assimilated (matched.) When data are not properly matched one can change to an iterative procedure. The user has to supply his or her own criteria as to what corresponds to an acceptable match of data. In the examples presented, here we use EnKF without iteration at the n th data assimilation step (time t_{i_n}) provided the predicted data at t_{i_n} gives a value of the normalized objective function (Gao (2005)), less than 5.

Examples

In this section, we compare the performance of the standard and iterative EnKF algorithms for two examples, essentially the same as the ones considered previously by Zafari and Reynolds (2005a). The first example is a one parameter toy problem and the second one is closer to a realistic reservoir problem.

Toy problem Here $p_0 = 1$, $d_n = g_n(m, p_n)$ is identical to p_n and the discrete evolutionary equation is given by

$$d_n = p_n = p_{n-1} - \frac{9\Delta t}{2} \left(m - \frac{2\pi}{3} \right)^2. \quad (41)$$

for $n = 1, 2, \dots$, where $\Delta t = 1$ and $t_0 = 0$. The true data are generated with $m_{true} = 1.88358$, at five times ($t_j = j, j = 1, 2, 3, 4, 5$). Because of the quadratic nature of the problem (Eq. 41), the same true data can be generated with $m = 2.30521$. For each time, synthetic observed data were generated by adding to the true d_n random noise sampled from $N(0, 0.01)$. The initial prior model for m is Gaussian, $m \sim N(2.1, 0.2)$. Fig. 1(a) shows the true posterior pdf for m conditional to the five observed data. The posterior pdf shows two peaks: one approximately at $m_{true} = 1.88$ and another at $m = 2.3$. The histogram of model parameters after assimilating all the 5 data with the standard EnKF is shown in Fig. 1(b). The corresponding results obtained from our iterative EnKF algorithms (IEnKF(1), IEnKF(2) and IEnKF(3)) are shown in Figs. 1(c), 1(d) and 1(e), respectively. The standard EnKF method provides a poor approximation to the true pdf, whereas the iterative methods give improved approximations to the true pdf. Among the iterative EnKF methods, IEnKF(2) (Fig. 1(d)) gives samples which generate a pdf which is almost equivalent to the true posterior pdf. Because we go back to time zero to assimilate all the data, this method here is actually the RML method. IEnKF(1)

(Fig. 1(c)) slightly over-predicts the uncertainty in m because in assimilating data sequentially, we are still forced to assume the prior at each assimilation step can be approximated by a Gaussian. IEnKF(3) (Fig. 1(e)) yields the poorest results and largest uncertainty in the model; although it gives a pdf with two modes, similar to EnKF, the two modes are not in the correct position. As we assimilate more data using IEnKF(3), we obtain a better approximation to the true pdf as illustrated in Fig. 1(f). (The true posterior pdf after assimilating 20 data will still have the two modes at the positions shown in Fig. 1(a) and in fact, the true posterior pdf after assimilating 20 data is very similar to that shown in Fig. 1(a).) On the other hand, the pdf generated by assimilating 20 data with EnKF does not give a better approximation to the true pdf than was obtained in Fig. 1(b).

IEnKF(3) is by far the most efficient iterative EnKF scheme presented in the paper due to the fact that at each iteration, it only requires generating the forward and adjoint solution on the interval $[t_{i_{n-1}}, t_{i_n}]$ when assimilating data at time t_{i_n} . However, it generates a less accurate characterization of uncertainty than the other iterative methods. Thus, more studies are needed to carefully delineate its utility for real applications.

Synthetic reservoir problem This example pertains to a 2-dimensional horizontal reservoir with a grid system of 20×30 . The size of the reservoir is 6000×9000 feet. A detailed description of the problem can be found in Zafari and Reynolds (2005a). The true horizontal log-permeability map with the location of the six wells is shown in Fig. 2(a). Well 1 is put on production on day 1 with all the other wells being shut-in. Well 2 is then put on production after Well 1 produces for 90 days. This sequence continues until all the wells are put on production. The well production constraint is $q_o = 100$ STB/day. After 540 days, Well 1 is switched from a producer to a water injector, the injection rate is set to 1000 STB/day and all the other wells produce at $q_o = 200$ STB/day. At 5580 days, the injection rate for Well 1 is increased to 2000 STB/day and the oil production rate for all the producers is increased to 300 STB/day. At day 5850, Well 5 is shut-in because of high WOR and all the other wells remain on the previous production schedule until 7290 days. After data assimilation up to 7290 days, we predict the reservoir performance to 10000 days by maintaining the injection rate at 2000 STB/day and setting the BHPs for Well 2, Well 3, Well 4 and Well 6 equal to 2500 psi, 1000 psi, 1500 psi and 300 psi, respectively. During the prediction period, Well 5 is kept shut-in.

The observation data in this study include the flowing bottom hole pressure, producing gas-oil ratio (GOR), and water-oil ratio (WOR). The true synthetic data are generated with the CLASS (Chevron's Limited Applications Simulation System) simulator and the observation data are obtained by adding Gaussian random noise to the true synthetic data. The random Gaussian noise added is

based on standard deviations of 5% for bottom hole flowing pressure and producing GOR, and 0.1% for WOR. The initial prior for m which consists of gridblock porosities and log-permeabilities is Gaussian; see Zafari and Reynolds (2005a) for additional details.

Because IEnKF(1) requires the gradient of the objective function obtained by running the simulator from time zero to the current data assimilation time and then running the adjoint code back to time zero, it can require considerable CPU time if apply this procedure for all ensemble members all the time and do several iterations. To avoid this situation, we only apply the iteration procedure when the standard EnKF fails to assimilate data sufficiently accurately so that the normalized objective function (Gao and Reynolds, 2006) at that assimilation time is greater than 5. We use a simplified version of the Gao and Reynolds (2006) line search procedure. However, we allow only a maximum of 5 iterations and this may not be enough for some problems. Data are assimilated sequentially at 90 day intervals to generate updated descriptions of the porosity field and the isotropic permeability field. For comparison purposes, we also apply IEnKF(3) whenever the normalized objective function obtained by the normal EnKF update is greater than 5.

Fig. 2(b) shows the average log-permeability after assimilating data up to 7290 days with standard EnKF. Figs. 2(c) and 2(d) show show the average log-permeability after assimilating data up to 7290 days with two of the iterative EnKF methods. All methods give an average log-permeability which shows the basic geological features of the true permeability field, i.e., a visual comparison suggests little improvement was obtained with the iterative EnKF methods.

Fig. 3 shows the production predictions by running the simulator forward from last data assimilation time, compared to the predictions from the true model. Note the EnKF predictions for cumulative oil and water production are biased and do not span the truth, and hence do not correctly characterize the uncertainty in performance predictions. Although not shown here, the EnKF ensemble predictions of the cumulative gas production are only slightly less biased. The data match during data assimilation and prediction for the producing WOR of Well 2 is shown in Fig. 3c. As we assimilated data up to 7290 days, this result shows that EnKF did not yield a match of the water cut data; all the ensemble members have later water breakthrough than the true case and give poor predictions until $t > 9000$ days. This late breakthrough is the reason EnKF gives a high estimate of cumulative oil production and a low estimate of cumulative water production. Fig. 4 compares the production predictions generated using the final ensemble with the predictions from the true model. To obtain these predictions, we ran the simulator from time zero to 10,000 days using the ensemble obtained after assimilating data up to 7290 days; these results are even more biased but the two sets of results are not radically

different qualitatively.

Fig. 5 shows the future performance predictions obtained by running the simulator forward from last data assimilation time with IEnKF(1). Fig. 6 shows the future performance predictions obtained by running the simulator forward from zero with the final ensemble obtained from IEnKF(1). Compared to the standard EnKF, this iterative EnKF method gives a much more accurate estimate of the true future performance prediction and appears to give a reasonable characterization of the uncertainty in predictions, but the last statement can not be verified without an exhaustive sampling of the true posterior pdf, which is not feasible. Figs. 5b and 6b show a stair-step increase in the cumulative water production. The water production of the first stage comes from Well 5, the flat part on the curves corresponds to times when Well 5 is shut-in but water breakthrough has not occurred at Well 2. The cumulative water production before the horizontal section (Fig. 5b) shows an almost exact agreement between the ensemble and the truth, which occurs because this iterative EnKF method matches the WOR data of Well 5 (not shown) almost exactly.

Fig. 7 shows the future performance predictions obtained by running the simulator forward from last data assimilation time with IEnKF(3). Better future performance predictions are obtained compared to the standard EnKF, but the results are worse than those from IEnKF(1), especially the cumulative water production. Most of the ensemble members under predict the true cumulative water production. Similar to the results from the IEnKF(1), the match on the water production from Well 5 is good. However, we obtained a relatively poor match of the WOR at Well 2 (not shown). The predictions from time zero with IEnKF(3) are shown in Fig. 8.

Conclusions

1. When the standard EnKF method fails to give an adequate data match and as a result poor estimates of future predictions, EnKF(1) can be applied to obtain an improved match of data and improved predictions.
2. Unfortunately, each iteration of IEnKF(1) requires an adjoint solution back to time zero and is thus requires far more computational time than the standard EnKF method. Because of this, a highly efficient iterative method, IEnKF(3), was developed. Non-exhaustive tests with IEnKF(3) suggests it gives better estimates of future performance predictions than standard EnKF, but far less accurate than IEnKF(1). Whether a more robust iterative method can be formulated that only requires an adjoint solution from the current data assimilation time to the prior data assimilation time is an open question.

Acknowledgements

The support of the member companies of TUPREP is very gratefully acknowledged.

References

- Chu, L., A. C. Reynolds, and D. S. Oliver, Computation of sensitivity coefficients for conditioning the permeability field to well-test data, *In Situ*, **19**(2), 179–223, 1995.
- Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research*, **99**, 10,143–10,162, 1994.
- Evensen, G., The combined parameter and state problem, in *Technical Report*, Norsk Hydro Research Center, 2005.
- Evensen, G., *Data Assimilation: The Ensemble Kalman Filter*, Springer, Berlin, 2006.
- Evensen, G., J. Hove, H. C. Meisingset, E. Reiso, K. S. Seim, and S.-O. Espelid, Using the EnKF for assisted history matching of a north sea reservoir, in *Proceeding of the SPE Reservoir Simulation Symposium*, 2007.
- Eydinov, D., G. Gao, G. Li, and A. C. Reynolds, Simultaneous estimation of relative permeability curves and porosity/permeability fields by history matching production data, in *Proceeding of the 2007 Canadian International Petroleum Conference*, 2007.
- Gao, G., *Data Integration and Uncertainty Evaluation for Large Scale Automatic History Matching Problems*, Ph.D. thesis, University of Tulsa, Tulsa, Oklahoma, 2005.
- Gao, G. and A. C. Reynolds, An improved implementation of the LBFGS algorithm for automatic history matching, *SPE Journal*, **11**(1), 5–17, 2006.
- Gao, G., M. Zafari, and A. C. Reynolds, Quantifying uncertainty for the PUNQ-S3 problem in a Bayesian setting with RML and EnKF (SPE-93324), in *SPE Reservoir Simulation Symposium*, p. 19, 2005.
- Gao, G., M. Zafari, and A. C. Reynolds, Quantifying uncertainty for the PUNQ-S3 problem in a Bayesian setting with RML and EnKF, *SPE Journal*, **11**(4), 506–51, 2006.
- Golub, G. H. and C. F. van Loan, *Matrix Computations*, second edn., The Johns Hopkins University Press, Baltimore, 1989.
- Gu, Y. and D. S. Oliver, History matching of the PUNQ-S3 reservoir model using the ensemble Kalman filter, *SPE-89942*, 2004.
- Li, G. and A. C. Reynolds, Iterative enfk, in *TUPREP Research Report 24*, pp. 14–56, The University of Tulsa, 2007.
- Li, R., *Conditioning Geostatistical Models to Three-Dimensional Three-Phase Flow Production Data by Automatic History Matching*, Ph.D. thesis, University of Tulsa, Tulsa, Oklahoma, 2001.
- Li, R., A. C. Reynolds, and D. S. Oliver, History matching of three-phase flow production data, *SPE J.*, **8**(4), 328–340, 2003.
- Liu, N. and D. S. Oliver, Critical evaluation of the ensemble Kalman filter on history matching of geologic facies, SPE-92867, in *Proceedings of the 2005 SPE Reservoir Simulation Symposium*, 2005.
- Naevdal, G., L. M. Johnsen, S. I. Aanonsen, and E. H. Vefring, Reservoir monitoring and continuous model updating using ensemble Kalman filter (SPE-84372), in *2003 SPE Annual Technical Conference and Exhibition*, 2003.
- Naevdal, G., T. Mannseth, and E. H. Vefring, Near-well reservoir monitoring through ensemble Kalman filter (SPE-75235), in *Proceeding of SPE/DOE Improved Oil Recovery Symposium*, 2002.
- Nocedal, J. and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.
- Oliver, D. S., On conditional simulation to inaccurate data, *Math. Geology*, **28**(6), 811–817, 1996.
- Oliver, D. S., N. He, and A. C. Reynolds, Conditioning permeability fields to pressure data, in *European Conference for the Mathematics of Oil Recovery*, V, pp. 1–11, 1996.
- Reynolds, A. C., N. He, and D. S. Oliver, Reducing uncertainty in geostatistical description with well testing pressure data, in *Reservoir Characterization—Recent Advances*, (edited by R. A. Schatzinger and J. F. Jordan), pp. 149–162, American Association of Petroleum Geologists, 1999.
- Reynolds, A. C., M. Zafari, and G. Li, Iterative forms of the ensemble kalman filter, *Proceedings of 10th European Conference on the Mathematics of Oil Recovery*, Amsterdam, 4–7 September 2006.
- Skjervheim, J. A., G. Evensen, S. I. Aanonsen, B. O. Ruud, and T. A. Johansen, Incorporating 4D seismic data in reservoir simulation models using ensemble Kalman filter, SPE-95789, in *SPE Annual Technical Conference and Exhibition, 9-12 October, Dallas, Texas*, 2005.
- Skjervheim, J. A., B. O. Ruud, S. I. Aanonsen, G. Evensen, S. I. Aanonsen, and T. A. Johansen, Using the ensemble Kalman filter with 4D data to estimate properties and

lithology of reservoir rocks, in *Proceedings of 10th European Conference on the Mathematics of Oil Recovery*, Amsterdam, 4–7 September 2006.

Tarantola, A., *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, Amsterdam, The Netherlands, 1987.

Thulin, K., G. Li, S. I. Aanonsen, and A. C. Reynolds, Estimation of initial fluid contacts by assimilation of production data with EnKF, (SPE-109975), in *Proceedings of the 2007 SPE Annual Technical Conference and Exhibition*, 2007.

Wen, X.-H. and W. H. Chen, Real-time reservoir model updating using ensemble Kalman filter (SPE-92991), in *SPE Reservoir Simulation Symposium*, 2005.

Zafari, M., Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter,, 2005.

Zafari, M. and A. C. Reynolds, Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter, (SPE-95750), in *Proceedings of the 2005 SPE Annual Technical Conference and Exhibition*, 2005a.

Zafari, M. and A. C. Reynolds, EnKF versus RML, theoretical comments and numerical experiments, *TUPREP Research Report 22*, pp. 195–228, 2005b.

Zhang, F. and A. C. Reynolds, Optimization algorithms for automatic history matching of production data, *Proceedings of 8th European Conference on the Mathematics of Oil Recovery*, pp. 1–10, 2002.

Zhang, F., J. A. Skjervheim, A. C. Reynolds, and D. S. Oliver, Automatic history matching in a Bayesian framework: Example applications, *SPERE*, **8**(3), 214–223, 2005.

Zhao, Y., G. Li, and A. C. Reynolds, Generating facies maps by assimilating production data with the ensemble Kalman filter, in *TUPREP Research Report 24*, The University of Tulsa, 2007.

Appendix A: EnKF Derivation and Assumptions

Although the fundamental assumptions and implementation equations for EnKF are well described in (Evensen, 2005, 2006), here we present a simple derivation starting from the basic Bayesian updating formula. The derivation is a generalization of one recently presented in Skjervheim et al. (2006). Throughout, we use f to denote all pdf's, m denotes a random N_m -dimensional column vector of model parameters and p^i denotes an N_p -dimensional column vector of dynamic variables at simulator time step t_i . In a reservoir simulation setting, p^i would include all

variables that could be primary variables solved by the simulator. In a black oil setting p^i would include the grid-block values of one phase pressure, two phase saturations and dissolved gas oil ratio and wellbore pressures. The initial data is represented as p^0 . We assume zero errors in the boundary conditions for the system of flow equations. The vector d_{obs}^n is a N_n dimensional column vector of observed data at time t_{i_n} . It is not necessary that observed data exist at every simulator time-step. The vector d^n denotes data predicted at t_{i_n} corresponding to the observed data vector d_{obs}^n . Times corresponding to simulation time steps are represented by $t_0 = 0, t_1, t_2, \dots, t_i, \dots$ whereas we have observed conditioning data at times, t_{i_j} , $j = 1, 2, \dots, n$. We make three key assumptions:

Assumption 1. Data measurement errors at different times are uncorrelated. If model errors are included, we also assume model errors are uncorrelated in time.

Assumption 2. For all j , predicted data, d^j at time t_{i_j} is independent of p^ℓ for $\ell \neq i_j$, i.e., d^j is functionally dependent only on m and p^{i_j} and possibly model error.

Assumption 3. The dynamical system represents a first order Markov process so that p^i depends only on p^{i-1} and m for all i , or more specifically,

$$f(p^i | p^{i-1}, p^{i-1}, \dots, p^0, m) = f(p^i | p^{i-1}, m). \quad (\text{A-1})$$

Under the above set of assumptions, it can be shown that (Evensen, 2005, 2006)

$$\begin{aligned} f(p^{i_n}, p^{i_n-1} \dots p^1, p^0, m | d^n, \dots, d^1) &= a f(d^n | p^{i_n}, m) \\ f(p^{i_n-1}, p^{i_n-1-1}, \dots, p^1, p^0, m | d^{n-1}, d^{n-2}, \dots, d^1) &= \prod_{j=i_{n-1}+1}^{i_n} f(p^j | p^{j-1}, m). \end{aligned} \quad (\text{A-2})$$

As we are focused here on EnKF rather than the ensemble Kalman smoother, we integrate Eq. A-2 with respect to $p^0, p^1, \dots, p^{i_n-1}$ to obtain

$$\begin{aligned} f(p^{i_n}, m | d^n, \dots, d^1) &= a f(d^n | p^{i_n}, m) \\ \int \left\{ f(p^{i_n-1}, p^{i_n-1-1}, \dots, p^1, p^0, m | d^{n-1}, d^{n-2}, \dots, d^1) \right. & \\ \left. \prod_{j=i_{n-1}+1}^{i_n} f(p^j | p^{j-1}, m) \right\} dp^0, dp^1, \dots, dp^{i_n-1} &= \\ a f(d^n | p^{i_n}, m) \int \left\{ \prod_{j=i_{n-1}+1}^{i_n} f(p^j | p^{j-1}, m) \right. & \\ \left. \left[\int f(p^{i_n-1}, p^{i_n-1-1}, \dots, p^1, p^0, m | d^{n-1}, d^{n-2}, \dots, d^1) \right. \right. & \\ \left. \left. dp^0, dp^1, \dots, dp^{i_n-1-1} \right] \right\} dp^{i_n-1} \dots dp^{i_n-1} &= \end{aligned}$$

$$\begin{aligned}
& af(d^n | p^{i_n}, m) \int \left(\prod_{j=i_{n-1}+1}^{i_n} f(p^j | p^{j-1}, m) \right. \\
& \quad \left. f(p^{i_{n-1}}, m | d^{n-1}, \dots, d^1) \right) dp^{i_{n-1}} \dots dp^{i_n-1} = \\
& af(d^n | p^{i_n}, m) \int \left(\prod_{j=i_{n-1}+1}^{i_n} f(p^j | p^{j-1}, m) \right. \\
& \quad \left. f(d^{n-1}, \dots, d^1 | p^{i_{n-1}}, m) f(p^{i_{n-1}}, m) \right) dp^{i_{n-1}} \dots dp^{i_n-1} \\
& = af(d^n | p^{i_n}, m) \int \left(f(p^{i_n}, \dots, p^{i_{n-1}}, m) \right. \\
& \quad \left. f(d^{n-1}, \dots, d^1 | p^{i_{n-1}}, m) \right) dp^{i_{n-1}} \dots dp^{i_n-1} \\
& = af(d^n | p^{i_n}, m) \int \left(f(p^{i_n}, \dots, p^{i_{n-1}}, m) \right. \\
& \quad \left. f(d^{n-1}, \dots, d^1 | p^{i_n}, \dots, p^{i_{n-1}+1}, p^{i_{n-1}}, m) \right) dp^{i_{n-1}} \dots dp^{i_n-1} \\
& = af(d^n | p^{i_n}, m) \times \\
& \quad \int f(p^{i_n}, \dots, p^{i_{n-1}}, m | d^{n-1}, \dots, d^1) dp^{i_{n-1}} \dots dp^{i_n-1}.
\end{aligned} \tag{A-3}$$

Finally doing the last integration in Eq. A-3 gives

$$\begin{aligned}
f(p^{i_n}, m | d^n, \dots, d^1) = \\
af(d^n | p^{i_n}, m) f(p^{i_n}, m | d^{n-1}, \dots, d^1). \tag{A-4}
\end{aligned}$$

or given specific observations of data

$$\begin{aligned}
f(p^{i_n}, m | d_{\text{obs}}^n, \dots, d_{\text{obs}}^1) = \\
af(d_{\text{obs}}^n | p^{i_n}, m) f(p^{i_n}, m | d_{\text{obs}}^{n-1}, \dots, d_{\text{obs}}^1). \tag{A-5}
\end{aligned}$$

Eq. A-5 is a general Bayesian updating equation and does not rely on any Gaussian or linearity assumptions. Note that $f(p^{i_n}, m | d_{\text{obs}}^{n-1}, \dots, d_{\text{obs}}^1)$ is the joint pdf for (p^{i_n}, m) conditional to all observed data prior to time t_{i_n} and represents a prior for the joint pdf for (p^{i_n}, m) conditional to all observed data up to and including time t_{i_n} . If we have generated proper samples of the pdf $f(p^{i_{n-1}} | d^{n-1}, d^{n-2}, \dots, d^1)$ correctly, then running the dynamical system forward from time $t_{i_{n-1}}$ for each of these samples gives correct samples of the prior pdf $f(p^{i_n}, m | d_{\text{obs}}^{n-1}, \dots, d_{\text{obs}}^1)$.

A critical assumption in EnKF is that

Assumption 4. this prior can be represented by a Gaussian distribution, i.e.,

$$\begin{aligned}
f(p^{i_n}, m | d_{\text{obs}}^{n-1}, \dots, d_{\text{obs}}^1) = \\
a \exp \left(-\frac{1}{2} \left[\frac{m - \overline{m}_{n,p}}{p^{i_n} - \overline{p}_{n,p}} \right]^T C_{Y^{i_n,p}}^{-1} \left[\frac{m - \overline{m}_{n,p}}{p^{i_n} - \overline{p}_{n,p}} \right] \right), \tag{A-6}
\end{aligned}$$

where the superscript T always denotes the transpose, an overbar represents the mean and the superscript p refers to prior so that $C_{Y^{i_n,p}}$ is the prior covariance matrix for $[(p^{i_n})^T, m^T]^T$.

Next, using the standard Bayesian approach, once a specific observation of data at time t_{i_n} is available, we interpret $f(d_{\text{obs}}^n | p^{i_n}, m)$ as the likelihood of (p^{i_n}, m) given d_{obs}^n and denote it by $L(p^{i_n}, m | d_{\text{obs}}^n)$.

Assumption 5. We assume that the measurement error that contaminates d_{obs}^n is a realization of a Gaussian distribution with zero mean and covariance C_{D^n} so that

$$\begin{aligned}
f(d_{\text{obs}}^n | p^{i_n}, m) = L(p^{i_n}, m | d_{\text{obs}}^n) = \\
a \exp \left(-\frac{1}{2} (d^n - d_{\text{obs}}^n)^T C_{D^n}^{-1} (d^n - d_{\text{obs}}^n) \right). \tag{A-7}
\end{aligned}$$

where d^n corresponds to data at time t_{i_n} generated from the dynamical system of equations. In our examples, we neglect modeling error and then, based on the assumptions we have made, the functional form of d^n is given by

$$d^n = g_n(p^{i_n}, m) \tag{A-8}$$

where in our examples d^n is calculated from the reservoir simulator.

Using Eqs. A-6 and A-7 in Eq. A-5 gives

$$\begin{aligned}
f(p^{i_n}, m | d_{\text{obs}}^n, \dots, d_{\text{obs}}^1) = \\
a \exp \left(-\frac{1}{2} \left[\frac{m - \overline{m}_{n,p}}{p^{i_n} - \overline{p}_{n,p}} \right]^T C_{Y^{i_n,p}}^{-1} \left[\frac{m - \overline{m}_{n,p}}{p^{i_n} - \overline{p}_{n,p}} \right] - \right. \\
\left. \frac{1}{2} (d^n - d_{\text{obs}}^n)^T C_{D^n}^{-1} (d^n - d_{\text{obs}}^n) \right). \tag{A-9}
\end{aligned}$$

Assuming the two covariance matrices in the previous equation can be calculated, we still need a procedure for sampling this pdf. Randomized maximum likelihood is a good candidate, but first we consider the standard procedure for implementing EnKF.

Standard EnKF Equations As commonly described, the implementation of EnKF defines a state vector

$$y^{i_j} = \begin{bmatrix} m \\ p^{i_j} \\ d^j \end{bmatrix}. \tag{A-10}$$

for $j = 1, 2, \dots, n$. Then, under appropriate assumptions, we show that for $j = n$, the EnKF analysis equation actually represents a legitimate sampling strategy for the pdf given by

$$\begin{aligned}
f(y^{i_n} | d_{\text{obs}}^n \dots d_{\text{obs}}^1) = \\
af(d_{\text{obs}}^n | y^{i_n}) f(y^{i_n} | d_{\text{obs}}^{n-1} \dots d_{\text{obs}}^1), \tag{A-11}
\end{aligned}$$

From this viewpoint, it seem clear that sampling the pdf of Eq. A-11 for y^{i_n} will not generally yield (m, p^{i_n}) components which are samples of Eq. A-9 unless the relation between predicted data and (m, p^{i_n}) is truly linear over the data assimilation time step. This is not to say that EnKF can not be motivated without resorting to Eq. A-11, for example, EnKF has the form of the covariance minimizing analysis equation as the Kalman filter (Evensen,

2006). On the other hand, if we view EnKF from the optimization viewpoint, then EnKF is essentially equivalent to sampling with randomized maximum likelihood (RML) where we use one iteration of the Gauss-Newton method with an average sensitivity matrix (Zafari and Reynolds, 2005b; Reynolds et al., 2006).

Although starting with Eq. A-11 may not be the correct approach for obtaining a theoretical understanding of EnKF, this equation yields an easy derivation of the EnKF equations. We use it here. Recall that the column vectors m , p^{i_n} , and d^n respectively have dimensions N_m , N_p and N_n . Letting N_y denote the dimension of each y^{i_n} , it follows that

$$N_y = N_m + N_p + N_n. \quad (\text{A-12})$$

Letting

$$H = [O \quad I_{N_n}], \quad (\text{A-13})$$

where O is the $N_n \times (N_m + N_p)$ null matrix and I_{N_n} is the $N_n \times N_n$ identity matrix, we see that

$$d^n = Hy^{i_n} \quad (\text{A-14})$$

so by the “trick” of adding d_n to the state vector, we have a linear relation between the random data vector d^n and y^{i_n} .

Now one must replace Assumption 4 by the following: **Assumption 4***. the prior for y^{i_n} is Gaussian and given by

$$f(y^{i_n} | d_{\text{obs}}^{n-1} \cdots d_{\text{obs}}^1) = a \exp \left(-\frac{1}{2} (y^{i_n} - \overline{y^{i_n,p}})^T C_{Y^{i_n,p}}^{-1} (y^{i_n} - \overline{y^{i_n,p}}) \right), \quad (\text{A-15})$$

The likelihood function for y^{i_n} is now given by

$$\begin{aligned} f(d_{\text{obs}}^n | y^{i_n}) &= L(y^{i_n} | d_{\text{obs}}^n) = \\ &a \exp \left(-\frac{1}{2} (d^n - d_{\text{obs}}^n)^T C_{D^n}^{-1} (d^n - d_{\text{obs}}^n) \right) = \\ &a \exp \left(-\frac{1}{2} (Hy^{i_n} - d_{\text{obs}}^n)^T C_{D^n}^{-1} (Hy^{i_n} - d_{\text{obs}}^n) \right). \end{aligned} \quad (\text{A-16})$$

Using the preceding two equations in Eq. A-11 gives

$$\begin{aligned} f(y^{i_n} | d_{\text{obs}}^n \cdots d_{\text{obs}}^1) &= \\ &a \exp \left(-\frac{1}{2} (y^{i_n} - \overline{y^{i_n,p}})^T C_{Y^{i_n,p}}^{-1} (y^{i_n} - \overline{y^{i_n,p}}) - \right. \\ &\quad \left. \frac{1}{2} (Hy^{i_n} - d_{\text{obs}}^n)^T C_{D^n}^{-1} (Hy^{i_n} - d_{\text{obs}}^n) \right). \end{aligned} \quad (\text{A-17})$$

Because of the linearity of the operator H and the assumption that the prior is Gaussian, the pdf of Eq. A-17 can be sampled using randomized maximum likelihood (RML) (Oliver et al., 1996; Oliver, 1996; Reynolds et al., 1999). That is if $y^{i_n,p}$ is a sample of the prior pdf for y^{i_n} (Eq. A-15) and d_{uc}^n is a sample of the Gaussian pdf for d^n which has mean d_{obs}^n and covariance C_{D^n} , then a y^{i_n}

which minimizes the objective function

$$\begin{aligned} O(y^{i_n}) &= \frac{1}{2} (y^{i_n} - y^{i_n,p})^T C_{Y^{i_n,p}}^{-1} (y^{i_n} - y^{i_n,p}) + \\ &\quad \frac{1}{2} (Hy_j^{i_n} - d_{\text{uc}}^n)^T C_{D^n}^{-1} (Hy_j^{i_n} - d_{\text{uc}}^n) \end{aligned} \quad (\text{A-18})$$

is a sample of the pdf of Eq. A-17. The minimizing model is denoted by $y^{i_n,u}$ and can be obtained by setting the gradient equal to zero which gives, after considerable but standard algebra (Tarantola, 1987),

$$\begin{aligned} y^{i_n,u} &= y^{i_n,p} + \\ &\quad C_{Y^{i_n,p}} H^T (C_{D^n} + H C_{Y^{i_n,p}}^p H^T)^{-1} (d_{\text{uc}}^n - Hy^{i_n,p}), \end{aligned} \quad (\text{A-19})$$

which is the standard EnKF analysis equation. Thus, EnKF provides samples of the pdf of Eq. A-17 but there is no assurance that this provides samples of the pdf of Eq. A-9.

Letting $m^{n,p}$ denote the parameter part of $y^{i_n,p}$, we let $C_{M^{n,p}}$ denote the covariance of $m^{n,p}$, let $C_{M^{n,p}P^{i_n}}$ denote the cross covariance matrix between $m^{n,p}$ and p^{i_n} and let $C_{M^{n,p}D^n}$ denote the cross covariance matrix between $m^{n,p}$ and d^n and use similar notation for other relevant cross-covariance matrices. Then

$$C_{Y^{i_n,p}} = \begin{bmatrix} C_{M^{n,p}} & C_{M^{n,p}Y^{i_n,p}} & C_{M^{n,p}D^n} \\ C_{P^{i_n}M^{n,p}} & C_{P^{i_n}} & C_{P^{i_n}D^n} \\ C_{D^n M^{n,p}} & C_{D^n P^{i_n}} & C_{D^n D^n} \end{bmatrix}. \quad (\text{A-20})$$

Note we have used $C_{D^n D^n}$ as the autocovariance of predicted data to distinguish it from the data measurement error covariance matrix, C_{D^n} . From Eqs. A-13 and A-20, it follows that

$$C_{Y^{i_n,p}} H^T = \begin{bmatrix} C_{M^{n,p}D^n} \\ C_{P^{i_n}D^n} \\ C_{D^n D^n} \end{bmatrix} \quad (\text{A-21})$$

Similarly, from Eqs. A-13 and A-21, it follows that

$$H C_{Y^{i_n,p}} H^T = [O \quad I_{N_n}] \begin{bmatrix} C_{M^{n,p}D^n} \\ C_{P^{i_n}D^n} \\ C_{D^n D^n} \end{bmatrix} = C_{D^n D^n}. \quad (\text{A-22})$$

Using Eqs. A-21 and A-22 in Eq. A-19 and noting that $Hy^{i_n,p}$ represents predicted data $d^{n,p}$ gives

$$\begin{aligned} y^{i_n,u} &= y^{i_n,p} + \begin{bmatrix} C_{M^{n,p}D^n} \\ C_{P^{i_n}D^n} \\ C_{D^n D^n} \end{bmatrix} \times \\ &\quad (C_{D^n} + C_{D^n D^n})^{-1} (d_{\text{uc}}^n - d^{n,p}). \end{aligned} \quad (\text{A-23})$$

From the last equation, we can obtain the individual update (analysis) equations

$$\begin{aligned} m^{n,u} &= m^{n,p} + \\ &\quad C_{M^{n,p}D^n} (C_{D^n} + C_{D^n D^n})^{-1} (d_{\text{uc}}^n - d^{n,p}), \end{aligned} \quad (\text{A-24})$$

and

$$p^{i_n, u} = p^{i_n, p} + C_{P^{i_n} D^n} (C_{D^n} + C_{D^n D^n})^{-1} (d_{uc}^n - d^{n, p}). \quad (\text{A-25})$$

Because only dynamical variables are calculated (advanced in time) between data assimilation steps, we have used the superscript n on the model to indicate this is the n th update of model parameters even though they correspond to time t_{i_n} .

We now let

$$y^{i_n} = \begin{bmatrix} m \\ p^{i_n} \end{bmatrix} = \begin{bmatrix} m^n \\ p^{i_n} \end{bmatrix}. \quad (\text{A-26})$$

and again use the superscript p to indicate the prior model so that

$$y^{i_n, p} = \begin{bmatrix} m^{n, p} \\ p^{i_n, p} \end{bmatrix}. \quad (\text{A-27})$$

With the y^{i_n} defined by Eq. A-26, Eqs. A-24 and A-25 can be recombined to obtain

$$y^{i_n, u} = y^{i_n, p} + C_{Y^{i_n, p}, D^n, p} (C_{D^n} + C_{D^n, p, D^n, p})^{-1} (d_{uc}^n - d^n(y^{i_n, p})), \quad (\text{A-28})$$

where we used the superscript p on D to indicate that covariances involving data generated based on $d^{n, p} = d^n(y^{i_n, p})$.

With the notation of Eqs. A-26 and A-27, we can rewrite the pdf that we wish to sample (Eq. A-9) as

$$f(y^{i_n} | d_{\text{obs}}^n, \dots, d_{\text{obs}}^1) = a \exp \left(-\frac{1}{2} \begin{bmatrix} m^n - \overline{m^{n, p}} \\ p^{i_n} - \overline{p^{i_n, p}} \end{bmatrix}^T C_{Y^{i_n, p}}^{-1} \begin{bmatrix} m^n - \overline{m^{n, p}} \\ p^{i_n} - \overline{p^{i_n, p}} \end{bmatrix} - \frac{1}{2} (d^n - d_{\text{obs}}^n)^T C_{D^n}^{-1} (d^n - d_{\text{obs}}^n) \right), \quad (\text{A-29})$$

or simply

$$f(y^{i_n} | d_{\text{obs}}^n, \dots, d_{\text{obs}}^1) = a \exp \left(-\frac{1}{2} [y^{i_n} - \overline{y^{i_n, p}}]^T C_{Y^{i_n, p}}^{-1} [y^{i_n} - \overline{y^{i_n, p}}] - \frac{1}{2} (d^n - d_{\text{obs}}^n)^T C_{D^n}^{-1} (d^n - d_{\text{obs}}^n) \right) \quad (\text{A-30})$$

Estimation of the Mean with EnKF. EnKF attempts to provide a correct estimation of the mean of y^{i_n} (Evensen, 2006). Taking expectations in Eq. A-28, we see that the estimated mean obtained from EnKF is given by

$$\overline{y^{i_n, u}} = \overline{y^{i_n, p}} + C_{Y^{i_n, p}, D^n, p} (C_{D^n} + C_{D^n, p, D^n, p})^{-1} (d_{\text{obs}}^n - \overline{d^n(y^{i_n, p})}), \quad (\text{A-31})$$

Next, we assume that the following first order Taylor series approximation applies:

$$d^n(y^{i_n}) = d^n(\overline{y^{i_n, p}}) + \overline{S_n} (y^{i_n} - \overline{y^{i_n, p}}), \quad (\text{A-32})$$

where $\overline{S_n}$ represents the transpose of the gradient of $[d^n(y^{i_n})]^T$ evaluate at $\overline{y^{i_n, p}}$. When Eq. A-31 holds, d^n is a linear function of y^{i_n} . In this circumstance, the pdf given in Eq. A-30 is Gaussian and then ((Tarantola, 1987) its mean can be obtained by setting the gradient of the objective function

$$O_s(y^{i_n}) = \frac{1}{2} [y^{i_n} - \overline{y^{i_n, p}}]^T C_{Y^{i_n, p}}^{-1} [y^{i_n} - \overline{y^{i_n, p}}] + \frac{1}{2} (d^n - d_{\text{obs}}^n)^T C_{D^n}^{-1} (d^n - d_{\text{obs}}^n) \quad (\text{A-33})$$

equal to zero. This corresponds to minimizing this objective function. Because the objective function is quadratic in y^{i_n} , the state vector which minimizes this objective function can be found by doing one iteration of the Gauss-Newton optimization algorithm using any initial guess. Using $y^{i_n, 0}$ to denote the initial guess and $y^{i_n, \min}$ to denote the result of one Gauss-Newton iteration, we have

$$(C_{Y^{i_n, p}}^{-1} + \overline{S_n}^T C_{D^n}^{-1} \overline{S_n}) [y^{i_n, \min} - y^{i_n, 0}] = -\{C_{Y^{i_n, p}}^{-1} [y^{i_n} - \overline{y^{i_n, p}}] + \overline{S_n}^T C_{D^n}^{-1} (d^n(y^{i_n, 0}) - d_{\text{obs}}^n)\}. \quad (\text{A-34})$$

Evaluation the Taylor series of Eq. A-32 at the initial guess $y^{i_n, 0}$ and using the result in Eq. A-34 gives

$$(C_{Y^{i_n, p}}^{-1} + \overline{S_n}^T C_{D^n}^{-1} \overline{S_n}) [y^{i_n, \min} - y^{i_n, 0}] = -\{C_{Y^{i_n, p}}^{-1} [y^{i_n} - \overline{y^{i_n, p}}] + \overline{S_n}^T C_{D^n}^{-1} (d^n(\overline{y^{i_n, p}}) + \overline{S_n} (y^{i_n, 0} - \overline{y^{i_n, p}}) - d_{\text{obs}}^n)\}. \quad (\text{A-35})$$

Choosing the initial guess $y^{i_n, 0} = \overline{y^{i_n, p}}$ and using the approximation

$$d^n(\overline{y^{i_n, p}}) \approx \overline{d^n(y^{i_n, p})} = \overline{d^{n, p}}, \quad (\text{A-36})$$

Eq. A-35 becomes

$$(C_{Y^{i_n, p}}^{-1} + \overline{S_n}^T C_{D^n}^{-1} \overline{S_n}) [y^{i_n, \min} - \overline{y^{i_n, p}}] = -\{C_{Y^{i_n, p}}^{-1} [y^{i_n} - \overline{y^{i_n, p}}] + \overline{S_n}^T C_{D^n}^{-1} (\overline{d^{n, p}} - d_{\text{obs}}^n)\}. \quad (\text{A-37})$$

Using standard matrix inversion lemmas Tarantola (1987), we can write Eq. A-37 as

$$y^{i_n, \min} = \overline{y^{i_n, p}} + C_{Y^{i_n, p}} \overline{S_n}^T (C_{D^n} + \overline{S_n} C_{Y^{i_n, p}} \overline{S_n}^T)^{-1} (d_{\text{obs}}^n - \overline{d^n(y^{i_n, p})}), \quad (\text{A-38})$$

Now, we wish to show the right sides of Eqs. A-31 and Eq. A-38 are identical so that $\overline{y^{i_n, u}} = y^{i_n, \min}$ provided we

use a finite ensemble to present the prior covariance in both equations. To show the two equations give the same result for a finite ensemble, we need only show that

$$C_{Y^{i_n,p}} \bar{S}_n^T = C_{Y^{i_n,p}, D^{n,p}} \quad (\text{A-39})$$

and

$$\bar{S}_n C_{Y^{i_n,p}} \bar{S}_n^T = C_{D^{n,p}, D^{n,p}}. \quad (\text{A-40})$$

In EnKF, $C_{D^{n,p}, D^{n,p}}$ is estimated as

$$C_{D^{n,p}, D^{n,p}} = \frac{1}{N_e - 1} \sum_{k=1}^{N_e} ((d_k^{n,p} - \bar{d}^{n,p})(d_j^{n,p} - \bar{d}^{n,p})^T). \quad (\text{A-41})$$

Assuming the approximations of Eqs. A-32 and A-36 hold, Eq. A-41 gives

$$\begin{aligned} C_{D^{n,p}, D^{n,p}} &= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} ((d_j^{n,p} - d^n(\bar{y}^{i_n,p}))(d_j^{n,p} - d^n(\bar{y}^{i_n,p}))^T = \\ &= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \bar{S}_n(y^{i_n} - \bar{y}^{i_n,p})(\bar{S}_n(y^{i_n} - \bar{y}^{i_n,p})^T = \\ &= \bar{S}_n \left\{ \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (y^{i_n} - \bar{y}^{i_n,p})(y^{i_n} - \bar{y}^{i_n,p})^T \right\} \bar{S}_n^T = \\ &= \bar{S}_n C_{Y^{i_n,p}} \bar{S}_n^T, \end{aligned} \quad (\text{A-42})$$

which establishes the validity of Eq. A-40. Eq. A-39 can be verified by the same type of proof. Under the same assumptions, the same argument can be applied to show that the equation for generating an individual sample by EnKF is equivalent to using Gauss-Newton in RML to generate samples of the pdf of Eq. A-30. Thus, we have shown that with the linear relation of Eq. A-32, EnKF provides an unbiased estimate of the mean provided the approximation of A-36 applies. Moreover, as the ensemble size goes to infinity, EnKF will provide a correct sample of the pdf of Eq. A-30.

Appendix B: Singular Value Decomposition

In each iteration of IEnKF(1) and IEnKF(3), we must evaluate a term of the form $z_\ell = C_{Y^{i_n,p}}^{-1}(y_j^{n,\ell} - y_j^{n,p})$ by solving

$$C_{Y^{i_n,p}} z_\ell = y_j^{n,\ell} - y_j^{n,p}, \quad (\text{B-1})$$

for z_ℓ where $C_{Y^{i_n,p}}$ is the $N_y \times N_y$ matrix estimated from the ensemble of predictions (forecasts) as

$$C_{Y^{i_n,p}} = \frac{1}{N_e - 1} \sum_{k=1}^{N_e} [(y_k^{i_n,p} - \bar{y}^{i_n,p})(y_k^{i_n,p} - \bar{y}^{i_n,p})^T], \quad (\text{B-2})$$

where $\bar{y}^{i_n,p}$ is the average of the N_e predictions, $y_k^{i_n,p}$, $k = 1, 2, \dots, N_e$. Letting Y^p be the $N_y \times N_e$ matrix which

has $y_k^{i_n,p}$ as its k th column and letting $\bar{Y}$ be the $N_y \times N_e$ matrix which has $\bar{y}^{i_n,p}$ as its k th column, Eq. B-2 can be written as

$$C_{Y^{i_n,p}} = \frac{1}{N_e} (Y^p - \bar{Y})(Y^p - \bar{Y})^T, \quad (\text{B-3})$$

or letting $A = Y^p - \bar{Y}$, and $b_\ell = (N_e - 1)(y_j^{n,\ell} - y_j^{n,p})$, it follows that solving Eq. B-1 is equivalent to solving

$$AA^T z_\ell = b_\ell. \quad (\text{B-4})$$

We assume that $N_e < N_y$ so the $N_y \times N_y$ matrix can not have rank greater than N_e and Eq. B-4 does not have a unique solution. Thus, we must apply singular value decomposition (SVD) to find a “solution.” Let

$$A = U\Lambda V^T, \quad (\text{B-5})$$

be the SVD of A and rewrite Eq. B-4 as

$$U\Lambda\Lambda^T U^T z_\ell = b_\ell. \quad (\text{B-6})$$

The $N_y \times N_e$ matrix Λ can be represented by

$$\Lambda = \begin{bmatrix} \Lambda_{N_e} \\ O \end{bmatrix}, \quad (\text{B-7})$$

where Λ_{N_e} is the $N_e \times N_e$ diagonal matrix of singular values and O 's represents the null $(N_y - N_e) \times N_e$ null matrix. The matrices U and V are orthogonal square matrices. Note that

$$\Lambda\Lambda^T = \Lambda_{N_e}^2 \quad (\text{B-8})$$

is an $N_e \times N_e$ diagonal matrix with its j th diagonal entry given by λ_j^2 , the squared j th singular value of A . As in standard SVD solutions, we define $[\Lambda_{N_e}^2]^{-1}$ to be the diagonal matrix whose j diagonal entry is equal to zero if $\lambda_j = 0$ and is otherwise equal to $1/\lambda_j^2$. With this notation, we can write the SVD solution of Eq. B-6 as

$$z_\ell = U [\Lambda_{N_e}^2]^{-1} U^T b_\ell. \quad (\text{B-9})$$

Appendix C: Consequences of a Limited Ensemble

In EnKF, covariances are represented by the members of the finite ensemble. Letting $Y^{i_n,u}$ denote the $N_y \times N_e$ matrix with its j th column equal to $y_j^{i_n,u}$, the j th updated ensemble member at the n th data assimilation step it is easy to show that Evensen (2006); Zafari (2005); Zafari and Reynolds (2005a) the matrix of updated (analyzed) ensemble members can be represented by

$$Y^{i_n,u} = Y^{i_n,p} (I_{N_e} + \delta D). \quad (\text{C-1})$$

Here, I_{N_e} is the $N_e \times N_e$ identity matrix and δD is the $N_e \times N_e$ matrix given by

$$\begin{aligned} \delta D &= \frac{1}{N_e - 1} (D^{n,p} - \bar{D}^{n,p})^T [C_{D^n} + \\ &\frac{1}{N_e - 1} (D^{n,p} - \bar{D}^{n,p})(D^{n,p} - \bar{D}^{n,p})^T]^{-1} (D_{uc}^n - D^{n,p}), \end{aligned} \quad (\text{C-2})$$

where D_{uc}^n denotes the $N_n \times N_e$ matrix with j th column equal to $d_{uc,j}^n$, $D^{n,p}$ denotes the $N_n \times N_e$ matrix with j th column equal to $d_j^{n,p}$ and $\overline{D^{n,p}}$ denotes the $N_n \times N_e$ matrix with j th column equal to $\overline{d_j^{n,p}}$. From Eq. C-1, (Evensen, 2006; Evensen et al., 2007) notes that each updated ensemble member at the n th data assimilation step is a linear combination of the members of the ensemble of predictions. Because the dynamical system is generally nonlinear, even if there is no model error, one can not apply Eq. C-2 recursively to show that an ensemble member is a linear combination of the initial ensemble members of the combined state vector, y^{in} . However, if we consider only the model parameter part of the Y vectors, we show below that the vector of model parameters obtained at any update (analysis) step is a linear combination of the initial set of realizations of the vectors of model parameters.

Let $M^{n,p}$ denote the $N_m \times N_e$ matrix which has $m_j^{n,p}$ as its j th column and let $\overline{M^{n,p}}$ denote the matrix which has each and every column equal to $\overline{m_j^{n,p}}$. Also let $M^{n,u}$ denote the $N_m \times N_e$ matrix which has $m_j^{n,u}$ as its j th column. Then the system of equations given by Eq. A-24 for $j = 1, 2, \dots, N_e$ can be written as

$$M^{n,u} = M^{n,p} + C_{M^{n,p}D^n} (C_{D^n} + C_{D^n D^n})^{-1} (D_{uc}^n - D^{n,p}), \quad (\text{C-3})$$

where Eq. A-42 can be written as

$$C_{D^n D^n} = \frac{1}{N_e - 1} (D^{n,p} - \overline{D^{n,p}}) (D^{n,p} - \overline{D^{n,p}})^T. \quad (\text{C-4})$$

Similarly, in EnKF, the covariance matrix $C_{M^{n,p}D^n}$ is represented by

$$C_{M^{n,p}D^n} = \frac{1}{N_e - 1} (M^{n,p} - \overline{M^{n,p}}) (D^{n,p} - \overline{D^{n,p}})^T. \quad (\text{C-5})$$

But using the definitions of $\overline{M^{n,p}}$, $D^{n,p}$ and $\overline{D^{n,p}}$, it follows that

$$\begin{aligned} \overline{M^{n,p}} (D^{n,p} - \overline{D^{n,p}})^T &= \sum_{j=1}^{N_e} [\overline{m_j^{n,p}} (d_j^{n,p} - \overline{d_j^{n,p}})^T] = \\ \overline{m_j^{n,p}} \left[\left(\sum_{i=1}^{N_e} d_i^{n,p} \right) - N_e \overline{d_j^{n,p}} \right]^T &= 0. \end{aligned} \quad (\text{C-6})$$

Using Eq. C-6 in Eq. C-5 yields

$$C_{M^{n,p}D^n} = \frac{1}{N_e - 1} (M^{n,p}) (D^{n,p} - \overline{D^{n,p}})^T. \quad (\text{C-7})$$

Substituting Eqs. C-4 and C-7 into Eq. C-3 and simplifying gives

$$\begin{aligned} M^{n,u} &= M^{n,p} + \frac{1}{N_e - 1} (M^{n,p}) (D^{n,p} - \overline{D^{n,p}})^T \left[C_{D^n} + \right. \\ &\quad \left. \frac{1}{N_e - 1} (D^{n,p} - \overline{D^{n,p}}) (D^{n,p} - \overline{D^{n,p}})^T \right]^{-1} (D_{uc}^n - D^{n,p}) \\ &= M^{n,p} \left\{ I_{N_e} + \frac{1}{N_e - 1} (D^{n,p} - \overline{D^{n,p}})^T \left[C_{D^n} + \right. \right. \\ &\quad \left. \left. \frac{1}{N_e - 1} (D^{n,p} - \overline{D^{n,p}}) (D^{n,p} - \overline{D^{n,p}})^T \right]^{-1} (D_{uc}^n - D^{n,p}) \right\}. \end{aligned} \quad (\text{C-8})$$

Using Eq. C-2, Eq. C-8 can be written as

$$M^{n,u} = M^{n,p} \{ I_{N_e} + \delta D \}. \quad (\text{C-9})$$

Letting the N_e -dimensional column vector

$$\alpha^j = \begin{bmatrix} \alpha_1^j \\ \alpha_2^j \\ \vdots \\ \alpha_{N_e}^j \end{bmatrix}, \quad (\text{C-10})$$

denote the j th column of $I_{N_e} + \delta D$, it follows from Eq. C-9 and C-10 that

$$m_j^{n,u} = M^{n,p} \alpha^j = \sum_{i=1}^{N_e} \alpha_i^j m_i^{n,p}, \quad (\text{C-11})$$

for $j = 1, 2, \dots, N_e$. Thus each updated (model) ensemble member at the n th assimilation step (time t_{in}) is a linear combination of the ensemble members for m obtained at the previous assimilation step, i.e. a linear combination of the $m_j^{n,p} = m_j^{n-1,u}$, $j = 1, 2, \dots, N_e$. As this holds for all steps, it follows that each realization of m obtained in the final ensemble obtained by assimilating to all data up to the current time is a linear combination of the initial ensemble for m , m_j^0 , $j = 1, 2, \dots, N_e$. Thus, to obtain either a realization equal to the truth or the average of all ensemble members equal to the truth requires that the true model is in the space spanned by m_j^0 , $j = 1, 2, \dots, N_e$.

Figure 1: Results from EnKF algorithms.

Figure 2: Horizontal log-permeability.

Figure 3: Production predictions from last data assimilation time with standard EnKF.

Figure 4: Production predictions from time zero with standard EnKF.

Figure 5: Production predictions from last data assimilation time with IEnKF(1).

Figure 6: Production predictions from time zero with IEnKF(1).

Figure 7: Production predictions from last data assimilation time with IEnKF(3).

Figure 8: Production predictions from time zero with IEnKF(3).