Math Geosci (2012) 44:1–26
DOI 10.1007/s11004-011-9376-z

Ensemble Randomized Maximum Likelihood Method as an Iterative Ensemble Smoother

Yan Chen · Dean S. Oliver

Received: 19 May 2011 / Accepted: 22 November 2011 / Published online: 22 December 2011

Abstract The ensemble Kalman filter (EnKF) is a sequential data assimilation method that has been demonstrated to be effective for history matching reservoir production data and seismic data. To avoid, however, the expense of repeatedly updating variables and restarting simulation runs, an ensemble smoother (ES) has recently been proposed. Like the EnKF, the ES obtains all information necessary to compute a correction to model variables directly from an ensemble of models without the need of an adjoint code. The success of both methods for history matching reservoir data without iteration is somewhat surprising since traditional gradient-based methods for history matching typically require 10 to 30 iterations to converge to an acceptable minimum. In this manuscript we describe a new iterative ensemble smoother (batch-EnRML) that assimilates all data simultaneously and compare the performance of the iterative smoother with the two non-iterative methods and the previously proposed sequential iterative ensemble filter (seq-EnRML). We discuss some aspects of the use of the ensemble estimate of sensitivity, and show that by sequentially assimilating data, the nonlinearity of the assimilation problem is substantially reduced. Although reasonably good data matches can be obtained using a non-iterative ensemble smoother, iteration was necessary to achieve results comparable to the EnKF for nonlinear problems.

Keywords History matching · Ensemble Kalman filter · Ensemble smoother · Data assimilation · Iterative ensemble filter · Iterative ensemble smoother

Y. Chen (✉)
International Research Institute of Stavanger, Bergen, Norway
e-mail: yan.chen@iris.no

D.S. Oliver
Uni Centre for Integrated Petroleum Research, Bergen, Norway
e-mail: dean.oliver@uni.no

1 Introduction

History matching in reservoir engineering is the process of finding reservoir models that are consistent with historical production observations. Classical methods for automatic or computer assisted history matching methods generally have used an adjoint system to compute the gradient of a data misfit function with respect to changes in model parameters. Data from the entire history are usually included in the cost function and the minimization is iterative. Past experience has been that synthetic examples can be history matched with approximately 10 iterations using a Gauss–Newton approach (Li et al. 2003) or with approximately 30 iterations using a quasi-Newton method (Gao and Reynolds 2006) when gradients can be computed accurately from an adjoint system. Recently, ensemble-based methods such as the ensemble Kalman filter (Evensen 1994a) have gained much attention in petroleum engineering. The EnKF does not require an adjoint system to obtain the gradient and has been demonstrated to be quite flexible in the types of variable that can be estimated and the types of measurement that can be used as data (see reviews in Aanonsen et al. (2009), Oliver and Chen (2011)). Surprisingly, the EnKF seems to be able to often obtain satisfactory history matches without the need for iteration despite the fact that many iterations were often required when the exact gradient is used in classical methods.

Iterative ensemble Kalman filters (Gu and Oliver 2007; Li and Reynolds 2009; Sakov et al. 2011), like the EnKF, have been used to sequentially assimilate data for problems in which the relationship between data and model parameters is highly nonlinear. The iterative EnKF methods are able to improve the quality of data match with additional simulation runs. One of the iterative EnKF methods, the ensemble randomized maximum likelihood (EnRML) method (Gu and Oliver 2007), uses an average sensitivity estimated from the ensemble to iteratively update the model parameters using the Gauss–Newton formulation. The state variables are computed by rerunning the simulation using the updated model parameters. In a practical implementation of the EnRML, a standard EnKF update is used unless the changes in the state variables are large, in which case, the method switches to iterative updating (Gu and Oliver 2007; Chen and Oliver 2010a). The EnKF update is equivalent to the first iteration of the EnRML with a full step length except that in the EnKF both model and state variables are included in estimation.

Different from the filtering problem, a smoothing problem updates both the current state and the past states when data are assimilated. The ensemble smoother was initially introduced for the weather prediction problem (van Leeuwen and Evensen 1996), where estimation is focused on dynamic states and none or very few model parameters are involved. The ensemble smoother has been tested on a quasigeostropic model (van Leeuwen and Evensen 1996) and on a Lorentz model (Evensen 1994b). In both cases the ensemble smoother performed poorer than the EnKF. van Leeuwen and Evensen (1996) attributed the better performance of the EnKF to its sequential reduction of uncertainty at every assimilation step. Skjervheim and Evensen (2011) introduced the use of an ensemble smoother (ES) for petroleum history matching. In their implementation of the ES, data at all times are assimilated simultaneously so that the analysis step of the EnKF is only applied once to update the model variables.

The name "ensemble smoother" was used because prediction of the production and dynamic states at different times can be computed by rerunning simulations from initialization using the updated model variables. The batch update of the ES is similar to the asynchronous EnKF in Sakov et al. (2010) in which the EnKF is used to assimilate data that are not all collected at the same time.

The ES requires only a single update step, so the use of it avoids the frequent modification of restart files and restarting of simulation runs that are required by sequential updating of the EnKF. In some real applications, the increase in computational efficiency from the ES compared to the EnKF is substantial, partially because it avoids updating of state variables which may cause convergence problems with the simulator. The leading order of magnitude of floating point operations (FLOPS) for the analysis step of the ES and the EnKF is $O(\max(nN_e^2, N_d N_e^2))$ (Evensen 2003), where n is the number of variables in the state vector, N_e is the size of the ensemble, and N_d is the number of data. In combined parameter and state estimation problems, n equals to the sum of the number of model and state variables and is typically much larger than the number of data N_d. Although the number of data at the analysis step is greatly increased in the ES compared to the number of data at each of the data assimilation times of the EnKF, the leading order of magnitude of the FLOPS is the same as the EnKF if N_d remains less than n. The two examples presented in Skjervheim and Evensen (2011) both showed comparable results for the ES and the EnKF, but the computational cost in their examples for the ES was much lower than for the EnKF due to the one-step update and smoother simulation runs for the ES.

In this paper, we introduce an iterative ensemble smoother (or batch-EnRML), in which all data are assimilated simultaneously as in the ES, but in which the model variables are iteratively updated as in the sequential version of the EnRML (Gu and Oliver 2007). In the following sections, we first review the EnRML formulation with emphases on its sequential and batch implementation. Because the iterative method requires estimation of the sensitivity, we investigate the ensemble estimate of sensitivity and compare it with the sensitivity from the adjoint method, and illustrate the effect of using ensemble sensitivity in the EnRML for high-dimensional history matching problems. Results from the EnKF, the ES, the seq-EnRML and the batch-EnRML are compared for two examples with varying degrees of nonlinearity which have been selected to illustrate the potential benefits of iteration, damping of updates, and sequential data assimilation for improving the skill of the predictions.

2 Ensemble Randomized Maximum Likelihood Method

An iterative form of the ensemble Kalman filter called the ensemble randomized maximum likelihood method (EnRML) was introduced in Gu and Oliver (2007). If sequential updating is not desirable, the EnRML can also be used with a slight modification to assimilate data from different times or to assimilate all data simultaneously.

2.1 Optimization Method to Sample Posterior pdf

In this section, we review the method that samples the posterior pdf by the mean of optimization without the ensemble approximation. Denote two independent sets

of data by $d^{\mathrm{I}}_{\mathrm{obs}}$ and $d^{\mathrm{II}}_{\mathrm{obs}}$ with no condition on observation time or from how many different times they are collected. Let $p(m \mid d^{\mathrm{I}}_{\mathrm{obs}})$ be the probability density function (pdf) of model variables m conditioned to the first set of data $d^{\mathrm{I}}_{\mathrm{obs}}$. The pdf $p(m \mid d^{\mathrm{I}}_{\mathrm{obs}})$ is the prior pdf for the assimilation of a new set of data $d^{\mathrm{II}}_{\mathrm{obs}}$. The Bayes theorem gives the probability density of model variables m after the assimilation of data $d^{\mathrm{II}}_{\mathrm{obs}}$ as

$$p\big(m \mid d^{\mathrm{I}}_{\mathrm{obs}}, d^{\mathrm{II}}_{\mathrm{obs}}\big) \propto p\big(d^{\mathrm{II}}_{\mathrm{obs}} \mid m\big)\, p\big(m \mid d^{\mathrm{I}}_{\mathrm{obs}}\big). \tag{1}$$

Efficient methods of sampling from the posterior pdf can be used if the prior pdf $p(m \mid d^{\mathrm{I}}_{\mathrm{obs}})$ before the assimilation of data $d^{\mathrm{II}}_{\mathrm{obs}}$ can be approximated as Gaussian with mean, m_{I}, and covariance, C_{M}, and the error in measurement $d^{\mathrm{II}}_{\mathrm{obs}}$ are normally distributed with zero mean and covariance, C_{D}, so that (1) becomes

$$p\big(m \mid d^{\mathrm{I}}_{\mathrm{obs}}, d^{\mathrm{II}}_{\mathrm{obs}}\big) \propto \exp\big[-S(m)\big], \tag{2}$$

where

$$S(m) = \frac{1}{2}\Big[\big(g(m) - d^{\mathrm{II}}_{\mathrm{obs}}\big)^{\mathrm{T}} C_{\mathrm{D}}^{-1}\big(g(m) - d^{\mathrm{II}}_{\mathrm{obs}}\big) + (m - m_{\mathrm{I}})^{\mathrm{T}} C_{\mathrm{M}}^{-1}(m - m_{\mathrm{I}})\Big]. \tag{3}$$

In (3), $g(\cdot)$ indicates the nonlinear relationship between data $d^{\mathrm{II}}_{\mathrm{obs}}$ and the model variables m.

For Monte Carlo methods, we sample from (2) as a way of representing the pdf of model variables conditional to both sets of data. An efficient, but approximate, method of sampling is to compute the values of model variables that minimize the stochastic objective function

$$S_*(m) = \frac{1}{2}\Big[\big(g(m) - d^{\mathrm{II}}_*\big)^{\mathrm{T}} C_{\mathrm{D}}^{-1}\big(g(m) - d^{\mathrm{II}}_*\big) + (m - m_*)^{\mathrm{T}} C_{\mathrm{M}}^{-1}(m - m_*)\Big], \tag{4}$$

where m_* is a sample from the prior pdf $p(m \mid d^{\mathrm{I}}_{\mathrm{obs}})$ and d^{II}_* is a sample from $p(d \mid d^{\mathrm{II}}_{\mathrm{obs}})$. Samples from $p(m \mid d^{\mathrm{I}}_{\mathrm{obs}})$ are available from the ensemble. If errors in the observations are additive, then samples from $p(d \mid d^{\mathrm{II}}_{\mathrm{obs}})$ are obtained by adding perturbations ϵ_D from the assumed pdf for observation error to the actual observations

$$d^{\mathrm{II}}_* = d^{\mathrm{II}}_{\mathrm{obs}} + \epsilon_D. \tag{5}$$

Using the Gauss–Newton method to minimize the stochastic objective function $S_*(m)$ in (4), the $(\ell+1)$th iterative estimate of a sample of model parameters is

$$\begin{aligned} m^{\ell+1} = {} & \beta_\ell m_* + (1-\beta_\ell) m^\ell - \beta_\ell C_{\mathrm{M}} G_\ell^{\mathrm{T}}\big(C_{\mathrm{D}} + G_\ell C_{\mathrm{M}} G_\ell^{\mathrm{T}}\big)^{-1} \\ & \times \big[g\big(m^\ell\big) - d^{\mathrm{II}}_* - G_\ell\big(m^\ell - m_*\big)\big], \end{aligned} \tag{6}$$

where G_ℓ is a linearization of $g(\cdot)$ at m^ℓ and β is the step length parameter that can be determined by standard line search. If β equals one, the iteration takes a full step size, otherwise the correction at the iteration is damped.

2.2 Ensemble-Based Methods

For methods in which the adjoint gradient is used in the Gauss–Newton minimization (6), it is typical that $d_{\mathrm{obs}}^{\mathrm{II}}$ contains all the data and C_{M} is the covariance of the model parameters prior to assimilation of any data, so that stationarity of C_{M} or singular value decomposition of C_{M} can be used to reduce the computation for large scale problems. In ensemble-based methods, both the covariance $C_{\mathrm{M}}^{\mathrm{e}}$ and the sensitivity G^{e} are obtained from a limited size ensemble. The stochastic objective function (4) then becomes

$$S_*^{\mathrm{e}}(m) = \frac{1}{2}\big[\big(g(m) - d_*^{\mathrm{II}}\big)^{\mathrm{T}} C_{\mathrm{D}}^{-1}\big(g(m) - d_*^{\mathrm{II}}\big) + (m - m_*)^{\mathrm{T}} C_{\mathrm{M}}^{\mathrm{e}\,-1}(m - m_*)\big]. \tag{7}$$

In some implementations of the EnKF, the covariance of noise in data C_{D} is also approximated by the ensemble. It is often, however, to assume uncorrelated data and use the theoretical variance of the noise in C_{D}. Benefits of the ensemble approximation are computational efficiency and ability to handle non-stationary covariance after data assimilation.

When the ensemble approximation $C_{\mathrm{M}}^{\mathrm{e}}$ and G^{e} are used and $\beta = 1$, the first iteration of the Gauss–Newton methods (6) becomes

$$m^1 = m_* - C_{\mathrm{M}}^{\mathrm{e}} G^{\mathrm{eT}}\big(C_{\mathrm{D}} + G^{\mathrm{e}} C_{\mathrm{M}}^{\mathrm{e}} G^{\mathrm{eT}}\big)^{-1}\big[g(m_*) - d_*^{\mathrm{II}}\big], \tag{8}$$

which is simply the perturbed-observation form of the EnKF update equations (Burgers et al. 1998; Houtekamer and Mitchell 1998) when applied to the updating of model parameters. If d^{II} includes all the data at various times, (8) represents the ES update. In both the EnKF and ES, the sensitivity G^{e} is never calculated explicitly, instead only the cross-covariance between model and state variables and data are used in the Kalman gain. The cross-covariance between model variables m and data d is

$$\begin{aligned} C_{md} &= E\big[(m - \bar{m})\big(g(m) - \bar{g}(m)\big)^{\mathrm{T}}\big] \\ &\approx E\big[(m - \bar{m})(m - \bar{m})^{\mathrm{T}} G^{\mathrm{eT}}\big] \\ &\approx C_{\mathrm{M}}^{\mathrm{e}} G^{\mathrm{eT}}, \end{aligned} \tag{9}$$

where the overbar indicates the mean, $E[\cdot]$ indicates the expectation. Similarly, in the EnRML the ensemble approximation of the covariance and sensitivity is used in (6) to iteratively minimize the objective function $S_*^{\mathrm{e}}(m)$ in (7):

$$\begin{aligned} m^{\ell+1} = {} & \beta_\ell m_* + (1 - \beta_\ell) m^\ell - \beta_\ell C_{\mathrm{M}}^{\mathrm{e}} G_\ell^{\mathrm{eT}}\big(C_{\mathrm{D}} + G_\ell^{\mathrm{e}} C_{\mathrm{M}}^{\mathrm{e}} G_\ell^{\mathrm{eT}}\big)^{-1} \\ & \times \big[g\big(m^\ell\big) - d_*^{\mathrm{II}} - G_\ell^{\mathrm{e}}\big(m^\ell - m_*\big)\big]. \end{aligned} \tag{10}$$

Different from the EnKF, however, G_ℓ^{e} and $C_{\mathrm{M}}^{\mathrm{e}}$ in (10) need to be calculated separately in the EnRML. $C_{\mathrm{M}}^{\mathrm{e}}$ is the covariance of model parameters conditioned to data $d_{\mathrm{obs}}^{\mathrm{I}}$ before the assimilation of the new set of data $d_{\mathrm{obs}}^{\mathrm{II}}$. This covariance, represented by the ensemble of models in the EnRML, provides weighting of prior information for assimilation of $d_{\mathrm{obs}}^{\mathrm{II}}$. The covariance $C_{\mathrm{M}}^{\mathrm{e}}$ does not change with iterations, while

the ensemble approximation of the sensitivity G_ℓ^{e} does. The Hessian preconditioning is also approximated from the ensemble, although the Hessian matrix is not shown explicitly in (10) due to matrix manipulation.

In Gu and Oliver (2007), line search is performed based on the average value of the objective function of the ensemble, so all ensemble members take the same step length β at each iteration. Zhao et al. (2008) and Wang et al. (2010) suggested that using separate line search for each ensemble member could improve the efficiency of the method. Although the same data are used multiple times during iterations, the EnRML obtains correct estimates of the mean and the posterior covariance for linear problems due to the prior term in (6) (Gu and Oliver 2007). Iterative methods that do not explicitly incorporate a prior term during iteration must either terminate the iteration after a small number of iterations (Wen and Chen 2007) or include a prior term in the stopping criteria of the iteration to prevent overfitting to the data and excessive reduction in ensemble variability (Lorentzen and Nævdal 2011). In the following two sections, we show two special forms of the EnRML depending on the content of the two sets of data $d_{\mathrm{obs}}^{\mathrm{I}}$ and $d_{\mathrm{obs}}^{\mathrm{II}}$: one form leads to an iterative EnKF and one leads to an iterative ensemble smoother. A pseudo-code is included in the Appendix as a reference for implementation.

2.3 Sequential EnRML (Iterative EnKF)

If $d_{\mathrm{obs}}^{\mathrm{I}}$ consists of data from time 1 to $k-1$, $d_{\mathrm{obs},1},\ldots,d_{\mathrm{obs},k-1}$ and $d_{\mathrm{obs}}^{\mathrm{II}}$ is the data collected at time k and the updating process repeats for all k, the EnRML performs as an iterative EnKF. If the prior pdf, before the assimilation of data at time k, is sufficiently close to Gaussian with mean, m_{k-1}, and covariance, $C_{\mathrm{M},k-1}$, such that the Gaussian approximation is adequate for weighting new data versus prior data, and the error in measurement $d_{\mathrm{obs},k}$ are normally distributed with zero mean and covariance, $C_{\mathrm{D},k}$, then each realization of the model parameters is updated using (10) as

$$\begin{aligned} m_j^{\ell+1} = {} & \beta_\ell m_{k-1,j} + (1-\beta_\ell) m_j^\ell - \beta_\ell C_{\mathrm{M},k-1}^{\mathrm{e}} G_{k,\ell}^{\mathrm{eT}} \left(C_{\mathrm{D},k} + G_{k,\ell}^{\mathrm{e}} C_{\mathrm{M},k-1}^{\mathrm{e}} G_{k,\ell}^{\mathrm{eT}}\right)^{-1} \\ & \times \left[g_k\left(m_j^\ell\right) - d_{\mathrm{obs},k,j} - G_{k,\ell}^{\mathrm{e}}\left(m_j^\ell - m_{k-1,j}\right)\right], \quad j = 1,2,\ldots,N_{\mathrm{e}}. \end{aligned} \tag{11}$$

In (11), the subscript j is the index for the ensemble members and N_{e} is the size of the ensemble, $g_k(\cdot)$ indicates the nonlinear relationship between data at time t_k and model variables, the sensitivity matrix $G_{k,\ell}^{\mathrm{e}}$ is a linearization of $g_k(\cdot)$ at the ℓth iteration. At each data assimilation time k, $C_{\mathrm{M},k-1}^{\mathrm{e}}$ is the prior covariance before the assimilation of data at time k, represented by the ensemble of model realizations $m_{k-1,j}$, $j = 1,2,\ldots,N_{\mathrm{e}}$.

2.4 Batch EnRML (Iterative ES)

The EnRML can also be used to incorporate all data collected at different times together to estimate the model variables, in which case $d_{\mathrm{obs}}^{\mathrm{I}}$ is an empty set and $d_{\mathrm{obs}}^{\mathrm{II}}$ consists of data collected at all data times $d_{\mathrm{obs},1},\ldots,d_{\mathrm{obs},N_{\mathrm{t}}}$, where N_{t} represents total number of data times. Assuming that the prior pdf before the assimilation of any

data is Gaussian with mean m_{pr}, and covariance C_{M}, and the error in measurement $d_{\mathrm{obs},1},\ldots,d_{\mathrm{obs},N_{\mathrm{t}}}$ are normally distributed with zero mean and covariance, C_{D}, then each realization of the model parameters is updated using (10) as

$$m_j^{\ell+1} = \beta_\ell m_{\mathrm{pr},j} + (1-\beta_\ell)m_j^\ell - \beta_\ell C_{\mathrm{M}}^{\mathrm{e}} G_\ell^{\mathrm{eT}}\left(C_{\mathrm{D}} + G_\ell^{\mathrm{e}} C_{\mathrm{M}}^{\mathrm{e}} G_\ell^{\mathrm{eT}}\right)^{-1} \times \left[g\left(m_j^\ell\right) - d_{\mathrm{obs},j} - G_\ell^{\mathrm{e}}\left(m_j^\ell - m_{\mathrm{pr},j}\right)\right], \quad j=1,2,\ldots,N_{\mathrm{e}}, \tag{12}$$

with

$$d_{\mathrm{obs}} = \begin{pmatrix} d_{\mathrm{obs},1} \\ \vdots \\ d_{\mathrm{obs},N_{\mathrm{t}}} \end{pmatrix}, \qquad g(\cdot) = \begin{pmatrix} g_1(\cdot) \\ \vdots \\ g_{N_{\mathrm{t}}}(\cdot) \end{pmatrix},$$

$$C_{\mathrm{D}} = \begin{pmatrix} C_{\mathrm{D},1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C_{\mathrm{D},N_{\mathrm{t}}} \end{pmatrix}, \qquad G^{\mathrm{e}} = \begin{pmatrix} G_1^{\mathrm{e}} \\ \vdots \\ G_{N_{\mathrm{t}}}^{\mathrm{e}} \end{pmatrix}.$$

The only difference between this batch implementation of the EnRML and the sequential implementation is that data from all times are used simultaneously to condition the model variables, so that the sensitivity matrix G^{e} is expanded to include sensitivity of data at different times. The batch implementation of the EnRML, in fact, strongly resembles the traditional iterative gradient-based history matching method (Li et al. 2003), but with the ensemble representation of the pdf and the ensemble approximation of the sensitivity. Similar to the sequential EnRML, $C_{\mathrm{M}}^{\mathrm{e}}$ and $m_{\mathrm{pr},j}$, $j=1,2,\ldots,N_{\mathrm{e}}$ in (12) represent the prior information before the assimilation of data and do not change with iteration.

2.5 Ensemble-Based Sensitivity

Note that in (11) and (12), there is no realization index for the sensitivity matrix G, because in the EnRML all realizations use the same approximation to the sensitivity. The ensemble average sensitivity G^{e} at the ℓth iteration, can be computed by solving

$$\Delta d^\ell = G_\ell^{\mathrm{e}} \Delta m^\ell \tag{13}$$

using singular value decomposition (SVD) (Gu and Oliver 2007). The columns of Δd^ℓ and Δm^ℓ are deviation realizations of the predicted data and deviation realizations of the model variables from the mean, respectively. Although the dimension of the model variables, N_m, is generally large (10^4–10^6), the size of the ensemble, N_{e}, is typically fairly small (10^2) so the effort required to compute the SVD is affordable.

Generally, we do not expect the sensitivity approximation from the ensemble to be identical to the accurate local gradient (e.g. analytical solution or adjoint sensitivity) unless the variability of the ensemble is very small and the size of the ensemble is large. The adjoint equation calculates the local derivative based on a numerical reservoir simulator, while the ensemble approximates an average relationship between the

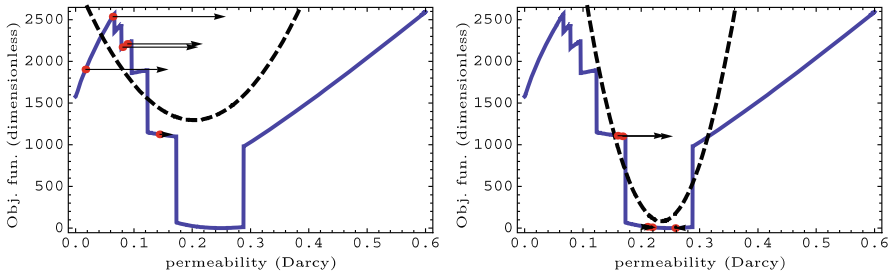


Fig. 1 Illustration of the first two iterations of the batch EnRML. The *solid blue curve* represents the change of objective function with the model variable (permeability of the bottom layer); the *five red dots* are the model ensemble prior to assimilation of data at the current iteration; the *black dashed curve* is the quadratic approximation of the objective function from the ensemble at the current iteration; the *black arrow* shows the direction of change for each realization and the length of the arrow represents the magnitude of change

change in predicted data and the change in model parameters through the correlation estimated from multiple simulation runs. We illustrate the use of the ensemble-based sensitivity in the EnRML using a simple three layer system (one-dimensional horizontal flow in each layer with no vertical communication between layers) with piston-like displacement of oil by water at constant inlet and outlet pressures with only one uncertain model parameter, the permeability of the bottom layer. Details of this example can be found in Oliver et al. (2011) where the permeability of all three layers are estimated using a gradient-based method. The data are observations of water cut and total flow rate at the outlet at seven different times. All 14 data from the seven times are assimilated simultaneously using the batch EnRML (12).

The solid blue line in both plots in Fig. 1 represents the change of the objective function (3) with respect to the model parameter. Because the water cut is measured at discrete intervals, this objective function has multiple local minima and is not differentiable everywhere. The local gradient, as would be obtained from the adjoint method, is the slope (if it exists) of the solid blue line. Depending on the starting point of the minimization, a gradient-based minimization method using the local gradient may converge to different minima. Figure 1(a) illustrates the first iteration of the batch EnRML using an ensemble of size five. First, the realizations are sampled from the prior pdf of model parameters and the values of the objective function are evaluated at each realization. The five realizations used in this one-parameter problem are shown as red dots in Fig. 1(a). From the values of the model parameter and their corresponding objective function, the ensemble approximates the objective function by a quadratic function (shown as the dashed black curve in Fig. 1(a)) based on the computed estimate of the Hessian and gradient of the objective function. The black arrows in Fig. 1(a) represent the direction of change for each realization and in this case all the realizations require an increase to the permeability of the bottom layer in order to reduce the objective function. The magnitude of change to each realization is represented by the length of the arrow in Fig. 1(a). Figure 1(b) shows the same procedure as in Fig. 1(a) for the second iteration with the starting model parameters

(five red dots in Fig. 1(b)) being the updated realizations after the first iteration shown in Fig. 1(a).

In high-dimensional problems, the ensemble approximation of the sensitivity G^{e} is often poor and corrupted by noise. The product of covariance and sensitivity $C_{\mathrm{M}}^{\mathrm{e}} G_{\ell}^{\mathrm{eT}}$, however, can be better approximated from the ensemble (will be shown later). Except for the last term in the updating equation of the EnRML, (10), the sensitivity G_{ℓ}^{e} appears together with $C_{\mathrm{M}}^{\mathrm{e}}$. The last term of (10), $G_{\ell}^{\mathrm{e}}(m^{\ell} - m_*)$, is generally small when the correction to the model parameters is small. As a result a good estimate of the product $C_{\mathrm{M}}^{\mathrm{e}} G_{\ell}^{\mathrm{eT}}$ is usually sufficient for minimizing the objective function using (10). At the first iteration of the EnRML, the covariance sensitivity product $C_{\mathrm{M}}^{\mathrm{e}} G_{\ell}^{\mathrm{eT}}$ is the same as the cross-covariance between the model parameters and the data as used in the EnKF (9). At later iterations, however, $C_{\mathrm{M}}^{\mathrm{e}} G_{\ell}^{\mathrm{eT}}$ cannot be replaced by the cross-covariance since G_{ℓ}^{e} changes with iteration and $C_{\mathrm{M}}^{\mathrm{e}}$ does not.

To show the similarity of the product $C_{\mathrm{M}}^{\mathrm{e}} G^{\mathrm{eT}}$ and $C_{\mathrm{M}} G^{\mathrm{T}}$, we compare both the sensitivity and the covariance sensitivity product from the adjoint method and from an ensemble of size 100 using a two-dimensional waterflood example of standard 5-spot pattern with 41×41 gridblocks. All the wells are under bottom-hole-pressure constraint. The ensemble mean of the log-permeability ($\ln k$) is shown in Fig. 2(a). The log-permeability is modeled as a Gaussian random field with two-dimensional spherical covariance function

$$C(r) = \begin{cases} \sigma^2 \left(1 - \frac{2r}{a\pi}\sqrt{1 - r^2/a^2} + \frac{2}{\pi} \arcsin \frac{r}{a}\right) & \text{for } 0 \le r \le a, \\ 0 & \text{for } r > a \end{cases} \tag{14}$$

with $\sigma = 1$ and $a = 30$ gridblocks. The coordinates are transformed to create anisotropy with the principal direction 30 degrees from the vertical direction and an anisotropy ratio of 3.5. The ensemble sensitivity (G^{e}) is computed using (13). The adjoint sensitivity is computed using the ECLIPSE 300 simulator for the ensemble mean $\ln k$ field. Figures 2(b) and 2(c) show the covariance of $\ln k$ at the injector cell to $\ln k$ at all other locations from the ensemble approximation and from the theoretical model (14), respectively. Note that for large scale models, it is computationally prohibitive to compute the full C_{M} and a lower rank representation from the ensemble is usually necessary. The theoretical covariance and its product with the sensitivity is shown here only for comparison. Figures 2(d) to (i) show the sensitivity and the covariance sensitivity product from both the adjoint and ensemble methods. Since it is the pattern of the sensitivity and covariance sensitivity product that determines the relative importance of each model variable (gridblock $\ln k$ in this case) to a particular data, we only use the dashed thick line in each subplot to show the zero contour. The warmer color indicates high magnitude and cooler color indicates low magnitude. The sensitivity of water production rate of P2 at the end of Year 8 to gridblock $\ln k$ from the adjoint method and from the ensemble are shown in Figs. 2(d) and (g), respectively. The adjoint-based sensitivity (Fig. 2(d)) shows a clear high sensitivity region corresponding to the high permeability path between P2 and the injector in the mean $\ln k$ field, while the ensemble-based sensitivity (Fig. 2(g)) appears to be quite noisy. Figures 2(e), (f), (h) and (i) show the corresponding covariance sensitivity product with both the ensemble estimation $C_{\mathrm{M}}^{\mathrm{e}}$ and the true covariance C_{M}.

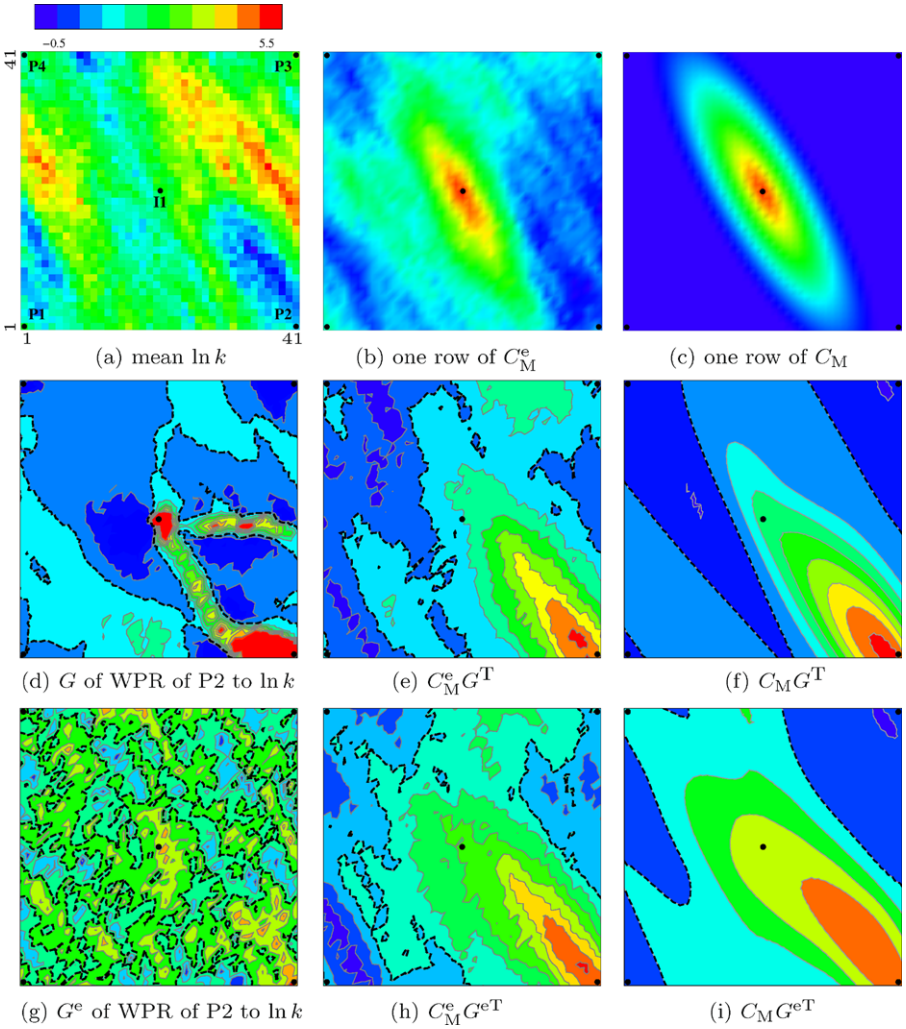


Fig. 2 Comparison between sensitivity and sensitivity covariance product from the ensemble-based and adjoint-based method. (**a**) ensemble mean of the $\ln k$, (**b**) covariance of $\ln k$ at the injector cell to $\ln k$ at all other locations from the ensemble, (**c**) covariance of $\ln k$ at the injector cell to $\ln k$ at all other locations from the theoretical model, (**d**) sensitivity of water rate of P2 at the end of Year 8 to $\ln k$ from the adjoint method, (**e**) product of the adjoint G with the ensemble estimate of covariance C_M^e, (**f**) product of the adjoint G with the theoretical covariance C_M, (**g**) sensitivity of water rate of P2 at the end of Year 8 to $\ln k$ from the ensemble, (**h**) product of the ensemble G^e with the ensemble estimate of covariance C_M^e, (**i**) product of the ensemble G^e with the theoretical covariance C_M. The *dashed thick line* in the *middle* and *bottom row* shows the zero contour line. The color scale is only for (**a**). The axis labels are shown in (**a**) for all plots

Despite the large difference in the sensitivity, the covariance sensitivity product obtained from the two methods (comparing Figs. 2(e), (f), (h) and (i)) shows similar patterns, with high correlation at areas that are expected to be influential to the water rate at P2.

Approximating sensitivity from an ensemble does not require the objective function to be differentiable as the adjoint-based method does. If the objective function has multiple small local minima but with a relatively well behaved large trend, using sensitivity approximated from an ensemble might be more efficient, because it has less chance of getting stuck at local minima (a good example of this is shown in Fig. 2 of Annan and Hargreaves (2004)). The quality of the ensemble-based sensitivity depends on the ensemble and the shape of the objective function, which is generally unknown for high-dimensional problems. Comparisons between the EnKF and estimation methods using the adjoint sensitivity (4DVAR and RML), however, suggest that the EnKF gives comparable performance as the 4DVAR in operational models for the weather system (Caya et al. 2005; Kalnay et al. 2007) and performs better than the RML for a problem of estimation of geologic facies boundary in history matching (Liu and Oliver 2005).

3 Illustrative Examples

Gu and Oliver (2007) showed that the EnRML obtains correct estimates of the mean and covariance for linear problems even when the updates are damped and many iterations are used. In this section, two nonlinear reservoir problems are used to compare results of the four methods introduced in the introduction: EnKF, ensemble smoother (ES), sequential EnRML (seq-EnRML) and batch EnRML (batch-EnRML). The first example is a five-spot waterflood problem, in which we compare the four methods using cases with different levels of nonlinearity. The experiment was repeated with multiple, independent ensembles for all four methods in order to reduce the effect of variability in the initial ensemble. The second example is the Brugge benchmark case with a relatively large number of model parameters (165,000), including horizontal and vertical permeability, porosity, net-to-gross ratio for each model gridblock and relative permeability curve, initial oil-water contact and layer mean of permeability. With large amounts of data available, appropriate selection of the model variables and localization of the Kalman gain, the EnKF has been shown previously to work well on this example (Peters et al. 2010). The focus of the Brugge example is the comparison of the sequential updating scheme of the EnKF to the batch assimilation methods, i.e. the non-iterative ensemble smoother and the batch EnRML. We also use this example to investigate the importance of appropriate tuning of the analysis step for the ES and the appropriate damping at early iterations for the iterative methods.

3.1 Five-Spot Waterflood

In this section, we consider a two-dimensional reservoir model (41×41 cells) with an injector at the center and four producers at the corners. The injector is constrained by bottom-hole pressure of 8000 psi and the producers are constrained by bottom-hole pressure of 500 psi. The initial reservoir pressure is 3600 psi. Porosity is 0.2 throughout the reservoir and log-permeability ($\ln k$) is a Gaussian random field with mean equal to 2.4 and standard deviation equal to 1.2, so the permeability field has a mode close to 12 mD and a range roughly from 5 mD to 150 mD. The reference $\ln k$

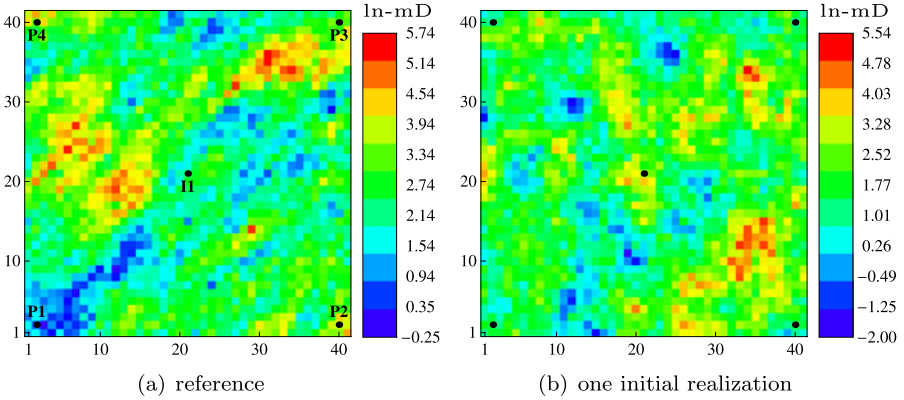
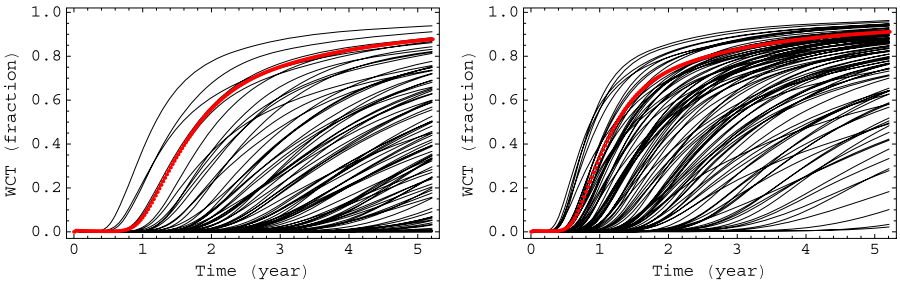


Fig. 3 Reference $\ln k$ and one realization from the biased initial ensemble

field is drawn from a multivariate normal distribution with an anisotropic exponential covariance function, with range equal to 18 gridblocks and 10 gridblocks, respectively, in the two principal directions. The reference $\ln k$ field and the well locations are shown in Fig. 3(a). The irreducible water saturation (S_{wir}) and the residual oil saturation (S_{or}) are both set as 0.2. Water cut of the producers at days 300, 600, 900, 1200, 1500 and 1900 are used as data for the assimilation, so the number of data at each data assimilation time is four and the total number of data is 24. The noise of the water cut data is 1%. We choose an ensemble size of 100, which is relatively large compared to the number of data for assimilation, so that localization is not necessary for the assimilation of data. Although the performance of the ensemble-based methods will still be affected by sampling error, the effect is expected to be similar for the four ensemble-based methods compared in this section. The focus of the comparison is on the benefit from the iterative schemes in the presence of strong nonlinearity.

Two scenarios were considered: the first is a scenario with biased initial ensemble, in which the mean of the initial ensemble is 1.5, so that the ensemble is biased compared to the reference case; the second is a scenario with unbiased initial ensemble, in which the mean of the initial ensemble is the same as the reference case. Other than the difference in the mean, the initial ensemble is identical in the two scenarios. An isotropic exponential covariance function with range 10 gridblocks is used to generate the initial realizations. One of the initial realizations from the biased initial ensemble is shown in Fig. 3(b). Figure 4 shows the prediction of water cut of Well P2 from the initial ensemble for both the scenario with biased and unbiased initial ensemble. The predictions from the biased initial ensemble show much later water breakthrough for all producers compared to the reference case and the unbiased scenario due to the lower mean of the initial ensemble. Bias due to incorrect specification of the prior is usually negligible in synthetic studies, but real field cases are generally more complex, and often bias cannot be completely avoided. The biased scenario is chosen in this example to represent a realistic situation where the initial assessment of uncertainty is poor and the amount of data is limited, so that the inverse problem is highly nonlinear.



(a) biased initial ensemble (b) unbiased initial ensemble

Fig. 4 The prediction of water cut of Well P2 from the biased and unbiased initial ensemble. The *red curve* is the reference, the *black curves* are ensemble forecasts

The four methods are compared based on results from 20 independent assimilation experiments with different initial ensembles of size 100 for both the biased and unbiased scenario. The performance of the four methods are compared in terms of the quality of data match after data assimilation and the computational cost. To differentiate the scenario with biased and unbiased initial ensemble, we use Γ (15) as a measure of nonlinearity of the inverse problem

$$\Gamma = \frac{1}{N_e N_d} \sum_{j=1}^{N_e} \big(g(m_{pr,j}) + G_{pr}^{e}(m_{post,j} - m_{pr,j}) - g(m_{post,j})\big)^{T} \times C_D^{-1} \big(g(m_{pr,j}) + G_{pr}^{e}(m_{post,j} - m_{pr,j}) - g(m_{post,j})\big), \tag{15}$$

where G_{pr}^{e} is the ensemble estimate of sensitivity of water cut data at the six data assimilation times to $\ln k$ based on the initial ensemble of $\ln k$ using (13). The samples from the prior and posterior pdf are represented by $m_{pr,j}$ and $m_{post,j}$, respectively. The nonlinearity measure Γ shows the difference between simulated data from samples of the posterior pdf and the linear projection of their magnitude using the simulated data from samples of the prior, the ensemble sensitivity from the prior and the correction required to change from the prior to posterior. This measure is appropriate for the ES and the batch EnRML in which data at all times are assimilated simultaneously. For the filtering problem, such as the EnKF, the nonlinearity changes with the sequential assimilation (this is discussed in Sect. 3.2.4), so a different measure should be used. The initial realizations are used to represent samples from the prior pdf, $m_{pr,j}$. The samples from the posterior pdf, $m_{post,j}$, are represented using the final updated realizations from the batch EnRML, because these updated realizations typically show reasonable updates and good match to data (as shown later in the section). The nonlinearity measure Γ for the scenario with biased and unbiased initial ensemble from the 20 runs are shown in Table 1. The nonlinearity for the scenario with biased initial ensemble is about twice as large as for the scenario with unbiased initial ensemble based on the measure in (15).

Table 1 Data mismatch of the initial ensemble and the updated ensemble and the computational cost for different methods in terms of the number of equivalent simulation runs for one history matched model. The nonlinear measure for the scenario with biased and unbiased initial ensemble is shown in the second row. The values in the table are given in terms of the mean plus and minus one standard deviation from the 20 independent runs

Nonlinearity (Γ)	Biased		Unbiased	
	294 ± 53		142 ± 27	
	Data mismatch	Number of runs	Data mismatch	Number of runs
initial	2581 ± 73	–	1373 ± 64	–
EnKF	96 ± 43	1 ± 0	51 ± 16	1 ± 0
ES	243 ± 64	1 ± 0	80 ± 10	1 ± 0
seq-EnRML	39 ± 11	2.4 ± 0.7	30 ± 8	2.7 ± 0.8
batch-EnRML	49 ± 16	7.9 ± 2.3	34 ± 9	6.5 ± 1.5

Data mismatch is computed using predictions from the updated ensembles,

$$S_{\mathrm{D}} = \frac{1}{N_{\mathrm{e}} N_{\mathrm{d}}} \sum_{j=1}^{N_{\mathrm{e}}} (d_{\mathrm{sim},j} - d_{\mathrm{obs}})^{\mathrm{T}} C_{\mathrm{D}}^{-1} (d_{\mathrm{sim},j} - d_{\mathrm{obs}}), \tag{16}$$

where d_{obs} is the vector of measured data and d_{sim} is the vector of simulated data from the model. Data mismatch for the four methods in the two scenarios are shown in Table 1. The computational cost is reported in terms of the number of equivalent simulation runs for each history matched model and the overhead for matrices manipulation at the analysis step is not considered, so that both the EnKF and ES require one simulation run for one history matched model. The overhead is negligible for this small example, however it can be considerable for large scale field studies as reported in Skjervheim and Evensen (2011). In Table 1, one simulation run is defined as running one simulation for the entire production length (from day 0 to day 1900). For the sequential EnRML, iterations at different data assimilation times require simulation runs for different lengths of time period. One simulation for iterations at the second data assimilation time (day 600) is considered as $600/1900 \approx 0.32$ equivalent simulation run. Iterations are triggered in the sequential EnRML in case of large corrections to the state variables. The indicator we used for this example is the maximum change to saturation (among all the gridblocks) at the analysis step of the EnKF. If the maximum change to saturation is greater than 0.3 but less than 0.5, iterations will start from the EnKF solution, meaning that the EnKF update is accepted as the first iteration of the sequential EnRML with a full step size ($\beta_1 = 1$ in (11)). If the maximum change to saturation is greater than 0.5, the EnKF solution is discarded and the iteration starts with $\beta = 0.3$ in (11) for the first iteration. The stopping criterion for the iteration for both sequential and batch EnRML is the reduction to the objective function at two consecutive iterations being less than 1%.

Both iterative methods, the sequential and batch EnRML, show lower data mismatch than the EnKF and ES. The reduction in data mismatch from iteration, however, is much more significant in the scenario with biased initial ensemble. Although the quality of data match is similar for the two iterative methods, the computational cost is much lower for the sequential EnRML since iterations are only triggered at the first one or two data assimilation times (for all the 20 runs) at which the cost of

iteration is relatively low. Note that the threshold for the change in saturation at the analysis step for triggering iteration is reduced to 0.2 for the sequential EnRML for the unbiased scenario, otherwise iteration was almost never triggered and the results from sequential EnRML are about the same as the EnKF.

3.2 Brugge Benchmark Study

In this section, history matching results for the first ten years of the Brugge field (Peters et al. 2010) are used to evaluate the performance of the EnKF, the ES and the batch EnRML. The results from the sequential EnRML are reported in Chen and Oliver (2010a) and are similar to the EnKF. The size of the Brugge model is $139 \times 48 \times 9$, with 10 injectors and 20 producers and with only oil and water phases present (for more description of the Brugge benchmark study see Peters et al. (2010)). The model parameters we chose to update include gridblock horizontal permeability (k_{h}) heterogeneity, vertical permeability (k_{v}) heterogeneity, porosity and net-to-gross ratio, three end points of Corey-type relative permeability curves, one initial water–oil contact and the mean horizontal and mean vertical permeability of each model layer. The total number of model parameters is approximately 165,000. The inclusion of the relative permeability parameters, depth of the initial contact and layer mean of the permeability allows correction to the model on a large scale, thus reduces the need for excessive changes to the gridblock properties. Typically a large parameterization provides the ability to make necessary updates to a variable at some location without inappropriately forcing a change at other locations. The use of large numbers of variables, even when not required for a match, can be beneficial in retaining plausible values of other parameters when an appropriate form of regularization is used. More details on the selection of model parameters can be found in Chen and Oliver (2010a, 2010b). The 104 realizations of the gridblock properties provided for the Brugge benchmark study are used for the three methods.

The primary constraints of the wells are fluid production rate (FPR) for the producers and water injection rate (WIR) for the injectors. All wells are subject to bottom-hole pressure (BHP) limits. Most of the producers are steered under FPR constraint of 2000 bbl/day and all the injectors are steered under WIR of 4000 bbl/day in the true case. Producer P9 was not able to make the target fluid production rate and the BHP of 725 psi was the active constraint throughout the ten years of production. The data used in assimilation include bottom-hole pressure, water cut, fluid production rate and water injection rate from 19 times. The number of data at each data assimilation time is 80. The primary constraints of the wells were also included as data with a low level of error to force the estimated models to honor the same active constraints as in the true case. The standard deviation of noise used for different types of data are 30 psi for BHP, 3.5% for water cut, 1 bbl/day for FPR and WIR if they are active constraint, 20 bbl/day for FPR of P9. Because the number of the data at each data assimilation time is relatively large and the initial ensemble of model parameters is not close to Gaussian, it was shown that localization is necessary to insure reasonable updates and data match (Chen and Oliver 2010a). The same distance-based localization as in Chen and Oliver (2010a) is used for all the three methods to constraint the updates of the gridblock properties. The updates to the global variables, i.e. parameters of the relative permeability curves, initial water–

Table 2 The average data mismatch (16) of the prediction from the updated ensemble for all the data and for different data types

	All data	BHP injectors	BHP producers	WCT	FPR	WIR
EnKF	1.1	0.3	0.9	0.8	1.0	0
ES	12.7	0.7	15.5	3.6	12.9	0
batch-EnRML	1.3	0.3	1.4	1.3	0.6	0

oil contact and the layer mean of permeability, are not localized, although the update of them could be regularized using adaptive localization methods (Anderson 2007; Zhang and Oliver 2010).

3.2.1 Comparison of Data Assimilation Results

Figure 5 shows a few selected production profiles predicted by the initial ensemble and the ensembles after updating by the three methods. For nonlinear problems, when the predictions from the initial ensemble are far from the observed data, iterations are necessary to obtain a satisfactory data match (compare ES results in the third row to batch-EnRML results in the bottom row of Fig. 5). The EnKF, however, gives a good data match without having to iterate as reported in early studies (summarized in Peters et al. (2010)). The prediction from the batch EnRML generally shows slightly higher variability than prediction from the EnKF. Table 2 shows the averaged data mismatch (16) for the EnKF, the ES and the batch EnRML for all the data and for different data types, i.e. bottom-hole-pressure measurements of the injectors and producers, water cuts, fluid production rates and water injection rates.

Figure 6 shows the estimated permeability heterogeneity of Layer 1 of one of the realizations using the EnKF and the corresponding initial realization on the same scale. The permeability in regions that are away from the wells remain mostly unchanged as a result of distance-based localization. Only small changes are made to the well coverage area and the updated realization maintains the channel structure as in the initial realization. The updated realizations from the ES and the batch EnRML are visually similar to that from the EnKF, but the realizations updated by the EnKF and the batch EnRML give much improved data match compared to those from the ES (as shown in Fig. 5 and Table 2).

Figure 7 shows the change of the data mismatch of the ensemble with iteration for the batch EnRML. The total number of iterations used in the batch EnRML is 26; iteration is terminated when the reduction of the average objective function becomes less than 0.5% . The black dots indicate the mean of the data mismatch of the ensemble, the dashed lines represent the 25% and 75% quantiles, the dotted dashed lines represent the minimum and maximum, and solid line represents the median. Iteration index 0 on the horizontal axis represents the initial ensemble. The variability of the data mismatch among the ensemble members is large at early iterations, and there exist realizations with extremely large data mismatch indicated by the mean being higher or close to the 75% quantile. With iterations all the realizations gradually converge to a low level of data mismatch.

Figures 8 and 9 show the evolution of the estimates of two of the global parameters, mean of the log-permeability of Layer 7 and initial water–oil contact, with the

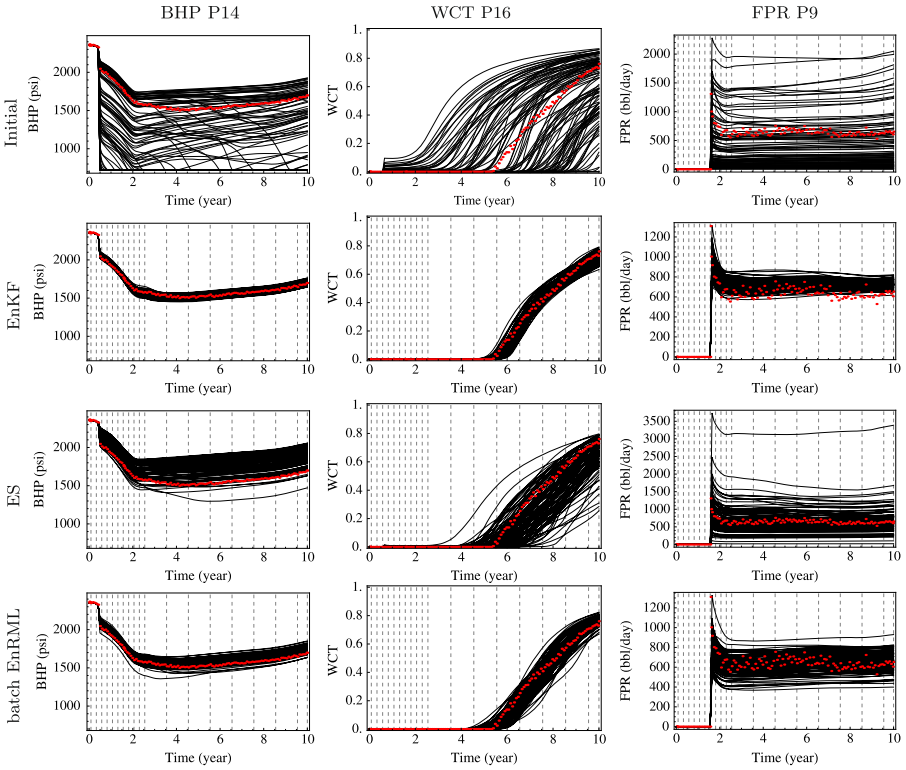


Fig. 5 Data match for a few selected wells for different methods. The *top row* shows prediction from the initial ensemble and is on different scale from the other rows. The predictions from the EnKF, ES and batch EnRML are made by rerunning simulations from initialization using the updated ensemble. The *red dots* are observations; the *black curves* are ensemble prediction and the *vertical dashed lines* indicate the times at which observations were assimilated

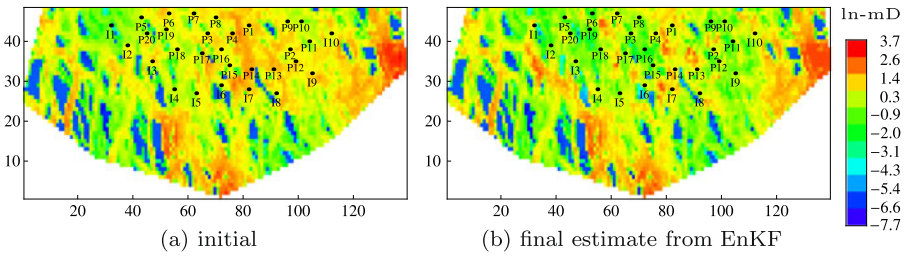


Fig. 6 The initial and the final estimate of the log-permeability heterogeneity of Layer 1 of one realization

data assimilation time for the EnKF and with iteration for the batch EnRML. In the EnKF, the estimate changes gradually with the assimilation of data at different times with also gradually reduced uncertainty. For the batch EnRML, the estimate and the associated uncertainty show gradual change in the first seven to 10 iterations then stabilize at about the same level in later iterations. The first 10 iterations correspond

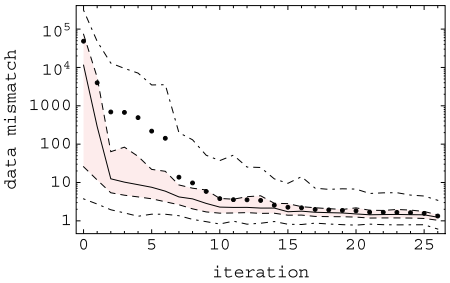
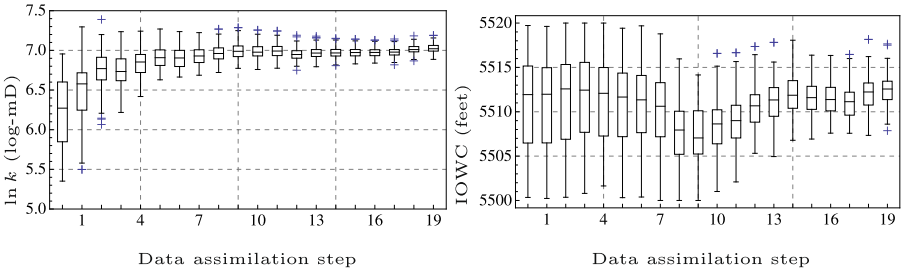
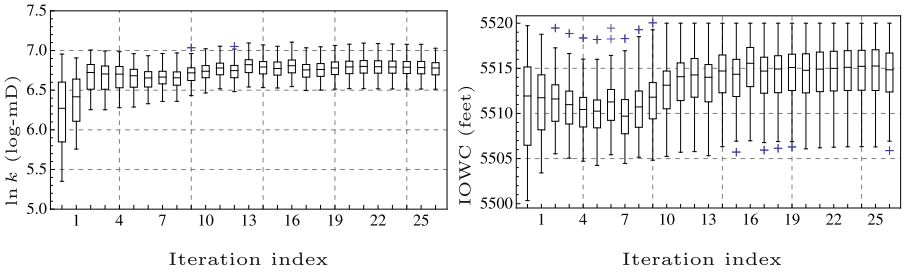


Fig. 7 Change of data mismatch from the updated realizations with iteration in the batch EnRML. The *black dots* indicate the mean of the data mismatch of the ensemble (16), the *dashed lines* represent the 25% and 75% quantiles, the *dotted dashed lines* represent the minimum and maximum, and *solid line* represents the median of the data mismatch of the ensemble. Iteration index 0 corresponds to the initial ensemble



(a) mean k_{h} of Layer 7 (b) initial oil-water contact

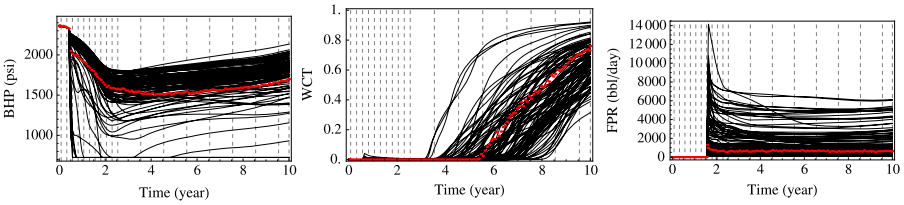
Fig. 8 Estimate of the global parameters with data assimilation time for the EnKF. The *bounds of the box* are 25% and 75% quantiles, the *whiskers* are the extremes, the *line in the box* is the median, and the *pluses* are outliers



(a) mean k_{h} of Layer 7 (b) initial oil-water contact

Fig. 9 Estimate of the global parameters with iteration for the batch EnRML. The *bounds of the box* are 25% and 75% quantiles, the *whiskers* are the extremes, the *line in the box* is the median, and the *pluses* are outliers

to the significant reduction of the objective function as shown in Fig. 7. The objective function of most of the realizations has reduced to less than 10 after the 10th iteration. Global parameters typically have high sensitivity to the overall behavior of



(a) BHP P14 (b) WCT P16 (c) FPR P9

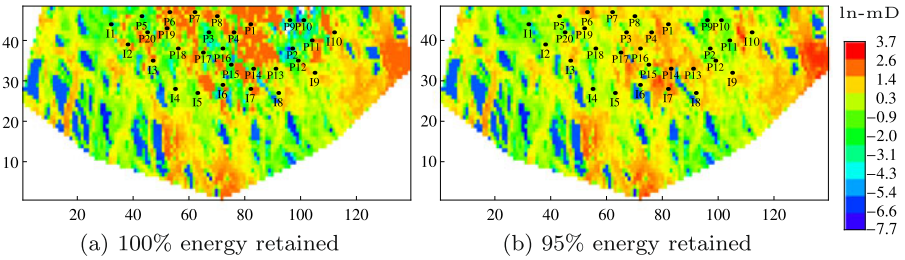
Fig. 10 Data match for a few selected wells for ES without truncation of singular values in the pseudo inversion. The *red dots* are observations; the *black curves* are prediction from the updated ensemble and the vertical dashed lines indicate the times at which observations were assimilated

the production, so that the change in global parameters could induce large change in the objective function. After the early iterations, the estimate of parameters with high sensitivity are stabilized and small changes to some of the gridblock properties at the later iteration further improve the data match for each of the wells. Note that the variability was maintained after the estimate of the global parameters has converged because of the prior term, $C_{\mathrm{M}}^{\mathrm{e}}$ and m_{pr} in (12).

3.2.2 Tuning Inversion for ES

It was shown in Evensen (2009, Chap. 14) that the truncation of singular values in the pseudo inversion of the matrix $C_{\mathrm{D}} + G^{\mathrm{e}} C_{\mathrm{M}}^{\mathrm{e}} G^{\mathrm{eT}}$ in (8) for the EnKF is important when the number of data is larger than the number of ensemble members or the matrix has poor conditioning. The number of data assimilated in the ES is generally much larger than the number of data at each data assimilation time in the EnKF, and is typically greater than the size of the ensemble. Even if the theoretical variance of the noise in data is used to construct C_{D}, the matrix $C_{\mathrm{D}} + G^{\mathrm{e}} C_{\mathrm{M}}^{\mathrm{e}} G^{\mathrm{eT}}$ is usually very ill-conditioned and the truncation in pseudo inversion becomes necessary to regularize the ES updates. Figure 10 shows the match to the same set of data as in Fig. 5 for the ES without truncation of singular values in the pseudo inversion. Compared to when 95% percent of energy is retained in the pseudo inversion for the ES (the 3rd row of Fig. 5), it is clear that retention of all singular values resulted in poor data match. Figure 11 compares the estimate of the log-permeability heterogeneity of Layer 1 of one of the realizations for the ES with 100% energy and 95% energy retained in the pseudo inversion. The corresponding initial realization is shown in Fig. 6(a). Without the tuning of inversion, the updates appear to be large and introduce roughness to the estimate.

The optimal truncation level, however, is case dependent. In this example we simply chose the best truncation level from a few trials. Table 3 shows the data match for the ES with different level of energy retained in the pseudo inversion. A large part of the data mismatch of the ES with 100% percent energy retained is a result of the violation of the fluid production rate constraint of the producers, meaning that not all the updated models from the ES are able to honor the production target (indicated by Fig. 10(a)). The reduction of data mismatch of other data types is also substantial from the tuning of inversion. Note that when truncation is used, scaling for different



(a) 100% energy retained (b) 95% energy retained

Fig. 11 The final estimate of the log-permeability heterogeneity of Layer 1 of one realization using the ES with different level of energy retained in the pseudo inversion

Table 3 The average data mismatch (16) of the prediction from the updated ensemble using the ES with different level of energy retained in the pseudo inversion. The data mismatch is shown for all the data and for different data types

Energy retained	All data	BHP injectors	BHP producers	WCT	FPR	WIR
80 percent	129.7	0.8	9.8	6.18	314.6	0
95 percent	12.7	0.7	15.5	3.6	12.9	0
99 percent	111.8	4.9	22	3.9	256.9	0
100 percent	2575.6	7.4	24.9	10.7	6535.2	0

types of data is important to avoid elimination of data with small magnitude, e.g. water cut (Wang et al. 2010).

3.2.3 Effect of Step Length at Early Iterations of EnRML

For the gradient-based history matching methods using the adjoint sensitivity, it was shown that it is important to damp updates at early iterations when the data mismatch is large (Li et al. 2003; Gao and Reynolds 2006). Without damping, the roughness introduced at early iterations will not always be removed at later iterations, and the minimization is more likely to be stuck at a local minimum. Damping large updates at early iterations is also important for the EnRML and other iterative ensemble-based assimilation methods. In this paper, we simply restrict the magnitude of the update without modifying the direction, so that the rate of convergence might not be optimal. Applying methods like Levenberg–Marquardt or truncation of singular values might be viable alternatives.

The effect of step length β in (12) to the data mismatch of the resulted ensemble at the first iteration is analyzed using the Brugge example. The data mismatch for water cut versus the step length β at the first iteration is shown in Fig. 12. We chose the mismatch of water cut data because there are fewer outliers in the distribution and because the same scale can be used for all choices of β. The mismatch of other data types show similar behavior. The leftmost box in Fig. 12 corresponds to $\beta = 0$, so it is the water cut mismatch of the initial ensemble. The bounds of the box are 25% and 75% quantiles, the whiskers are the extremes, the line in the box is the median, the red dot is the mean, and the pluses are outliers. The choice of $\beta = 1$ shows the highest mean data mismatch among all the choices in plot (excluding $\beta = 0$). Although a choice of β between 0.4 to 0.5 would give the lowest data mismatch after the first

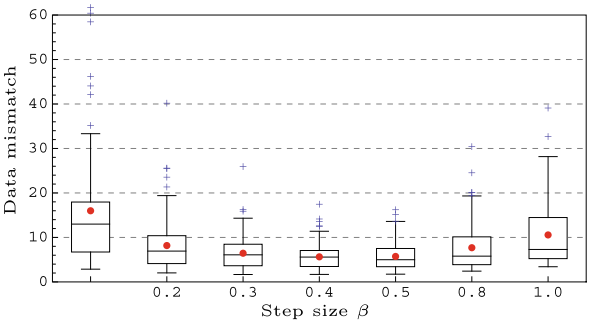


Fig. 12 Data mismatch of the water cut with respect to the step size β at the first iteration of the batch EnRML. The *bounds of the box* are 25% and 75% quantiles, the *whiskers* are the extremes, the *line in the box* is the median, the *red dot* is mean (16), and the *pluses* are outliers. The *first box* represents the initial ensemble

iteration, generally a more conservative choice of β at early iterations benefits the overall performance. In the results represented for the Brugge example, we chose $\beta = 0.3$ at the first iteration of the batch EnRML.

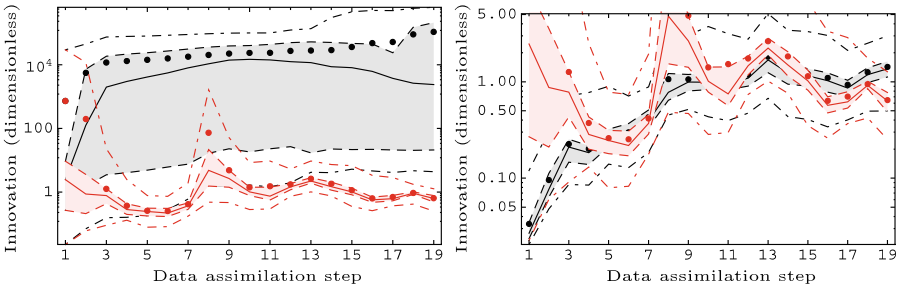
3.2.4 Benefit of Sequential Update of EnKF

For the Brugge example, the EnKF showed comparable results as the batch EnRML without the need to iterate. The good performance of the EnKF as a non-iterative method could be largely due to its sequential updating in which each sequential prediction from one data time to the next are from the models that have been conditioned to earlier data so that the nonlinearity at the subsequent updates are gradually reduced. We use the magnitude of innovation at the analysis step as an indicator of potential nonlinearity at each data assimilation time k

$$I_{k,j} = \frac{1}{N_\mathrm{d}}\big(g_k(m_j) - d_{\mathrm{obs},k}\big)^\mathrm{T} C_{\mathrm{D},k}^{-1}\big(g_k(m_j) - d_{\mathrm{obs},k}\big), \quad \text{for } j = 1, 2, \ldots, N_\mathrm{e}, \quad (17)$$

where the notation is the same as in (11). The innovation of the realizations reflects the distance between the model prediction and the data and distance for which a quadratic approximation needs to be made to determine the updates to the model parameters. Generally, small distance reduces the error in the quadratic approximation. The same measure is computed for both the EnKF and the batch EnRML for all data times, $k = 1, 2, \ldots, N_\mathrm{t}$. Although observations from all data assimilation times (N_t = 19 in this case) are incorporated simultaneously in the batch EnRML, we compute the innovation for data at different times separately for comparison with the EnKF.

Figure 13 compares the innovation $I_{k,j}$ for the batch EnRML at the first and last iteration with the innovation for the EnKF. The black color represents the batch EnRML at the first (Fig. 13(a)) and the last (Fig. 13(b)) iteration. The red color represents the EnKF and is the same in the two subplots, but on different scales. The dots indicate the mean, the dashed lines represent the 25% and 75% quantiles, the dotted dashed lines represent the minimum and maximum, and solid line represents the median. The innovation is the same for the first iteration of the batch EnRML and the EnKF at the first data time. The innovation at the subsequent data times for the EnKF are much lower than that of the batch EnRML at the first iteration due to the sequential assimilation of data. Note that the increase in the innovation around data time 7 for the EnKF is due to water breakthrough at various wells, which provides new information for data assimilation. The innovation at the last iteration of



(a) first iteration of batch EnRML (b) last iteration of batch EnRML

Fig. 13 Innovation (17) at each data assimilation time for the EnKF and batch EnRML. The *dots* represent the mean, the *dashed lines* represent the 25% and 75% quantiles, the *dotted dashed lines* represent the minimum and maximum, and *solid line* represents the median. The *black* color represents the batch EnRML at the first (**a**) and last (**b**) iteration. The *red* color represents the EnKF and is the same in (**a**) and (**b**), but on different scale

the batch EnRML at early data times (data time 1 to 10) is lower than the EnKF due to the updates at earlier iterations, but innovation at late data times (data time 11 to 19) is similar to the EnKF indicating the sequential assimilation of the EnKF gives similar effect as the iterative process of the batch EnRML (Fig. 13(b)).

4 Conclusions

In this paper, we propose the use of ensemble randomized maximum likelihood method as an iterative ensemble smoother (batch-EnRML) to improve the data match for problems in which the relationship between data and model parameters is highly nonlinear. The batch EnRML is compared with a sequential iterative method (seq-EnRML) and with two non-iterative methods, the ensemble Kalman filter (EnKF) and the ensemble smoother (ES) using two illustrative examples. The EnKF often works quite well at assimilating data into reservoir models, especially if the ensemble is sufficiently large or localization is used. A major reason for the good performance of the EnKF as a non-iterative methods appears to be a result of the sequential updating of variables. In the Brugge example, the sequential assimilation of data was observed to steadily reduce the nonlinearity of the problem. In the five-spot example with a biased prior, however, the nonlinearity was sufficiently large that the EnKF was unable to achieve an acceptably small data mismatch.

If simulator restarts complicate the data assimilation workflow or substantially increase the computational cost, the ES may be the appropriate choice, as it avoids the need for restarts and works well when the required updates are not large (Skjervheim and Evensen 2011). In our examples, however, we show that it is generally necessary to iterate in order to achieve satisfactory matches to data when the required changes to model variables are relatively large, while maintaining the flexibility of the smoother setup. Because of the benefits of the sequential adjustment of model variables, the sequential EnRML appears to be the preferred choice of iterative algorithms, unless the cost of restarts is substantial in which case the batch EnRML may be able to attain

small mismatch with a larger number of simulation runs. Localization is necessary to address the problem of assimilation of large amounts of data and spurious correlations for the ensemble-based methods, but does not directly address the issues of nonlinearity.

When large amounts of data are incorporated simultaneously, reduction in the magnitude of the corrections from the Gauss–Newton step in early iterations appears to be important. Appropriate damping increases the chance of convergence to an acceptable model and reduces the number of iterations that are required to converge. We also demonstrated that the use of the ensemble-based sensitivity might have advantages over the local accurate gradient for large scale complex inverse problems that are non-differentiable and have small local minima.

Appendix: Pseudo-code for EnRML

In this section, we detail the iterative algorithms for the sequential and batch EnRML methods. The notation is generally consistent with that used in Sect. 2:

Scalar	
N_e	number of realizations
N_m	number of model variables
N_p	number of dynamic variables
N_t	number of data assimilation times
k	index for data times and $k = 1, 2, \ldots, N_\mathrm{t}$
N_d	number of data
β	step length parameter for iterations
ℓ	iteration index
$c1$	max change at EnKF analysis to trigger iterations
$c2$	max change at EnKF analysis to decline EnKF update and start EnRML with a reduced step size at the 1st iteration
$c3$	min reduction of objective function for terminating iteration
$c4$	min update for terminating iteration
ℓ_max	max number of iterations for terminating iteration
Matrix	
m, Δm	$N_m \times N_\mathrm{e}$, model variables and deviation from mean
p, Δp	$N_p \times N_\mathrm{e}$, state variables and deviation from mean
d, Δd	$N_\mathrm{d} \times N_\mathrm{e}$, simulated data and deviation from mean, N_d is different for seq-EnRML and batch-EnRML
$\overline{m}$	$N_m \times N_\mathrm{e}$, each column is the mean of the model variables
$\overline{p}$	$N_p \times N_\mathrm{e}$, each column is the mean of the state variables
$\overline{d}$	$N_\mathrm{d} \times N_\mathrm{e}$, each column is the mean of the simulated data
$d_{\mathrm{obs},k,j}$	$N_\mathrm{d} \times 1$, jth realization of perturbed obs at data time k
$d^\mathrm{e}_{\mathrm{obs},k}$	$N_\mathrm{d} \times N_\mathrm{e}$, observation ensemble at data time k
Function	
$p = f_{t1\to t2}(m, p_{t1})$	function of model variables and state variables at time $t1$ to state variable between time $t1$ and $t2$
$d = g(m, p)$	function of model and state variables to predicted data

Functional	
$C(\delta p)$	maximum change of saturation (or other possible indicators)
$S_{\mathrm{D}}(m)$	average data mismatch of the ensemble

for $k \leftarrow 1$ **to** N_{t} **do**
  $d^{\mathrm{e}}_{\mathrm{obs},k} = [d_{\mathrm{obs},k,1}, d_{\mathrm{obs},k,2}, \ldots, d_{\mathrm{obs},k,N_{\mathrm{e}}}]$
end
if *(batch-EnRML)* **then** $d^{\mathrm{e}}_{\mathrm{obs}} = [d^{\mathrm{eT}}_{\mathrm{obs},1}, d^{\mathrm{eT}}_{\mathrm{obs},2}, \ldots, d^{\mathrm{eT}}_{\mathrm{obs},N_{\mathrm{t}}}]^{\mathrm{T}}$; $N = 1$;
if *(seq-EnRML)* **then** $N = N_{\mathrm{t}}$;
for $k \leftarrow 1$ **to** N **do**
  if *(seq-EnRML)* **then** $d^{\mathrm{e}}_{\mathrm{obs}} = d^{\mathrm{e}}_{\mathrm{obs},k}$; $t2 = k$;
  if *(batch-EnRML)* **then** $t2 = N_{\mathrm{t}}$;
  $p = f_{k-1\rightarrow t2}(m_{k-1}, p_{k-1})$; $d = g(m_{k-1}, p)$;
  $\Delta m_{k-1} = (m_{k-1} - \overline{m_{k-1}})/\sqrt{N_{\mathrm{e}} - 1}$;
  $\Delta p = (p - \overline{p})/\sqrt{N_{\mathrm{e}} - 1}$;
  $\Delta d = (d - \overline{d})/\sqrt{N_{\mathrm{e}} - 1}$;
  $\delta m = \Delta m_{k-1} \Delta d^{\mathrm{T}} (\Delta d \Delta d^{\mathrm{T}} + C_{\mathrm{D}})^{-1} (d^{\mathrm{e}}_{\mathrm{obs}} - d)$;
  $\delta p = \Delta p \Delta d^{\mathrm{T}} (\Delta d \Delta d^{\mathrm{T}} + C_{\mathrm{D}})^{-1} (d^{\mathrm{e}}_{\mathrm{obs}} - d)$;
  if $C(\delta p) > c1$ **then**
    if $C(\delta p) > c2$ **then**
      $\ell = 0$; $m^{\ell} = m_{k-1}$;
    else
      $\ell = 1$; $m^{\ell} = m_{k-1} + \delta m$;
    end
    while $\ell < \ell_{\max}$ **do**
      $p = f_{0\rightarrow t2}(m^{\ell}, p_0)$; $d^{\ell} = g(m^{\ell}, p)$;
      $G_{\ell} = (d^{\ell} - \overline{d^{\ell}})(m^{\ell} - \overline{m^{\ell}})^{+}$; $A_{\ell} = G_{\ell} \Delta m_{k-1}$;
      $\delta m = \beta m_{k-1} - \beta m^{\ell} - \beta \Delta m_{k-1} A_{\ell}^{\mathrm{T}} (C_{\mathrm{D}} + A_{\ell} A_{\ell}^{\mathrm{T}})^{-1} \ldots$
      $(d^{\ell} - d^{\mathrm{e}}_{\mathrm{obs}} - G_{\ell}(m^{\ell} - m_{k-1}))$;
      if $S_{\mathrm{D}}(m^{\ell} + \delta m) < S_{\mathrm{D}}(m^{\ell})$ **then**
        $m^{\ell+1} = m^{\ell} + \delta m$;
        if $(S_{\mathrm{D}}(m^{\ell}) - S_{\mathrm{D}}(m^{\ell+1}))/S_{\mathrm{D}}(m^{\ell}) < c3$ *or* $\|\delta m\| < c4$ **then**
          $m_k = m^{\ell+1}$; $p_k = p$; exit
        else
          $\ell = \ell + 1$;
        end
      else
        line search for β;
      end
    end
  else
    $m_k = m_{k-1} + \delta m$; $p_k = p + \delta p$
  end
end

References

Aanonsen SI, Nævdal G, Oliver DS, Reynolds AC, Vallès B (2009) Ensemble Kalman filter in reservoir engineering—a review. SPE J 14(3):393–412

Anderson JL (2007) Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. Physica D, Nonlinear Phenom 230(1–2):99–111

Annan JD, Hargreaves JC (2004) Efficient parameter estimation for a highly chaotic system. Tellus A 56(5):520–526

Burgers G, van Leeuwen PJ, Evensen G (1998) Analysis scheme in the ensemble Kalman filter. Mon Weather Rev 126(6):1719–1724

Caya A, Sun J, Snyder C (2005) A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. Mon Weather Rev 133(11):3081–3094

Chen Y, Oliver DS (2010a) Ensemble-based closed-loop optimization applied to Brugge Field. SPE Reserv Eval Eng 13(1):56–71

Chen Y, Oliver DS (2010b) Parameterization techniques to improve mass conservation and data assimilation for ensemble Kalman filter (SPE 133560). In: SPE Western regional meeting, Anaheim, California, USA, 27–29 May 2010

Evensen G (1994a) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res 99(C5):10143–10162

Evensen G (1994b) Advanced data assimilation for strongly nonlinear dynamics. Mon Weather Rev 125(6):1342–1354

Evensen G (2003) The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean Dyn 53:43–367

Evensen G (2009) Data assimilation: the ensemble Kalman filter, 2nd edn. Springer, Berlin

Gao G, Reynolds AC (2006) An improved implementation of the LBFGS algorithm for automatic history matching. SPE J 11(1):5–17

Gu Y, Oliver DS (2007) An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. SPE J 12(4):438–446

Houtekamer PL, Mitchell HL (1998) Data assimilation using an ensemble Kalman filter technique. Mon Weather Rev 126(3):796–811

Kalnay E, Li H, Miyoshi T, Yang SC, Ballabrera-Poy J (2007) 4-D-Var or ensemble Kalman filter? Tellus A 59(5):758–773

Li G, Reynolds AC (2009) Iterative ensemble Kalman filters for data assimilation. SPE J 14(3):496–505

Li R, Reynolds AC, Oliver DS (2003) History matching of three-phase flow production data. SPE J 8(4):328–340

Liu N, Oliver DS (2005) Critical evaluation of the ensemble Kalman filter on history matching of geologic facies. SPE Reserv Eval Eng 8(6):470–477

Lorentzen RJ, Nævdal G (2011) An iterative ensemble Kalman filter. IEEE Trans Autom Control 56(8):1990–1995

Oliver DS, Chen Y (2011) Recent progress on reservoir history matching: a review. Comput Geosci 15(1):185–221

Oliver DS, Zhang Y, Phale HA, Chen Y (2011) Distributed parameter and state estimation in petroleum reservoirs. Comput Fluids 46(1):70–77

Peters L, Arts R, Brouwer G, Geel C, Cullick S, Lorentzen R, Chen Y, Dunlop K, Vossepoel F, Xu R, Sarma P, Alhutali A, Reynolds A (2010) Results of the Brugge benchmark study for flooding optimization and history matching. SPE Reserv Eval Eng 13(3):391–405

Sakov P, Evensen G, Bertino L (2010) Asynchronous data assimilation with the ensemble Kalman filter. Tellus A 62(1):24–29

Sakov P, Oliver DS, Bertino L (2011) An iterative EnKF for strongly nonlinear systems. Mon Weather Rev (submitted)

Skjervheim JA, Evensen G (2011) An ensemble smoother for assisted history matching (SPE–141929). In: SPE reservoir simulation symposium, The Woodlands, Texas, 21–23 February

van Leeuwen PJ, Evensen G (1996) Data assimilation and inverse methods in terms of a probabilistic formulation. Mon Weather Rev 124(12):2898–2913

Wang Y, Li G, Reynolds AC (2010) Estimation of depths of fluid contacts and relative permeability curves by history matching using iterative ensemble Kalman smoothers (SPE 119056). SPE J 15(2):509–525

Wen XH, Chen WH (2007) Some practical issues on real-time reservoir model updating using ensemble Kalman filter. SPE J 12(2):156–166

Zhang Y, Oliver DS (2010) Improving the ensemble estimate of the Kalman gain by bootstrap sampling. Math Geosci 42(3):327–345

Zhao Y, Reynolds AC, Li G (2008) Generating facies maps by assimilating production data and seismic data with the ensemble Kalman filter, SPE-113990. In: Proceedings of the 2008 SPE improved oil recovery symposium, Tulsa, Oklahoma, 21–23 April