Comput Geosci (2012) 16:639–659
DOI 10.1007/s10596-012-9275-5

ORIGINAL PAPER

History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations

Alexandre A. Emerick · Albert C. Reynolds

Received: 25 July 2011 / Accepted: 6 January 2012 / Published online: 26 January 2012

Abstract The ensemble Kalman filter (EnKF) has become a popular method for history matching production and seismic data in petroleum reservoir models. However, it is known that EnKF may fail to give acceptable data matches especially for highly nonlinear problems. In this paper, we introduce a procedure to improve EnKF data matches based on assimilating the same data multiple times with the covariance matrix of the measurement errors multiplied by the number of data assimilations. We prove the equivalence between single and multiple data assimilations for the linear-Gaussian case and present computational evidence that multiple data assimilations can improve EnKF estimates for the nonlinear case. The proposed procedure was tested by assimilating time-lapse seismic data in two synthetic reservoir problems, and the results show significant improvements compared to the standard EnKF. In addition, we review the inversion schemes used in the EnKF analysis and present a rescaling procedure to avoid loss of information during the truncation of small singular values.

Keywords Ensemble Kalman filter · Multiple data assimilations · Time-lapse seismic

A. A. Emerick (✉) · A. C. Reynolds
University of Tulsa, 800 South Tucker Drive,
Tulsa, OK 74104, USA
e-mail: aemerick@gmail.com

A. C. Reynolds
e-mail: reynolds@utulsa.edu

1 Introduction

Reservoir simulation plays an important role in the entire hydrocarbon recovery process. In a reservoir simulation model, rock and fluid properties are characterized, and the physical process of fluid flow in the porous media is modeled in order to predict the reservoir's future performance. However, the reservoir data available are typically inaccurate, inconsistent, and insufficient, so that reservoir models are built with uncertain parameters, which means that predictions based on these models are also uncertain. In order to improve the reliability of reservoir predictions, the dynamic information available from historical field production data and seismic acquisitions needs to be incorporated into these reservoir models. This process is known in the oil industry as history matching. Bayesian statistics provides an adequate framework to incorporate field observations in reservoir simulation models in a way that allows one to describe uncertainty in the reservoir parameters and simulations predictions.

The EnKF, which can be derived from Bayesian statistics [13, Chap. 9], represents an attractive method for reservoir history matching because it is easy to implement and computationally efficient. Aanonsen et al. [1] present a comprehensive review of EnKF applications to reservoir problems. Some recent field applications of EnKF can be found in [6, 9, 15, 39]. Even though the EnKF was originally proposed as an alternative to the extended Kalman filter [13, Chap. 4] for applications in nonlinear dynamical systems, the update step in the EnKF is still linear. This linear update may result in a sub-optimal performance for highly nonlinear problems. Although the measurement errors do not directly impact the nonlinear relation

between predicted data and the EnKF state vector, it is well-known that they influence the level of correction in the states. For example, with EnKF, it is typically more difficult to assimilate very reliable data, i.e., with small measurement errors, than data with larger measurement errors. As shown in this paper, for the linear case with Gaussian prior, assimilating data multiple times with the covariance matrix of the measurement errors multiplied by the number of data assimilations is equivalent to assimilating data only once with the original covariance matrix. For the nonlinear case, however, this equivalence does not hold. In fact, the results presented in this paper indicate that it may be beneficial to assimilate data multiple times with an inflated measurement error covariance matrix in the nonlinear case. The rationality is that better estimates of the states can be obtained by doing successive smaller updates instead of one single large update.

This work was motivated by the comments made on pages 44, 45, 86, and 109 of Rommelse [37]. Rommelse suggested that when assimilation of accurate data requires a "large jump" between the forecast and analyzed states, the magnitude of the jump is overestimated by the linear update used in EnKF. He suggested that this overestimation of the magnitude of the jump could be reduced by assimilating the same data multiple times with increased measurement errors. Rommelse derived equations for determining the number of times that the data should be assimilated with increased measurement errors in order to obtain the same posterior variance for the state. His derivation, however, was based on a one-dimensional forecast vector, with a single datum and sensitivity matrix equal to the 1×1 identity matrix. Rommelse applied this idea in a multi-dimensional example, but he did not compare multiple assimilations of data with standard EnKF. He also did not discuss important implementation details. In a sense, this paper extends the theoretical equations on pages 44 and 45 of Rommelse [37] to multi-dimensional problems with an arbitrary sensitivity matrix and a general measurement error covariance matrix. However, our derivations assume that we multiply the covariance matrix of the measurement errors by the number of data assimilations, whereas it appears that Rommelse assimilated data one by one using an ensemble square root filter [49]. In this case, each covariance matrix of the measurement errors at each data assimilation is 1×1. Thus, he could assimilate each individual datum a different number of times.

Reservoir production data are typically scarce spatially, but dense in time. As shown in Reynolds et al. [36], the application of EnKF is similar to one Gauss–Newton (GN) iteration using an average sensitivity matrix and a full step in the search direction. Based on this analogy, we conjecture that sequential assimilation is one of the reasons that EnKF gives acceptable results when assimilating production data that are fairly closely spaced in time. The process works such that at each data assimilation time step, one GN correction is done to each realization in the ensemble of states, but because we have production data for several consecutive data assimilation time steps, we accumulate several GN corrections, keeping the ensemble conditioned to the production history. In contrast, seismic data are characterized by the large amount of data, spatially dense, but available at only a few times. Thus, we no longer have the beneficial effect of several consecutive data assimilation time steps as observed in the production data case. While our conjecture that assimilating production data with EnKF provides an approximation of multiple GN iterations is based on intuition, it is well-known that taking a full step at early iterations in the GN method when matching production data can lead to significant overcorrection (overshooting and undershooting) of reservoir parameters and result in unrealistically rough rock property fields [19, 29]. This overcorrection can be avoided either by applying some form of damping or constraining changes in model parameter at early iterations or by inflating the covariance matrix associated with measurement errors. This fact provides a strong motivation for the procedure considered in this paper, in which we apply EnKF to assimilate the same data multiple times using an inflated covariance matrix.

This paper is organized as follows: In the next sections, we briefly present the Kalman filter and EnKF equations. After that, we show the equivalence between single and multiple data assimilations (MDA) for the linear-Gaussian case followed by an interpretation of MDA for the nonlinear case. Then, we describe the implementation of the proposed procedure. After that, we present a brief discussion about time-lapse seismic data in the reservoir history-matching context. In the subsequent section, we present two test cases, which correspond to a small synthetic reservoir problem and the Brugge benchmark case [34]. The last section of the paper presents the conclusions. The paper also includes two appendices. In Appendix A, we review the inversion process required in the EnKF analysis. In this appendix, we discuss the inversion scheme based on truncated singular value decomposition (SVD) and the subspace inversion proposed by Evensen [12]. In both cases, we present a rescaling procedure to avoid loss of information during truncation of small singular values. In Appendix B, we present some matrix manipulations omitted in the main body of the paper.

2 Kalman filter

Under the restrictions of a Gaussian prior, a linear relation between state and predicted data, Gaussian noise in the measurements and Gaussian model error, the KF [26] is the optimal sequential data assimilation scheme. In the KF, the mean, μ_y^n, and the covariance, C_{Y}^n, of a state vector, y^n, are updated sequentially in time using

$$\mu_y^{n,a} = \mu_y^{n,f} + K_n \left(d_{\mathrm{obs}}^n - H_n \mu_y^{n,f} \right) \tag{1}$$

and

$$C_{\mathrm{Y}}^{n,a} = \left(I_{N_y} - K_n H_n \right) C_{\mathrm{Y}}^{n,f}, \tag{2}$$

where

$$K_n \equiv C_{\mathrm{Y}}^{n,f} H_n^{\mathrm{T}} \left(C_{\mathrm{D}}^n + H_n C_{\mathrm{Y}}^{n,f} H_n^{\mathrm{T}} \right)^{-1}. \tag{3}$$

Equations 1–3 are known as the KF analysis equations. In these equations, n denotes the data assimilation time-step index. K_n is the Kalman gain matrix. H_n is the $N_n \times N_y$ sensitivity matrix, which defines the linear relation between the state vector and predicted data, i.e.,

$$d^{n,f} = H_n y^{n,f}. \tag{4}$$

C_{D}^n is the covariance matrix of the measurement errors. I_{N_y} is the $N_y \times N_y$ identity matrix, where N_y denotes the dimension of the state vector y^n. The superscripts a and f denote analysis and forecast, respectively.

3 Ensemble Kalman filter

EnKF [3, 11, 24] is a sequential data assimilation method in which an ensemble of realizations is employed to construct Monte Carlo approximations of the mean and the covariance of the state vector. A convenient way to present the EnKF equations is by defining an augmented state vector, y^n, in which the predicted data vector, d^n, is also included, i.e.,

$$y_j^n = \begin{bmatrix} m_j^n \\ p_j^n \\ d_j^n \end{bmatrix}, \tag{5}$$

where the subscript j denotes the jth ensemble member. Equation 5 pertains to the parameter-state estimation problem [14]. Here, m^n is the N_m-dimensional column vector of model parameters, and p^n is the N_p-dimensional column vector representing the state of the dynamical system (primary variables of the reservoir simulator). Then, we define the matrix H_n as

$$H_n \equiv \left[O \; I_{N_n} \right], \tag{6}$$

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. Using the definition 6, we can write the predicted data vector, $d_j^{n,f}$, using the same linear relation used in the KF (Eq. 4).

This "trick" of augmenting the state vector with the predicted data allows to derive the EnKF equations in a similar fashion of the KF equations. The trick turns a nonlinear relation between data and the original state vector into a linear relationship between the vector of predicted data and the augmented state vector, making it possible to write down the formula for the analysis step analytically [28]. However, this trick does not remove the effect of the nonlinearity, and it only disguises the problem. As shown in Li and Reynolds [28], augmenting the state vector with data results in a correct procedure for sampling the posterior probability density function (pdf) if, and only if, at every data assimilation time step, the predicted data vector is a linear function of the unaugmented state vector. Using the augmented state vector, the EnKF analysis equations can be written in a similar form as the KF equations. However, unlike the KF, where we update directly the mean and covariance matrix, the EnKF updates each ensemble member using

$$y_j^{n,a} = y_j^{n,f} + C_{\mathrm{Y}}^{n,f} H_n^{\mathrm{T}} \left(C_{\mathrm{D}}^n + H_n C_{\mathrm{Y}}^{n,f} H_n^{\mathrm{T}} \right)^{-1} \times \left(d_{\mathrm{uc},j}^n - d_j^{n,f} \right), \tag{7}$$

for $j = 1, \ldots, N_e$, where N_e is the number of state vectors in the ensemble, i.e., the ensemble size and $d_{\mathrm{uc},j}^n$ is a sample from the Gaussian distribution $\mathcal{N}(d_{\mathrm{obs}}^n, C_{\mathrm{D}}^n)$. In order to simplify the notation, we drop the time-step index n in the remaining equations of this paper, noting that all equations refer to the same data assimilation time step. In Section 5, we make an exception to this notational simplification as the time index is necessary to explain the MDA procedure.

4 Multiple data assimilations

4.1 Multiple data assimilations for the linear case

In this section, we show the equivalence between single and multiple data assimilations for the linear case using the KF, where the full-rank covariance C_{Y}^f is updated.

Proposition *For the linear case, with a Gaussian prior and Gaussian noise in the measurements (linear-Gaussian problem), applying the KF to assimilate data N_a times with the measurement error covariance matrix multiplied by N_a is equivalent to assimilating the same data only once with the original measurement error covariance matrix.*

Proof We start the proof by noting that for the linear-Gaussian problem the posterior pdf of model parameters conditional to the observations is also Gaussian [33, Chap. 7]. Hence, to prove the proposition, it is necessary only to show that both procedures lead to the same posterior covariance matrix and posterior mean. Using results from [33, Chap. 7], the KF equations (Eqs. 1–3) can be written as

$$\begin{aligned}\mu_y^a &= \mu_y^f + C_{\mathrm{Y}}^f H^{\mathrm{T}} \left(C_{\mathrm{D}} + H C_{\mathrm{Y}}^f H^{\mathrm{T}}\right)^{-1} \left(d_{\mathrm{obs}} - H\mu_y^f\right) \\ &= C_{\mathrm{Y}}^a \left(\left(C_{\mathrm{Y}}^f\right)^{-1} \mu_y^f + H^{\mathrm{T}} C_{\mathrm{D}}^{-1} d_{\mathrm{obs}}\right) \end{aligned} \tag{8}$$

and

$$\begin{aligned}C_{\mathrm{Y}}^a &= C_{\mathrm{Y}}^f - C_{\mathrm{Y}}^f H^{\mathrm{T}} \left(C_{\mathrm{D}} + H C_{\mathrm{Y}}^f H^{\mathrm{T}}\right)^{-1} H C_{\mathrm{Y}}^f \\ &= \left(\left(C_{\mathrm{Y}}^f\right)^{-1} + H^{\mathrm{T}} C_{\mathrm{D}}^{-1} H\right)^{-1}. \end{aligned} \tag{9}$$

□

To assimilate data N_a times with the measurement error covariance multiplied by N_a, let us define

$$\widetilde{d}_{\mathrm{obs}} \equiv \begin{bmatrix} d_{\mathrm{obs}} \\ \vdots \\ d_{\mathrm{obs}} \end{bmatrix}, \tag{10}$$

$$\widetilde{H} \equiv \begin{bmatrix} H \\ \vdots \\ H \end{bmatrix} \tag{11}$$

and

$$\widetilde{C}_{\mathrm{D}} \equiv \begin{bmatrix} N_a C_{\mathrm{D}} & 0 & \cdots & 0 \\ 0 & N_a C_{\mathrm{D}} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & N_a C_{\mathrm{D}} \end{bmatrix}. \tag{12}$$

In the above definitions, we simply repeated the vector d_{obs} and the matrices H and C_{D} N_a times. For assimilating data multiple times, the linear relation (Eq. 4) becomes

$$\widetilde{d}^f = \widetilde{H} y^f. \tag{13}$$

Now, we develop an expression for $\widetilde{C}_{\mathrm{Y}}^a$, which denotes the posterior covariance for the MDA case. From Eq. 9, we can write $\widetilde{C}_{\mathrm{Y}}^a$ as

$$\begin{aligned}\widetilde{C}_{\mathrm{Y}}^a &= \left(\left(C_{\mathrm{Y}}^f\right)^{-1} + \widetilde{H}^{\mathrm{T}} \widetilde{C}_{\mathrm{D}}^{-1} \widetilde{H}\right)^{-1} \\ &= C_{\mathrm{Y}}^f - C_{\mathrm{Y}}^f \widetilde{H}^{\mathrm{T}} \left(\widetilde{C}_{\mathrm{D}} + \widetilde{H} C_{\mathrm{Y}}^f \widetilde{H}^{\mathrm{T}}\right)^{-1} \widetilde{H} C_{\mathrm{Y}}^f. \end{aligned} \tag{14}$$

Using definitions 10, 11, and 12 in the second equality of Eq. 14, we can write this posterior covariance matrix as

$$\widetilde{C}_{\mathrm{Y}}^a = C_{\mathrm{Y}}^f - \left[C_{\mathrm{Y}}^f H^{\mathrm{T}} \cdots C_{\mathrm{Y}}^f H^{\mathrm{T}}\right] C^{-1} \begin{bmatrix} H C_{\mathrm{Y}}^f \\ \vdots \\ H C_{\mathrm{Y}}^f \end{bmatrix}. \tag{15}$$

In Eq. 15, the matrix C is defined by

$$\begin{aligned} C &\equiv \begin{bmatrix} N_a C_{\mathrm{D}} & 0 & \cdots & 0 \\ 0 & N_a C_{\mathrm{D}} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & N_a C_{\mathrm{D}} \end{bmatrix} \\ &\quad + \begin{bmatrix} H C_{\mathrm{Y}}^f H^{\mathrm{T}} & H C_{\mathrm{Y}}^f H^{\mathrm{T}} & \cdots & H C_{\mathrm{Y}}^f H^{\mathrm{T}} \\ H C_{\mathrm{Y}}^f H^{\mathrm{T}} & H C_{\mathrm{Y}}^f H^{\mathrm{T}} & \cdots & H C_{\mathrm{Y}}^f H^{\mathrm{T}} \\ \vdots & & \ddots & \vdots \\ H C_{\mathrm{Y}}^f H^{\mathrm{T}} & \cdots & & H C_{\mathrm{Y}}^f H^{\mathrm{T}} \end{bmatrix} \\ &= \begin{bmatrix} N_a C_{\mathrm{D}} + C_{\mathrm{DD}}^f & C_{\mathrm{DD}}^f & \cdots & C_{\mathrm{DD}}^f \\ C_{\mathrm{DD}}^f & N_a C_{\mathrm{D}} + C_{\mathrm{DD}}^f & \cdots & C_{\mathrm{DD}}^f \\ \vdots & & \ddots & \vdots \\ C_{\mathrm{DD}}^f & \cdots & & N_a C_{\mathrm{D}} + C_{\mathrm{DD}}^f. \end{bmatrix}. \end{aligned} \tag{16}$$

In Eq. 16, we defined $C_{\mathrm{DD}}^f \equiv H C_{\mathrm{Y}}^f H^{\mathrm{T}}$ to simplify notation.

We need to develop an expression for C^{-1} in Eq. 15. As C is a real symmetric positive definite matrix, its

inverse is also real symmetric positive definite. Hence, we can write the product $CC^{-1} = I$ as

$$
\begin{aligned}
I &= CC^{-1} \\
&= \begin{bmatrix} N_a C_{\mathrm{D}} + C^f_{\mathrm{DD}} & C^f_{\mathrm{DD}} & \cdots & C^f_{\mathrm{DD}} \\ C^f_{\mathrm{DD}} & N_a C_{\mathrm{D}} + C^f_{\mathrm{DD}} & \cdots & C^f_{\mathrm{DD}} \\ \vdots & & \ddots & \vdots \\ C^f_{\mathrm{DD}} & \cdots & & N_a C_{\mathrm{D}} + C^f_{\mathrm{DD}} \end{bmatrix} \\
&\times \begin{bmatrix} A & B & \cdots & B \\ B & A & \cdots & B \\ \vdots & & \ddots & \vdots \\ B & \cdots & & A \end{bmatrix} = \begin{bmatrix} I_{N_n} & 0 & \cdots & 0 \\ 0 & I_{N_n} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & I_{N_n} \end{bmatrix}. \qquad (17)
\end{aligned}
$$

Thus, in order to compute C^{-1}, it is only necessary to develop expressions for the submatrices A and B. From Eq. 17, it is straightforward to obtain

$$\left(N_a C_{\mathrm{D}} + C^f_{\mathrm{DD}}\right) A + (N_a - 1)\, C^f_{\mathrm{DD}} B = I_{N_n} \qquad (18)$$

and

$$C^f_{\mathrm{DD}} A + \left(N_a C_{\mathrm{D}} + C^f_{\mathrm{DD}}\right) B + (N_a - 2)\, C^f_{\mathrm{DD}} B = 0. \qquad (19)$$

Subtracting Eq. 19 from Eq. 18, we obtain

$$A = \frac{1}{N_a} C_{\mathrm{D}}^{-1} + B \qquad (20)$$

and using this result in Eq. 19, we obtain

$$B = -\frac{1}{N_a^2} \left(C_{\mathrm{D}} + C^f_{\mathrm{DD}}\right)^{-1} C^f_{\mathrm{DD}} C_{\mathrm{D}}^{-1}. \qquad (21)$$

Using Eqs. 20 and 21 in Eq. 15, it is straightforward, but tedious, to show that

$$\widetilde{C}^a_{\mathrm{Y}} = C^f_{\mathrm{Y}} - C^f_{\mathrm{Y}} H^{\mathrm{T}} \left(C_{\mathrm{D}} + H C^f_{\mathrm{Y}} H^{\mathrm{T}}\right)^{-1} H C^f_{\mathrm{Y}} = C^a_{\mathrm{Y}}. \qquad (22)$$

In Appendix B, we present the matrix manipulations required to establish Eq. 22. This equation shows that the posterior covariance matrix obtained by assimilating data N_a times with the measurement error covariance matrix multiplied by N_a is the same as the posterior covariance matrix obtained assimilating data only once with the original measurement error covariance matrix. Following the same procedure, it is straightforward to show the equivalence of the posterior mean by starting with Eq. 8 and using Eqs. 20 and 21.

It is known that EnKF becomes equivalent to the KF for the linear-Gaussian case as the size of the ensemble goes to infinity [13, Chap. 4]. Thus, the equivalence between single and multiple data assimilations established in this section guarantees that as $N_e \to \infty$, EnKF with multiple data assimilations (EnKF-MDA) is still consistent with the KF for the linear-Gaussian case.

4.2 Sampling the posterior pdf with multiple data assimilations for the linear case

In the previous section, we showed the equivalence between single and multiple data assimilation for the linear-Gaussian case using the KF. In the KF, the mean and covariance are updated sequentially in time. For the EnKF, on the other hand, we sequentially update an ensemble of augmented state vectors generated by sampling the prior distribution in order to obtain a sampling of the posterior distribution. In this section, we demonstrate that for the linear-Gaussian case, EnKF-MDA samples the posterior pdf correctly, if we use the full-rank forecast covariance matrix C^f_{Y}. The derivation presented here explicitly shows that for EnKF-MDA, we need to perturb the observations based on the inflated covariance matrix of the measurement errors. The derivation presented in this section, follows Reynolds et al. [35] where it is shown that the randomized maximum likelihood (RML) method samples the posterior pdf correctly for the linear-Gaussian case.

For sampling the posterior pdf with EnKF-MDA, we start with a sample from the prior (forecast) pdf, denoted by y^f, i.e., $y^f \sim \mathcal{N}(\mu^f_y, C^f_{\mathrm{Y}})$. In addition, we perturb the observations by sampling $d_{\mathrm{uc}} \sim \mathcal{N}(d_{\mathrm{obs}}, N_a C_{\mathrm{D}})$. As before, to assimilate data N_a times, we define

$$\widetilde{d}_{\mathrm{uc}} \equiv \begin{bmatrix} d^1_{\mathrm{uc}} \\ \vdots \\ d^{N_a}_{\mathrm{uc}} \end{bmatrix}, \qquad (23)$$

where $d^\ell_{\mathrm{uc}} \sim \mathcal{N}(d_{\mathrm{obs}}, N_a C_{\mathrm{D}})$, for $\ell = 1, 2, \cdots N_a$.

The derivation has two parts. In the first part, we define the vector $\widetilde{y}^a$ as the minimizer of the RML objective function [35] modified for the MDA case, i.e.,

$$\widetilde{y}^a = \arg\min_y \widetilde{O}(y), \qquad (24)$$

where

$$\widetilde{O}(y) = \frac{1}{2} \left(y - y^f\right)^{\mathrm{T}} \left(C^f_{\mathrm{Y}}\right)^{-1} \left(y - y^f\right) + \frac{1}{2} \left(\widetilde{H} y - \widetilde{d}_{\mathrm{uc}}\right)^{\mathrm{T}} \widetilde{C}_{\mathrm{D}}^{-1} \left(\widetilde{H} y - \widetilde{d}_{\mathrm{uc}}\right). \qquad (25)$$

Minimizing $\widetilde{O}(y)$ is equivalent to assimilating $\widetilde{d}_{\mathrm{uc}}$ using $\mu_y^f = y^f$ with the KF [52]. In the second part of the derivation, we show that $\widetilde{y}^a$ has the correct posterior pdf. Note that $\widetilde{y}^a$ is a Gaussian random vector because y^f and $\widetilde{d}_{\mathrm{uc}}$ are Gaussian random vectors.

4.2.1 Part 1: Finding $\widetilde{y}^a$

Requiring the gradient of $\widetilde{O}(y)$ to vanish, solving the resulting expression for y and denoting the result as $\widetilde{y}^a$, we obtain

$$\widetilde{y}^a = C_{\mathrm{Y}}^a \left(\left(C_{\mathrm{Y}}^f \right)^{-1} y^f + \widetilde{H}^{\mathrm{T}} \widetilde{C}_{\mathrm{D}}^{-1} \widetilde{d}_{\mathrm{uc}} \right). \tag{26}$$

Appendix B presents the steps required to obtain Eq. 26.

The vectors y^f and $\widetilde{d}_{\mathrm{uc}}$ are samples of Gaussian distributions and can be obtained using the square roots of the respective covariance matrices, i.e.,

$$y^f = \mu_y^f + \left(C_{\mathrm{Y}}^f \right)^{1/2} z_y \tag{27}$$

and

$$\widetilde{d}_{\mathrm{uc}} = \widetilde{d}_{\mathrm{obs}} + \widetilde{C}_{\mathrm{D}}^{1/2} \widetilde{z}_d. \tag{28}$$

In Eqs. 27 and 28, z_y and $\widetilde{z}_d$ are normally distributed random vectors, i.e., $z_y \sim \mathcal{N}(0, I_{N_y})$ and $\widetilde{z}_d \sim \mathcal{N}(0, I_{N_d})$, where I_{N_y} is the $N_y \times N_y$ identity matrix and I_{N_d} is the $N_d \times N_d$ identity matrix. Here, $N_d = N_n \times N_a$ denotes the number of data points at the nth data assimilation time step multiplied by the number of times these data are assimilated.

Using Eqs. 27 and 28 in Eq. 26 results in our final expression for $\widetilde{y}^a$:

$$\begin{aligned} \widetilde{y}^a = {} & C_{\mathrm{Y}}^a \left(C_{\mathrm{Y}}^f \right)^{-1} \left(\mu_y^f + \left(C_{\mathrm{Y}}^f \right)^{1/2} z_y \right) \\ & + C_{\mathrm{Y}}^a \widetilde{H}^{\mathrm{T}} \widetilde{C}_{\mathrm{D}}^{-1} \left(\widetilde{d}_{\mathrm{obs}} + \widetilde{C}_{\mathrm{D}}^{1/2} \widetilde{z}_d \right). \end{aligned} \tag{29}$$

4.2.2 Part 2: Proving that the pdf for $\widetilde{y}^a$ is the correct posterior pdf

In the second part of the derivation, we show that sampling $\widetilde{y}^a$ using Eq. 29 is equivalent to sampling the posterior pdf, i.e., $\widetilde{y}^a \sim \mathcal{N}(\mu_y^a, C_{\mathrm{Y}}^a)$. Because the posterior pdf is Gaussian, we only need to show that

$$\mathrm{E}\left[\widetilde{y}^a \right] = \mu_y^a \tag{30}$$

and

$$\mathrm{cov}\left[\widetilde{y}^a \right] = \mathrm{E}\left[\left(\widetilde{y}^a - \mu_y^a \right) \left(\widetilde{y}^a - \mu_y^a \right)^{\mathrm{T}} \right] = C_{\mathrm{Y}}^a. \tag{31}$$

Using $\mathrm{E}[z_y] = 0$ and $\mathrm{E}[\widetilde{z}_d] = 0$ and Eq. 8, it follows that the expectation of $\widetilde{y}^a$ is given by

$$\begin{aligned} \mathrm{E}\left[\widetilde{y}^a \right] &= C_{\mathrm{Y}}^a \left(C_{\mathrm{Y}}^f \right)^{-1} \left(\mu_y^f + \left(C_{\mathrm{Y}}^f \right)^{1/2} \mathrm{E}[z_y] \right) \\ &\quad + C_{\mathrm{Y}}^a \widetilde{H}^{\mathrm{T}} \widetilde{C}_{\mathrm{D}}^{-1} \left(\widetilde{d}_{\mathrm{obs}} + \widetilde{C}_{\mathrm{D}}^{1/2} \mathrm{E}\left[\widetilde{z}_d \right] \right) \\ &= C_{\mathrm{Y}}^a \left(C_{\mathrm{Y}}^f \right)^{-1} \mu_y^f + C_{\mathrm{Y}}^a \widetilde{H}^{\mathrm{T}} \widetilde{C}_{\mathrm{D}}^{-1} \widetilde{d}_{\mathrm{obs}} \\ &= C_{\mathrm{Y}}^a \left(C_{\mathrm{Y}}^f \right)^{-1} \mu_y^f \\ &\quad + C_{\mathrm{Y}}^a \left[H^{\mathrm{T}} \cdots H^{\mathrm{T}} \right] \begin{bmatrix} \frac{1}{N_a} C_{\mathrm{D}}^{-1} d_{\mathrm{obs}} \\ \vdots \\ \frac{1}{N_a} C_{\mathrm{D}}^{-1} d_{\mathrm{obs}} \end{bmatrix} \\ &= C_{\mathrm{Y}}^a \left(\left(C_{\mathrm{Y}}^f \right)^{-1} \mu_y^f + H^{\mathrm{T}} C_{\mathrm{D}}^{-1} d_{\mathrm{obs}} \right) \\ &= \mu_y^a. \end{aligned} \tag{32}$$

Note that embedded in Eq. 32 is the equality

$$\mu_y^a = C_{\mathrm{Y}}^a \left(\left(C_{\mathrm{Y}}^f \right)^{-1} \mu_y^f + \widetilde{H}^{\mathrm{T}} \widetilde{C}_{\mathrm{D}}^{-1} \widetilde{d}_{\mathrm{obs}} \right). \tag{33}$$

To obtain the posterior covariance, we first compute $\widetilde{y}^a - \mu_y^a$ by subtracting Eq. 33 from Eq. 29, which leads to

$$\widetilde{y}^a - \mu_y^a = C_{\mathrm{Y}}^a \left[\left(C_{\mathrm{Y}}^f \right)^{-\mathrm{T}/2} z_y + \widetilde{H}^{\mathrm{T}} \widetilde{C}_{\mathrm{D}}^{-\mathrm{T}/2} \widetilde{z}_d \right]. \tag{34}$$

In Eq. 34, we allowed the possibility that the square root of the covariance matrices C_{Y}^f and $\widetilde{C}_{\mathrm{D}}$ are based on the Cholesky decomposition, i.e., $C_{\mathrm{Y}}^f = (C_{\mathrm{Y}}^f)^{1/2} (C_{\mathrm{Y}}^f)^{\mathrm{T}/2}$ and $\widetilde{C}_{\mathrm{D}} = \widetilde{C}_{\mathrm{D}}^{1/2} \widetilde{C}_{\mathrm{D}}^{\mathrm{T}/2}$, instead of requiring the symmetric square roots from the spectral decomposition of C_{Y}^f and $\widetilde{C}_{\mathrm{D}}$. For this reason, we obtained the transposes in Eq. 34.

Using Eq. 34 in Eq. 31, we obtain

$$\begin{aligned} \mathrm{cov}\left[\widetilde{y}^a \right] = {} & \mathrm{E}\left[\left(\widetilde{y}^a - \mu_y^a \right) \left(\widetilde{y}^a - \mu_y^a \right)^{\mathrm{T}} \right] \\ = {} & C_{\mathrm{Y}}^a \left(C_{\mathrm{Y}}^f \right)^{-1/2} \mathrm{E}\left[z_y z_y^{\mathrm{T}} \right] \left(C_{\mathrm{Y}}^f \right)^{-\mathrm{T}/2} C_{\mathrm{Y}}^a \\ & + C_{\mathrm{Y}}^a \left(C_{\mathrm{Y}}^f \right)^{-1/2} \mathrm{E}\left[z_y \widetilde{z}_d^{\mathrm{T}} \right] \widetilde{C}_{\mathrm{D}}^{-\mathrm{T}/2} \widetilde{H} C_{\mathrm{Y}}^a \\ & + C_{\mathrm{Y}}^a \widetilde{H}^{\mathrm{T}} \widetilde{C}_{\mathrm{D}}^{-1/2} \mathrm{E}\left[\widetilde{z}_d z_y^{\mathrm{T}} \right] \left(C_{\mathrm{Y}}^f \right)^{-\mathrm{T}/2} C_{\mathrm{Y}}^a \\ & + C_{\mathrm{Y}}^a \widetilde{H}^{\mathrm{T}} \widetilde{C}_{\mathrm{D}}^{-1/2} \mathrm{E}\left[\widetilde{z}_d \widetilde{z}_d^{\mathrm{T}} \right] \widetilde{C}_{\mathrm{D}}^{-\mathrm{T}/2} \widetilde{H} C_{\mathrm{Y}}^a. \end{aligned} \tag{35}$$

Noting that

$$\mathrm{E}\left[z_y z_y^{\mathrm{T}}\right] = I_{N_y}, \tag{36}$$

$$\mathrm{E}\left[\widetilde{z}_d \widetilde{z}_d^{\mathrm{T}}\right] = I_{N_d}, \tag{37}$$

$$\mathrm{E}\left[z_y \widetilde{z}_d^{\mathrm{T}}\right] = 0 \tag{38}$$

and

$$\mathrm{E}\left[\widetilde{z}_d z_y^{\mathrm{T}}\right] = 0, \tag{39}$$

Eq. 35 reduces to

$$\mathrm{cov}\left[\widetilde{y}^a\right] = C_{\mathrm{Y}}^a \left(\left(C_{\mathrm{Y}}^f\right)^{-1} + \widetilde{H}^{\mathrm{T}} \widetilde{C}_{\mathrm{D}}^{-1} \widetilde{H}\right) C_{\mathrm{Y}}^a. \tag{40}$$

In the previous section, we showed that

$$\widetilde{C}_{\mathrm{Y}}^a = \left(\left(C_{\mathrm{Y}}^f\right)^{-1} + \widetilde{H}^{\mathrm{T}} \widetilde{C}_{\mathrm{D}}^{-1} \widetilde{H}\right)^{-1} = C_{\mathrm{Y}}^a. \tag{41}$$

Using this result in Eq. 40, we obtain

$$\mathrm{cov}\left[\widetilde{y}^a\right] = C_{\mathrm{Y}}^a \left(C_{\mathrm{Y}}^a\right)^{-1} C_{\mathrm{Y}}^a = C_{\mathrm{Y}}^a, \tag{42}$$

which completes the proof.

This result shows that in order to sample the posterior pdf for the linear-Gaussian case with MDA, we need to perturb the observations using the inflated covariance matrix, $\widetilde{C}_{\mathrm{D}}$. Again, the proof was presented assuming the correct full-rank matrix C_{Y}^f but will apply for EnKF when $N_e \to \infty$.

4.3 Interpretation of multiple data assimilations for the nonlinear case

In the previous sections, we established the equivalence between single and multiple data assimilation for the linear-Gaussian case. However, for the nonlinear case, this equivalence does not hold. Intuitively, we can expect some benefit from assimilating data multiple times because we replace one potentially large update in the state vector by multiple smaller updates.

It is well-known that when using the GN method for history matching reservoir models, convergence problems can occur due to overcorrection in the model parameters at early iterations. This overcorrection may result in unreasonably small or large values of some model parameters [29, 50]. This is particularly true when the initial guess for the GN iterative process gives predicted data far from the observations. Reynolds et al. [36] showed that EnKF is similar to applying GN sequentially, with a full step and replacing the sensitivity matrix by an average sensitivity matrix obtained from the ensemble. Because of this similarity between the EnKF and GN method, it is reasonable to expect that EnKF analysis may also result in overcorrection in the states, especially if the predicted data are far from the observations at a particular data assimilation time step.

One way to ameliorate overcorrection when applying gradient-based minimization is to increase the variance of the measurement errors [19, 50]. By using an artificially high value of data measurement errors during early iterations, the objective function minimized becomes more nearly quadratic so Newton-type methods work well. As the algorithm improves the data matches, we can eventually use the correct measurement error covariance matrix [19].

Li et al. [29] showed that the Levenberg–Marquardt (LM) algorithm provides a natural way to avoid these convergence difficulties observed in the GN method, but it is clear that any trust-region method [31, 32] could be used to achieve the same result. For the case where the number of measurements is less than the number of model parameters, a convenient way to write the LM update equation for sampling with RML is [53]

$$\begin{aligned} y_j^{\ell+1} = y_j^{\ell} &+ \frac{y_j^f - y_j^{\ell}}{1+\lambda_{\ell}} + C_{\mathrm{Y}}^f H_{\ell}^{\mathrm{T}} \\ &\times \left[(1+\lambda_{\ell}) C_{\mathrm{D}} + H_{\ell} C_{\mathrm{Y}}^f H_{\ell}^{\mathrm{T}}\right]^{-1} \\ &\times \left\{\frac{H_{\ell}\left(y_j^{\ell} - y_j^f\right)}{1+\lambda_{\ell}} + d_{\mathrm{uc},j} - d_j^{\ell}\right\}, \end{aligned} \tag{43}$$

where $\lambda_{\ell} \geq 0$ is the LM parameter. Note that by choosing $\lambda_{\ell} = 0$, Eq. 43 becomes the GN update equation with a full step. Because we can interpret EnKF as one GN iteration with initial guess $y_j^0 = y_j^f$ and average sensitivity matrix, $\overline{H}$, we can write the EnKF analogous LM equation as

$$\begin{aligned} y_j^1 = y_j^f &+ C_{\mathrm{Y}}^f \overline{H}^{\mathrm{T}} \left[(1+\lambda_0) C_{\mathrm{D}} + \overline{H} C_{\mathrm{Y}}^f \overline{H}^{\mathrm{T}}\right]^{-1} \\ &\times \left(d_{\mathrm{uc},j} - d_j^f\right). \end{aligned} \tag{44}$$

The similarity between Eq. 44 and the EnKF analysis equation (Eq. 7) is evident. Using $y_j^1 = y_j^a$, Eq. 44 represents the EnKF analysis equation with the covariance of measurement errors increased by the factor $1 + \lambda_0$. Hence, we can interpret EnKF-MDA as applying the first iteration of the LM method N_a times with $\lambda_0 = N_a - 1$. Note that because Eq. 44 is missing terms involving $y_j^{\ell} - y_j^f$ (Eq. 43), this is not the same as applying multiple consecutive LM iterations.

Another well-known advantage of the LM algorithm over the GN method is that increasing λ_{ℓ} tends to

decrease the condition number of the matrix inverted in Eq. 43. A similar argument can be used for EnKF-MDA. To illustrate this argument, consider the case where the covariance matrix of the measurement errors is a diagonal matrix given by $C_D = \sigma_d^2 I_{N_n}$. It is straightforward to show that in this case, the condition number of the matrix $C = C_D + HC_Y^f H^T$ is given by

$$\kappa(C) = \frac{\beta_{max} + \sigma_d^2}{\beta_{min} + \sigma_d^2}, \tag{45}$$

where β_{max} and β_{min}, respectively, are the largest and the smallest eigenvalues of $HC_Y^f H^T$. For EnKF-MDA, we have a matrix $\widetilde{C} = N_a C_D + HC_Y^f H^T$ with condition number given by

$$\kappa(\widetilde{C}) = \frac{\beta_{max} + N_a\sigma_d^2}{\beta_{min} + N_a\sigma_d^2}, \tag{46}$$

which is a decreasing function of N_a. Thus, $\kappa(\widetilde{C}) < \kappa(C)$ for $N_a > 1$. Also note that $\kappa(\widetilde{C}) \to 1$ as $N_a \to \infty$.

5 Computational implementation

The general EnKF-MDA algorithm follows:

1. Choose the number of data assimilations, N_a.
2. Set $m_j^{n,1} = m_j^{n,f}$ and $p_j^{n-1,1} = p_j^{n-1,f}$.
3. For $\ell = 1$ to N_a do:
   For $j = 1$ to N_e do:
   (a) Run the forward model from the last data assimilation time step, t_{n-1}, until the next data assimilation time step, t_n, and store the predicted data vector $d_j^{n,\ell}$.
   (b) If $\ell < N_a$, then build the forecast state vector, $y_j^{n,\ell}$, as

$$y_j^{n,\ell} = \begin{bmatrix} m_j^{n,\ell} \\ p_j^{n-1,\ell} \\ d_j^{n,\ell} \end{bmatrix}, \tag{47}$$

   else, build $y_j^{n,\ell}$ as

$$y_j^{n,\ell} = \begin{bmatrix} m_j^{n,\ell} \\ p_j^{n,\ell} \\ d_j^{n,\ell} \end{bmatrix}. \tag{48}$$

   (see comment 4 of Section 5.1.)
   (c) Perturb the observation vector using

$$d_{uc,j}^n = d_{obs}^n + \sqrt{N_a}\left(C_D^n\right)^{1/2} z_n, \tag{49}$$

   where $z_n \sim \mathcal{N}(0, I_{N_n})$.
   (d) Update the state vector using

$$y_j^{n,\ell+1} = y_j^{n,\ell} + C_Y^{n,\ell} H_n^T \times \left(N_a C_D^n + H_n C_Y^{n,\ell} H_n^T\right)^{-1} \times \left(d_{uc,j}^n - d_j^{n,\ell}\right). \tag{50}$$

   end (for).
   end (for).
4. Set $y_j^{n,a} = y_j^{n,\ell}$ and go to the next data assimilation time step.

5.1 Comments about the EnKF-MDA algorithm

1. The algorithm presented above refers to one data assimilation time step. The same algorithm is applied for all time steps.
2. Perhaps, the most important thing to clarify about the implementation of EnKF-MDA algorithm is that $d_j^{n,\ell}$ is included in $y_j^{n,\ell}$ to keep the notation consistent with Eqs. 5 and 7, where we used the augmented state vector to present EnKF equations. However, $d_j^{n,\ell}$ and $p_j^{n,\ell}$ are computed by running the forward model from time step t_{n-1} to t_n using the reservoir state $p_j^{n-1,\ell}$ at time step t_{n-1}. Additional clarification is given below.
3. With EnKF-MDA, at each data assimilation time step, we do not assimilate the observed data N_a times simultaneously. Instead, we assimilate data N_a times consecutively and, after each of the N_a data assimilations, we rerun the forward model (reservoir simulator) starting from the previous time step for the ensemble with the updated state vectors. With this procedure, we are effectively updating the "average sensitivity" before the next data assimilation. In a sense, EnKF-MDA can be interpreted as an iterative form of EnKF where the number of iterations is chosen a priori. Note that in the derivations presented for the linear-Gaussian case, we assimilate data N_a times simultaneously. However, consecutive and simultaneous data assimilation are equivalent [13, Chap. 7]. Hence, the derivations are still valid for assimilating data N_a times consecutively for the linear-Gaussian case.
4. In order to use MDA sequentially in time, it is necessary to introduce an additional modification in the data assimilation process to keep the updated ensemble of model parameters statistically consistent with the state of the dynamical system (primary variables of the reservoir simulator). The rigorous way to do this would be to rerun the reservoir simulator from time zero after each data

assimilation, but this is computationally expensive. Instead, in our EnKF-MDA algorithm, we update the primary variables ($p_j^{n-1,\ell}$) at time step t_{n-1} with Eq. 50 for the first $N_a - 1$ data assimilations. These primary variables are used to restart the reservoir simulations during the MDA loop. In the last data assimilation, we update the vector of primary variables at the time step t_n, i.e., $p_j^{n,\ell}$, which is used to restart simulations in the next time step.

5. As in standard implementations of EnKF, $d_j^{n,\ell}$ does not need to be updated when using Eq. 7.
6. The computational cost of the proposed method is roughly N_a times the computational cost of data assimilation with standard EnKF.
7. For each of the N_a data assimilations, we recompute the perturbed observation vector, i.e., we resample $d^n_{\mathrm{uc},j} \sim \mathcal{N}(d^n_{\mathrm{obs}}, N_a C^n_{\mathrm{D}})$ instead of using the same $d^n_{\mathrm{uc},j}$ for all N_a data assimilations. Intuitively, we expect that this procedure improves sampling because we reduce bias possibly introduced by matching "outliers" generated when sampling $d^n_{\mathrm{uc},j}$. Recall that the development presented in Section 4.2 indicates that we should resample d_{uc} at each of the N_a consecutive data assimilations.

6 Time-lapse seismic data

History matching time-lapse seismic data require the capability to compute seismic data from a given reservoir model. This task can be accomplished by introducing a rock–fluid model to convert the reservoir properties and the simulator primary variables (pressure and saturations) into modeled elastic properties. The most widely used model to predict the seismic response of a reservoir due to production is the Gassmann model [20]. Here, we use a petroelastic model (PEM) based on Gassmann's equation to compute synthetic seismic data. Details about the PEM implementation can be found in Emerick et al. [8]. Among the elastic properties typically used for seismic data history matching, the most common choices are pressure-wave impedance (P-impedance or acoustic impedance) and Poisson's ratio; see, e.g., [15, 21, 22, 40, 47]. However, other seismic attributes such as amplitudes [23] and time shifts [27, 41] have also been used. Fahimuddin et al. [16] investigated different kinds of seismic data for history matching with EnKF. They concluded that time-difference impedance data performed better than time-difference amplitude data.

Assimilation of seismic data introduces some challenges to the EnKF methodology. First, it may not be feasible to directly compute or "invert" the $N_n \times N_n$ matrix ($C_{\mathrm{D}} + HC^f_{\mathrm{Y}}H^{\mathrm{T}}$) in Eq. 7. In this case, the subspace inversion procedure proposed by Evensen [12] and discussed in the Appendix A can be used. Second, integrating time-lapse seismic data together with production data typically requires the use of the ensemble Kalman smoother [13, Chap. 9] because of the time-difference characteristic of the data. In this case, it is necessary to update the state vector at the current data assimilation time step, and all other time steps corresponding to times where measured seismic data are available. A simple alternative procedure is to rerun the reservoir simulation from time zero immediately before the assimilation of the time-lapse data. This procedure removes potential problems related to statistical inconsistency between the updated model and primary variables but increases the computational cost of the process. For the two synthetic cases presented in the next sections of this paper, this inconsistency is not an issue because the seismic data correspond to the time difference between a monitor survey, obtained after a period of production, and a base survey, obtained at the initial condition of the reservoir, in which the states are assumed to be known accurately. Third and perhaps the most challenging problem when assimilating seismic data with EnKF is the large amount of data available. Effectively, there are roughly the same number of seismic data points as the number of reservoir simulator gridblocks. In EnKF, the size of the ensemble limits the number of degrees of freedom available to assimilate data. Thus, the ensemble may not provide enough degrees of freedom to assimilate all the seismic information for large reservoir problems. In the extreme case, the variability of the ensemble is reduced to the point that the ensemble collapses to a single realization. Standard procedures to reduce this problem include local analysis [13] and covariance localization [25]. Sakov and Bertino [38] showed that if we apply a correlation function to taper the updates in the local analysis procedure, then local analysis gives similar results to covariance localization. Both procedures, local analysis and covariance localization, require choosing a "localization region," and the performance of these methods is dependent on this choice. There exist also other "localization" procedures which do not explicitly require the definition of a "localization region"; see, e.g., [2, 17, 44]. Nevertheless, because the objective of this paper is only to investigate the effect of MDA, we did not consider any localization procedure in the two test cases presented in the next sections. For these two test cases, we did not observe ensemble collapse after assimilation of the seismic data. Finally, as discussed before, because seismic data are available only at very few and widely spaced time instances, we no longer

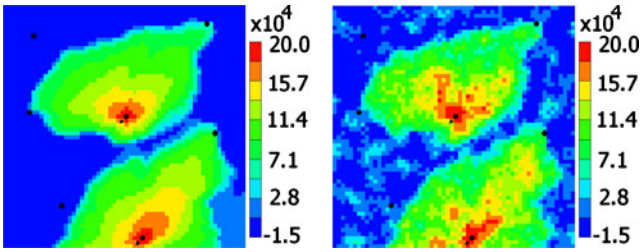
have the potentially beneficial effect of applying several consecutive data assimilation steps of data that have overlapping information content. Thus, assimilation of seismic data represents a problem in which MDA may improve the history matching results.

7 Synthetic case 1

The first test case is a two-phase (oil and water) synthetic reservoir model on a 2D uniform grid with 60×60 gridblocks. The dimensions of the gridblocks are $150 \times 150 \times 25$ ft. The model parameters are gridblock log-permeabilities, $\ln(k)$'s. The true model was generated from an anisotropic exponential correlation function with major correlation length of 3,750 ft (which corresponds to the width of 25 gridblocks) and minor correlation length of 1,050 ft (i.e., seven gridblocks) oriented at 45°. The prior mean of $\ln(k)$ is 5.0 and the prior variance is 1.0 for all gridblocks. Figure 1 shows the true permeability field used as the reference to generate the observed data. The porosities are constant and equal to 0.25 for all gridblocks. The compressibility of the rock, oil, and water is also constant and equal to 5×10^{-6}, 10^{-5}, and 10^{-6} psi^{-1}, respectively. In this model, there are five producing wells and two water injectors. All producing wells are controlled by a fixed bottomhole pressure of 1,000 psi. The injectors are controlled by a fixed bottomhole pressure of 3,000 psi.

7.1 Assimilation of seismic data

Synthetic time-lapse seismic data were generated based on the true reservoir model. The seismic data correspond to the P-impedance difference (ΔI_P) between a monitor survey after 3,900 days of production and a base survey before the beginning of the production. Figure 2a presents the true ΔI_P data. Correlated random noise was added to the true seismic data to generate the synthetic observed data. The noise was generated using an isotropic spherical covariance function with range of 750 ft (i.e., five gridblocks) and standard deviation corresponding to 30% of the average ΔI_P data ($\sigma_{d,s} = 17,320$ lb/ft^{2}s). Figure 2b shows the resulting seismic data with noise added. The seismic data used for history matching correspond to one datum per simulation gridblock.



(a) (b)

Fig. 2 P-impedance difference (lb/ft^{2}s). **a** True seismic. **b** Seismic with noise

Data assimilations with standard EnKF and EnKF-MDA assimilating data two (EnKF 2×), four (EnKF 4×), and eight (EnKF 8×) times were performed to history match the time-lapse seismic. The ensemble size is 100 and the models of the initial ensemble were generated using the same prior mean and covariance function used to generate the true model. During data assimilations, we used subspace inversion with rescaling, as discussed in the Appendix A. Truncation was done by retaining the largest singular values corresponding to 99.9% of the sum of the nonzero singular values when applying Eq. 76.

Figure 3 presents the mean ensemble predictions of ΔI_P obtained with the prior ensemble and the final ensembles after EnKF and EnKF-MDA. For visual comparison, we also include the true ΔI_P in this figure. For EnKF and EnKF-MDA, the predicted ΔI_P data were obtained by running the final ensembles from time zero after data assimilation. This figure shows clearly the improvement in the predicted ΔI_P compared to the prior ensemble. However, it is difficult to see the differences between the EnKF and EnKF-MDA cases. These differences are better visualized in the cross-plot presented in Fig. 4, which shows that increasing the number of data assimilations improved the match of time-lapse seismic. For four and eight data assimilations, the differences between the true and predicted seismic from the updated permeability fields are less than twice the standard deviation of the measurement errors for almost all reservoir gridblocks. Figure 5 presents the final mean permeability fields obtained for each case. It is interesting to note that all permeability fields present some features of the true permeability, e.g., the low permeability in the region between the wells I1 (at the center of the reservoir) and P4 (at

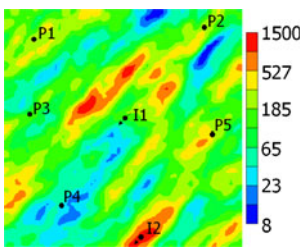


Fig. 1 True permeability field (millidarcy)

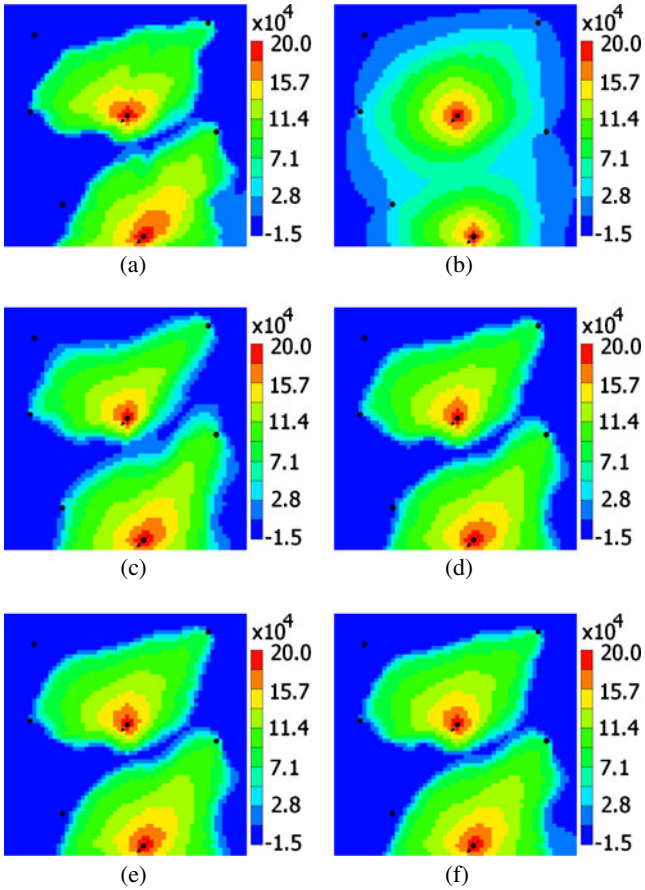


Fig. 3 P-impedance difference (pounds per square foot second). True seismic data and mean ensemble predictions. **a** True. **b** Prior. **c** EnKF 1×. **d** EnKF 2×. **e** EnKF 4×. **f** EnKF 8×

the lower left corner of the reservoir). However, the mean permeability fields obtained with EnKF-MDA are smoother than the one obtained with single data assimilation. Note, for example, that the case with single data assimilation (Fig. 5a) resulted in a relatively

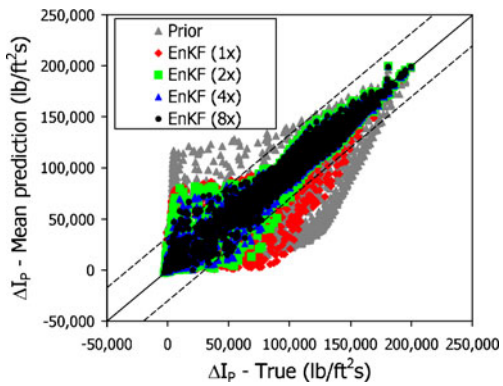


Fig. 4 Cross-plots between true and mean predicted P-impedance changes (pounds per square foot second). The *dashed lines* correspond to ±2 standard deviations of the measurement errors

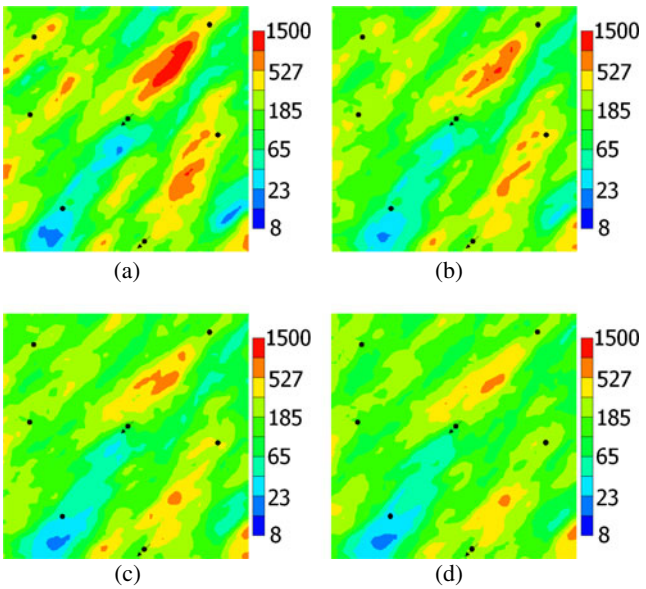


Fig. 5 Mean permeability fields (millidarcy) after assimilation of seismic data. **a** EnKF 1×. **b** EnKF 2×. **c** EnKF 4×. **d** EnKF 8×

high permeability region between wells I1 and P2 (at the upper right corner of the reservoir), whereas after four data assimilations, the permeability in this region is considerably reduced.

In order to further compare the data matches, we repeated the data assimilations with ten different initial ensembles. For each ensemble, we computed the data mismatch objective function normalized by the number of seismic data points, $O_{N,s}$, i.e.,

$$O_{N,s} = \frac{O_d(m)}{N_{d,s}}, \tag{51}$$

where $N_{d,s} = 3{,}600$ is the number of data points and $O_d(m)$ is the data mismatch objective function given by

$$O_d(m) = \frac{1}{2}\,(d - d_{\text{obs}})^{\text{T}}\, C_{\text{D}}^{-1}\,(d - d_{\text{obs}})\,. \tag{52}$$

Some comments on Eqs. 51 and 52 are in order. For the linear-Gaussian case, we define m_{c} as the random variable obtained by minimizing

$$O(m) = \frac{1}{2}(m - m_{\text{uc}})^{\text{T}} C_{\text{M}}^{-1}(m - m_{\text{uc}}) + \frac{1}{2}\,(Hm - d_{\text{uc}})^{\text{T}}\, C_{\text{D}}^{-1}(Hm - d_{\text{uc}}), \tag{53}$$

where $m_{\text{uc}} \sim \mathcal{N}(m_{\text{pr}}, C_{\text{M}})$ and $d_{\text{uc}} \sim \mathcal{N}(d_{\text{obs}}, C_{\text{D}})$. In the above equation, m is the vector of model parameters (log-permeability), m_{pr} is the vector with the prior mean, and C_{M} is the full-rank prior model covariance matrix. The remaining terms were defined before. As discussed in Oliver et al. [33, Chap. 10] as an extension of the theoretical results of Tarantola [42], $O(m_{\text{c}})$ has

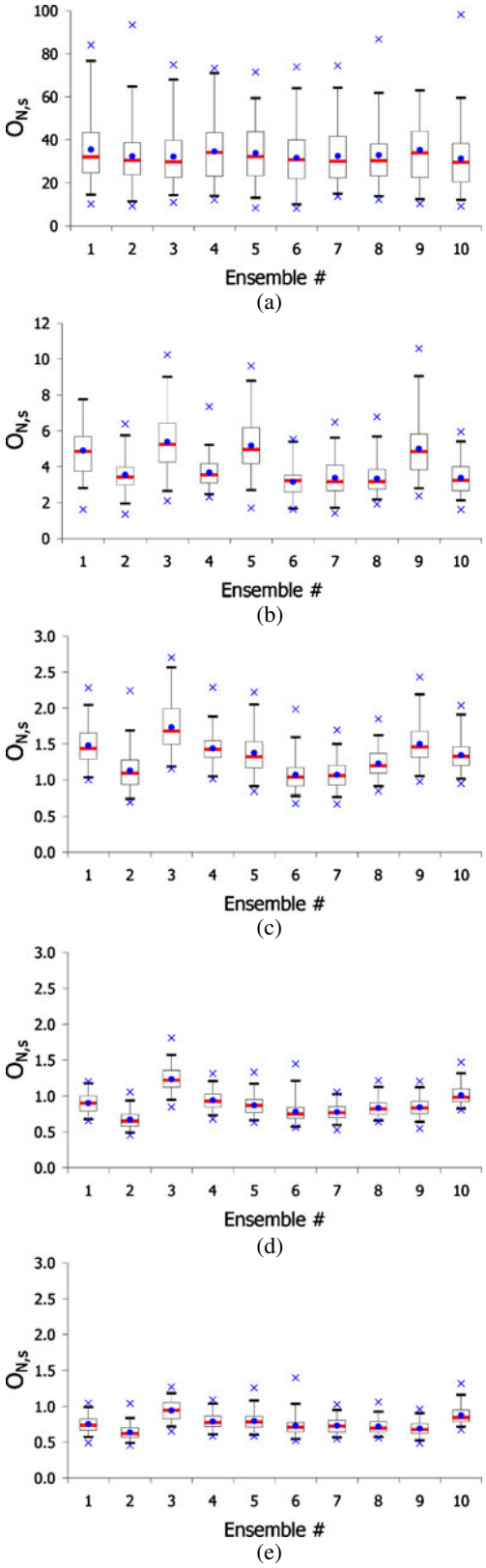


Fig. 6 Box plots of normalized seismic objective function for ten different initial ensembles. Note that the vertical scale in the plots are different. **a** Prior. **b** EnKF 1×. **c** EnKF 2×. **d** EnKF 4×. **e** EnKF 8×

a chi-squared distribution with N_d degrees of freedom, where N_d denotes the total number of data points. For N_d large, this chi-squared distribution can be approximated by a Gaussian distribution with expectation N_d and variance $2N_d$. Although computational evidence [18] suggests that this result on the distribution of $O(m_c)$ also applies for the Gaussian-nonlinear case, it has not been extended even for the linear case when m_c is restricted to a subspace spanned by N_e realization in the initial ensemble. Thus, when the conditional realizations, m_c's, are generated by EnKF, it is not clear what the expected value of $O(m_c)$ is. Thus, we have simply computed the value of $O_{N,s}$. In the special case that m is the true model and $d_{obs} \sim \mathcal{N}(d_{true}, C_D)$, with d_{true} denoting the data predicted by the true model, $O_{N,s}$ would have a chi-squared distribution with expectation 0.5. Thus, on average, we do not expect to be able to obtain a value of $O_{N,s}$ less then 0.5, but whether we can achieve an average $O_{N,s}$ of 0.5 is unclear as the conditional realizations obtained by EnKF will always lie in a subspace of dimension N_e.

Figure 6 presents the box plots of $O_{N,s}$ obtained for each ensemble. Table 1 summarizes the results for the ten ensembles. The results presented in Fig. 6 and Table 1 show that increasing the number of data assimilations resulted in consistently better seismic data matches. The biggest improvement occurred when the number of data assimilations was increased from one (average $O_{N,s} = 4.086$) to two (average $O_{N,s} = 1.335$).

Unfortunately, improving the data matches has an undesirable side effect. The variability of the final ensemble is also reduced. Figure 7 illustrates this fact by showing the standard deviation of log-permeability after data assimilation for the first of the ten ensembles considered in this section. According to results in this figure, we observe a significant reduction in the standard deviation when we increase the number of data assimilations from one (Fig. 7a) to two (Fig. 7b). Compared to EnKF 2×, EnKF 4× and EnKF 8× did not result in significant additional reductions in the standard deviation. The reduction in the variability of the final ensemble is related to the fact that EnKF

Table 1 Mean and standard deviation of $O_{N,s}$ for ten different initial ensembles

Case	Mean	Standard deviation
Prior	33.195	14.261
EnKF 1×	4.086	1.504
EnKF 2×	1.335	0.323
EnKF 4×	0.885	0.198
EnKF 8×	0.764	0.144

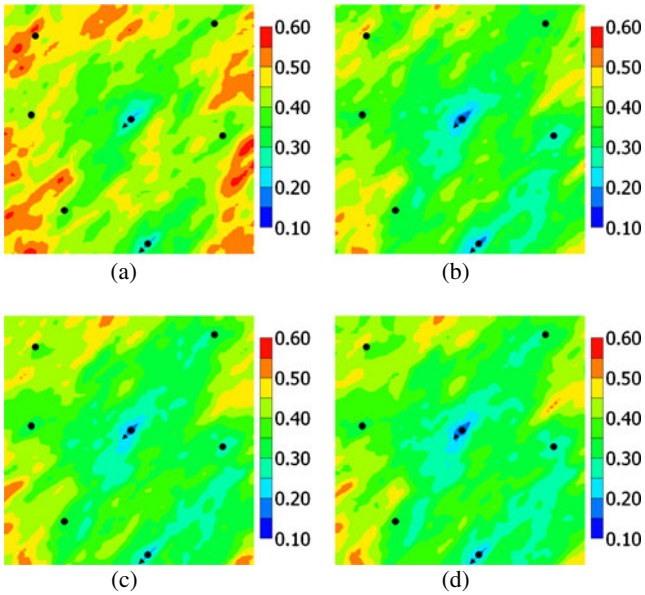


Fig. 7 Standard deviation of log-permeability after data assimilation. **a** EnKF 1×. **b** EnKF 2×. **c** EnKF 4×. **d** EnKF 8×

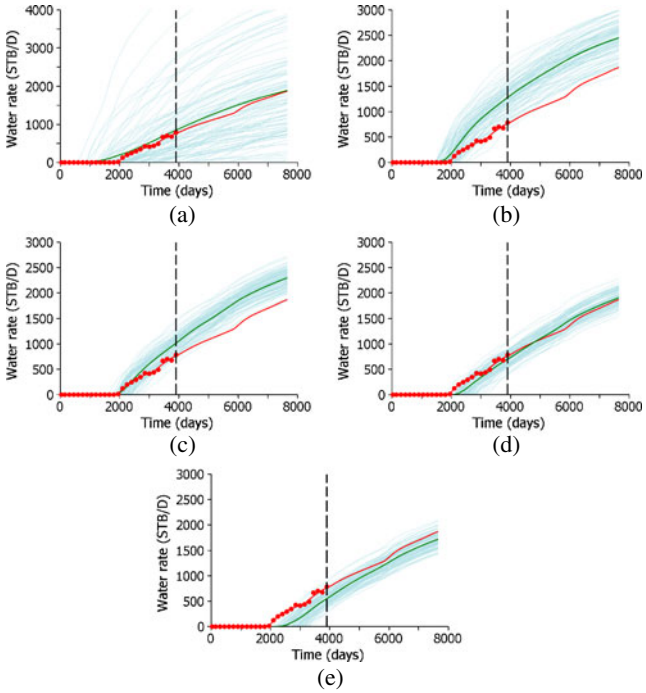


Fig. 8 Field water production rate (stock tank barrels per day) after assimilation of seismic data. The *vertical dashed line* indicates the end of the history. The *red dots* represent the production historical data, the *red curve* is the prediction from the true model, the *green curve* is the mean ensemble prediction, and the *light blue curves* are the predictions from the ensemble members. **a** Prior. **b** EnKF 1×. **c** EnKF 2×. **d** EnKF 4×. **e** EnKF 8×

"searches for solutions" in a subspace restricted by the ensemble. If we become more restrictive and search for models with better data matches, it is conceivable that there will be fewer independent models in this subspace satisfying the required level of matches, resulting in loss of variability. It is important to note that this is not a problem that occurs exclusively with EnKF-MDA. For example, in [10], we used a different iterative method based on EnKF and Markov chain Monte Carlo and obtained the same conclusion that improving the data matches by iterating with the ensemble results in additional loss of variability. In fact, it is very likely that additional loss of the ensemble variability is observed with any iterative form of EnKF which aims to improve the level of data matches. We believe that in order to obtain a reliable characterization of uncertainty in reservoir applications with ensemble-based methods, it is necessary to repeat the data assimilation many times with different initial ensembles because each ensemble tends to sample only a small region of the posterior pdf [10].

7.2 Production data prediction after assimilation of seismic data

Figure 8 presents the field water production rate obtained by simulation from time zero with the prior ensemble and the final ensembles after assimilation of the seismic data for the first of the ten ensembles considered in the previous section. In this figure, besides the historical period (3,900 days), we also include 3,750 days of forecast. It is important to emphasize that, even though Fig. 8 includes the production historical data, only seismic data were assimilated. Note that the

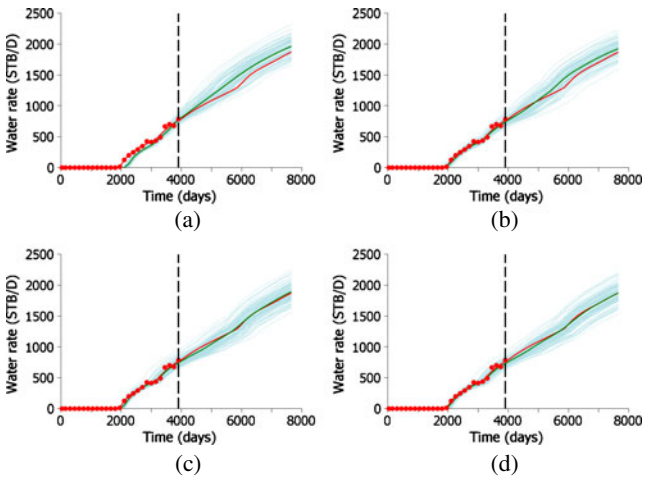


Fig. 9 Field water production rate (stock tank barrels per day) after assimilation of seismic and production data. The *colors* mean the same as in Fig. 8. **a** EnKF 1×. **b** EnKF 2×. **c** EnKF 4×. **d** EnKF 8×. Note that EnKF 2×, 4×, and 8× refer to multiple assimilations of seismic data only. The production data were assimilated with standard EnKF for all cases

Table 2 Mean and standard deviation of $O_{N,p}$ for ten different initial ensembles

Case	Mean	Standard deviation
EnKF 1×	2.893	2.946
EnKF 2×	2.230	1.806
EnKF 4×	1.780	1.460
EnKF 8×	1.533	1.135

Table 3 Mean and standard deviation of $O_{N,m}$ for ten different initial ensembles

Case	Mean	Standard deviation
EnKF 1×	0.501	0.134
EnKF 2×	0.388	0.078
EnKF 4×	0.365	0.058
EnKF 8×	0.358	0.056

assimilation of ΔI_P data with standard EnKF resulted in an ensemble which overestimated the field water production (Fig. 8b). Increasing the number of times we assimilate seismic data reduced the variance in the ensemble predictions of field water rate. After eight data assimilations, the ensemble underestimates the field water production. This may be surprising considering that we obtained a better match of seismic data with eight data assimilations. Note, however, that the results presented in Fig. 8 are after assimilating seismic data only, where no production data were used.

7.3 Assimilation of production data

In order to further investigate the effect of a better seismic data match on the production predictions, after matching seismic data, we assimilated production data using the standard EnKF (i.e., single data assimilation) starting from the final ensembles obtained after assimilation of seismic data. The observed production data correspond to 3,900 days of oil and water production rates and water injection rates. The frequency of data assimilation corresponds to one data assimilation every 150 days. Random normally distributed noise with zero mean and standard deviation equal to 5% of the true data was added to the true data to define the noisy observations.

Figure 9 presents the field water production rate obtained by running the ensembles from time zero after assimilation of the production data. The results in this figure indicate slightly better matches of production data for the ensembles with multiple assimilations of the seismic data, although the difference are not large. Table 2 presents the mean and standard deviation of the data mismatch objective function of the production data normalized by the number of production measurements ($O_{N,p} = O(m)/N_{d,p}$, with $N_{d,p} = 299$) obtained for the ten ensembles. According to results presented in the Table 2, the ensembles obtained with multiple assimilations of seismic data resulted in better production data matches as well. Figure 10 presents the final mean permeability fields for the first of the ten ensembles. Comparing Fig. 10 with Fig. 5, we observe that for the case with single assimilation of the seismic data, it was necessary to make larger changes in the permeability field to history match the production data, resulting in a rougher mean permeability field. Table 3 presents the mean and standard deviation of the normalized model mismatch objective function ($O_{N,m}$) obtained for the ten ensembles. $O_{N,m}$ was computed using

$$O_{N,m} = \frac{1}{2N_d}(m - m_{\mathrm{pr}})^{\mathrm{T}} C_{\mathrm{M}}^{-1}(m - m_{\mathrm{pr}}),$$

where $N_d = N_{d,s} + N_{d,p}$ is the total number of data points. The results of Table 3 indicate that EnKF-MDA resulted in smoother permeability fields.

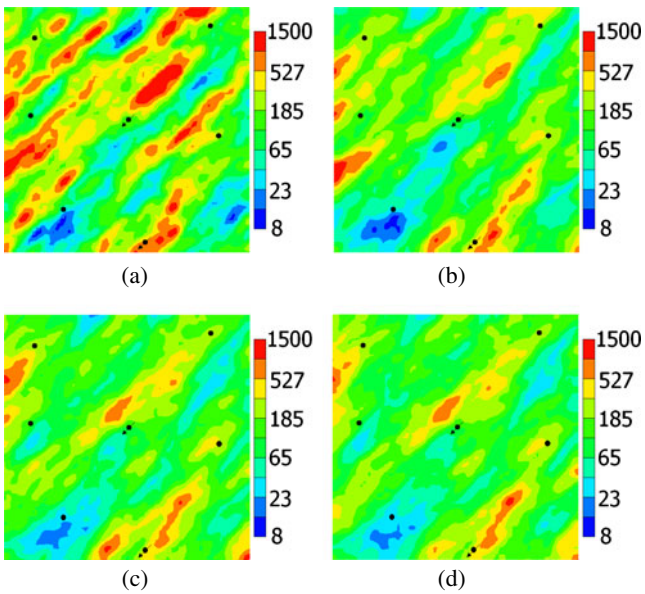


Fig. 10 Mean permeability fields (millidarcy) after assimilation of seismic and production data. **a** EnKF 1×. **b** EnKF 2×. **c** EnKF 4×. **d** EnKF 8×. Note that EnKF 2×, 4×, and 8× refer to multiple assimilations of seismic data only. The production data were assimilated with standard EnKF for all cases

Table 4 Mean and standard deviation of $O_{N,s}$ for ten different initial ensembles after assimilation of production data

Case	Mean	Standard deviation
EnKF 1×	4.104	0.997
EnKF 2×	2.043	0.679
EnKF 4×	1.343	0.430
EnKF 8×	1.187	0.416

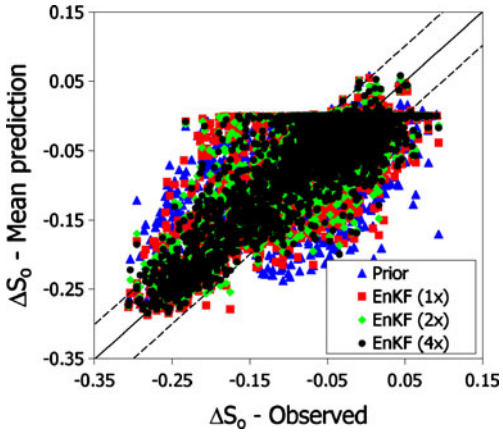


Fig. 11 Cross-plot between observed and mean predicted oil saturation changes. The *dashed lines* correspond to ±2 standard deviations of the measurement errors

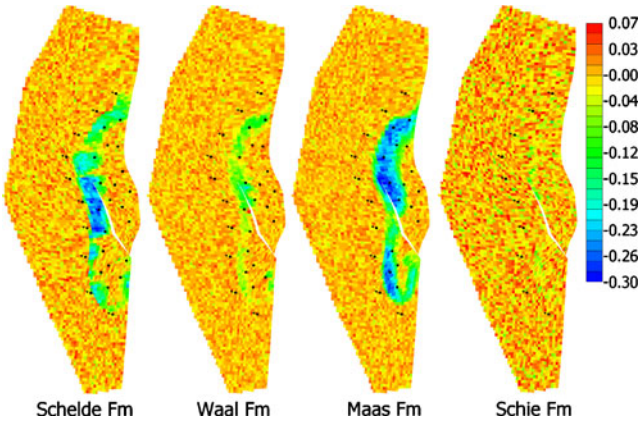


Fig. 12 Observed seismic (ΔS_o)

Table 4 presents the mean and standard deviation of the normalized seismic data mismatch objective function obtained after the assimilation of the production data. The results presented in this table indicate only a small deterioration in the seismic data matches. However, the EnKF-MDA cases still present better seismic data matches compared to standard EnKF.

8 Brugge case

We also tested EnKF-MDA on the Brugge case [34]. The Brugge case is a synthetic reservoir designed as a benchmark problem for evaluating methods for waterflooding optimization combined with history matching in a closed-loop workflow. A description of the case can be found in [34]. In the original Brugge case dataset, there are 104 realizations of rock properties (porosity, horizontal, and vertical permeabilities and net-to-gross ratio), 10 years of production history and a synthetic time-lapse seismic. The time-lapse seismic corresponds to pressure and oil saturation changes due to the 10 years of production. According to Peters et al. [34], from the nine research groups that participated in the original benchmark study, six used the seismic data in the history matching. However, no results or discussions about the seismic data matches are presented in [34] nor in other papers published by some of the groups [4, 5, 30].

According to Peters et al. [34], an unintentional bias was introduced in the pressure data because the seismic data were calculated based on an upscaled model. For this reason, here, we consider only the oil saturation data. The seismic data were provided as vertically averaged data, corresponding to the four geological zones of the reservoir. The provided seismic data were corrupted with an unknown level of noise. Here, we estimated the noise level by smoothing the observed data with a window averaging for each of the four reservoir zones and computing the residual between observed and smoothed data. We tried different sizes of the averaging window and, for each case, we computed empirical variograms. The results indicated that the noise added to the seismic seems to be spatially uncorrelated in the horizontal plane with an average standard deviation of 0.025.

Table 5 Mean and standard deviation of $O_{N,s}$ for the Brugge case

Case	Mean	Standard deviation
Prior	0.766	0.053
EnKF 1×	0.606	0.005
EnKF 2×	0.540	0.002
EnKF 4×	0.510	0.001

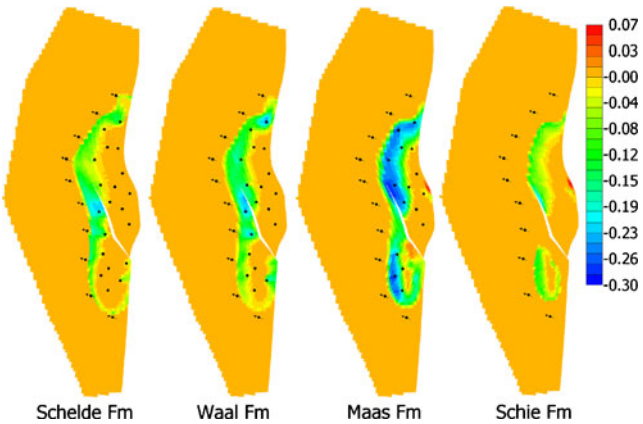


Fig. 13 Average predicted seismic from the prior ensemble (ΔS_o)

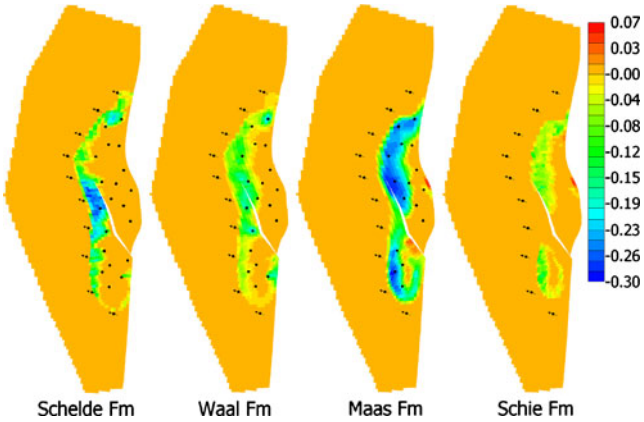


Fig. 14 Average predicted seismic from the final ensemble after EnKF (ΔS_o)

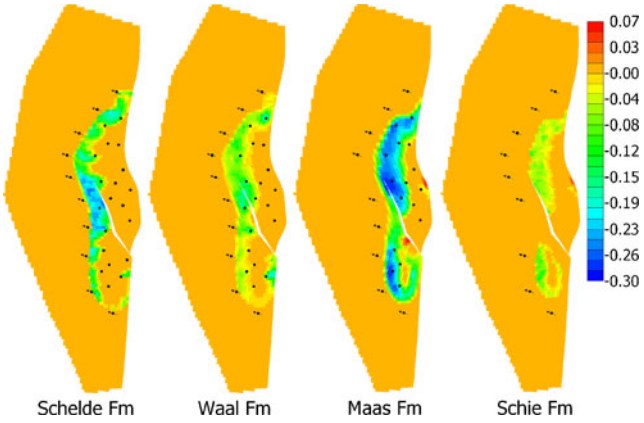


Fig. 16 Average predicted seismic from the final ensemble after EnKF with four data assimilations (ΔS_o)

In this example, we assimilated only the seismic data (ΔS_o) using standard EnKF and EnKF-MDA with two and four data assimilations (no production data are included). Figure 11 presents a cross-plot between observed and the mean predicted ΔS_o obtained with the prior ensemble and the ensembles after data assimilation. This figure indicates a slightly improvement in the predicted ΔS_o using two and four data assimilations. Table 5 presents the average values of $O_{N,s}$ for each case. Note that the values of the normalized objective function for the Brugge case are relatively small, even for the prior ensemble. This happens mainly because in the Brugge case, there is a large aquifer, in which the oil saturation change is zero, and thus seismic data pertaining to the aquifer is automatically well matched. Nevertheless, the results in Table 5 show reductions on the order of 10% (EnKF 2×) and 15% (EnKF 4×) in the values of $O_{N,s}$ compared to the standard EnKF. Figures 12, 13, 14, 15, and 16 present the observed ΔS_o and the mean predicted ΔS_o obtained with the prior ensemble and the posterior ensembles obtained with EnKF assimilating data one, two, and four times. In these figures, the oil saturation changes are presented for the four reservoir zones (Schelde, Waal, Maas, and Schie). In these figures, the negative values (green to blue colors) represent a decrease in oil saturation due to the displacement of oil by injected water.

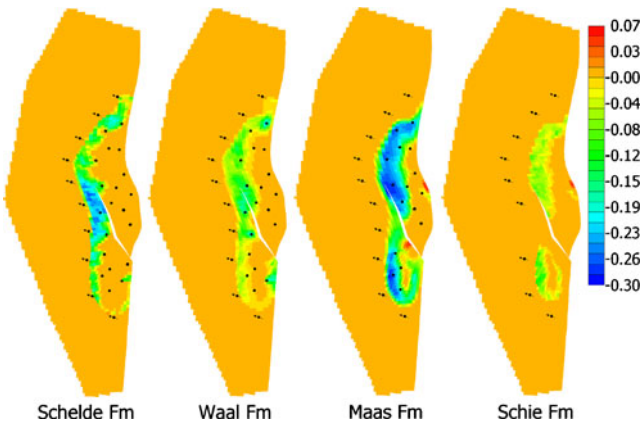


Fig. 15 Average predicted seismic from the final ensemble after EnKF with two data assimilations (ΔS_o)

9 Discussion

The proposed EnKF-MDA procedure requires very few modifications into a standard EnKF implementation. It can be interpreted as a simple, but still efficient, iterative form of EnKF. In this paper, we investigated EnKF-MDA for the case of assimilating time-lapse seismic data. However, the proposed procedure is general, and it can be used in different applications. For example, EnKF-MDA can be useful for assimilating production-logging data [7]. Production logging consists of reservoir layer rate data but collected only at a specific time. In this sense, data from production logging are similar to seismic data. In fact, Coutinho et al. [7] found that it was challenging to obtain good data matches of production-logging data using EnKF in a field case.

Although we conjecture that EnKF performs reasonably well when history matching production data because of the sequential assimilation approximately corresponds to accumulate several GN corrections, we

do not discard the possibility that EnKF-MDA can still be useful in this case. In our limited set of tests, however, we observed only small improvements in data matches when EnKF-MDA was applied to assimilate regular production data. However, the performance of MDA needs further investigation for the production data case.

Another potential application of MDA is for the ensemble smoother (ES) [45, 46]. The ES was recently applied for reservoir history matching by Skjervheim et al. [41]. In the ES, all data are assimilated at once, which means that only a single approximate GN iteration is done to history match all data. One advantage of ES over the standard EnKF is that in ES, there is no need to restart reservoir simulations every data assimilation time step. This makes ES much easier to implement and significantly faster than EnKF. Skjervheim et al. [41], for example, reported that ES required approximately 10% of the CPU time required by EnKF. Besides that, ES is an attractive option for data assimilation workflows which integrate different part of the reservoir modeling process, including seismic, structural, and geological modeling with flow simulation, as in the workflow presented by Zachariassen et al. [51]. These workflows typically require the integration of different geomodeling softwares and may include upscaling of the rock properties, which makes the simulation restarts required by EnKF very inconvenient, if not impossible. However, because in ES all data are assimilated at once, it may also result in poor data matches even compared to standard EnKF. In this situation, MDA seems to be an easy way to improve data matches obtained by the ES.

10 Conclusions

In this paper, we introduced the EnKF with multiple data assimilations. We proved that single and multiple data assimilations are equivalent for the linear-Gaussian case and presented computational evidence that multiple data assimilations can improve EnKF estimates for the nonlinear case. We applied the proposed procedure to two synthetic cases to assimilate time-lapse seismic data, and the results show better data matches when compared to standard EnKF.

Acknowledgements The support of the member companies of TUPREP is gratefully acknowledged. The first author acknowledges the financial support from Petrobras. The authors also thank to CMG for providing the licenses of IMEX reservoir simulator and the two anonymous reviewers for their valuable comments.

Appendix A: Inversion and rescaling

Pseudo-inverse

The EnKF analysis requires the inversion of the $N_n \times N_n$ matrix C given by

$$C = HC_Y^f H^T + C_D = C_{DD}^f + C_D. \tag{54}$$

Because C_{DD}^f is a real-symmetric positive semi-definite matrix, C given by Eq. 54 will be real symmetric positive definite as long as we choose C_D positive definite. However, this matrix may be poorly conditioned [13, Chap. 14]. Hence, EnKF implementations typically use a pseudo-inverse of C computed using a truncated SVD, i.e.,

$$C^+ = U_r \Lambda_r^{-1} U_r^T, \tag{55}$$

where C^+ denotes the pseudo-inverse of C. U_r is the $N_n \times N_r$ matrix with its jth column equal to the left singular vector of C corresponding to the jth singular value. Λ_r is a diagonal matrix containing the N_r largest nonzero singular values of C. N_r is typically defined by sorting the singular values, λ_i, in a decreasing order and finding the maximum N_r such that

$$\frac{\sum_{i=1}^{N_r} \lambda_i}{\sum_{i=1}^{N_n} \lambda_i} \leq \xi, \tag{56}$$

where ξ is a number typically between 0.9 and 1.0. However, C may also be poorly scaled as it may be constructed based on data with different magnitudes, e.g., pressure and water-cut data. In this case, one may lose the information necessary to match data when truncating small singular values; see, for example, Wang et al. [48] where an example is shown in which water-cut data could not be matched because of truncation. For this reason, it is important to rescale the components of the matrix C before calculating the truncated SVD. This rescaling can be done by using the Cholesky decomposition of C_D, i.e.,

$$C_D = C_D^{1/2} C_D^{T/2} \tag{57}$$

and writing Eq. 54 as

$$C = C_D^{1/2} \widetilde{C} C_D^{T/2}, \tag{58}$$

where

$$\widetilde{C} = C_D^{-1/2} C_{DD}^f C_D^{-T/2} + I_{N_n}. \tag{59}$$

Truncated SVD is now applied to the matrix $\widetilde{C}$ to obtain

$$\widetilde{C} = \widetilde{U}_r \widetilde{\Lambda}_r \widetilde{U}_r^{\mathrm{T}}, \tag{60}$$

and the pseudo-inverse of C becomes

$$C^{+} = C_{\mathrm{D}}^{-\mathrm{T}/2} \widetilde{U}_r \widetilde{\Lambda}_r^{-1} \widetilde{U}_r^{\mathrm{T}} C_{\mathrm{D}}^{-1/2}. \tag{61}$$

This rescaling procedure can be justified by defining the dimensionless sensitivity matrix, H_{D}, [54] as

$$H_{\mathrm{D}} = C_{\mathrm{D}}^{-1/2} H C_{\mathrm{Y}}^{1/2}, \tag{62}$$

For the linear case, we can write the vector of predicted data d^f in terms of the state vector y^f as

$$d^f = H y^f, \tag{63}$$

so that the cross-covariance matrix C_{DD}^f can be written as

$$C_{\mathrm{DD}}^f = H C_{\mathrm{Y}}^f H^{\mathrm{T}}. \tag{64}$$

Using Eqs. 62 and 64 in Eq. 59, we can write $\widetilde{C}$ as

$$\widetilde{C} = H_{\mathrm{D}} H_{\mathrm{D}}^{\mathrm{T}} + I_{N_n}. \tag{65}$$

Let ω_i be the ith singular value of H_{D}. Thus, it is straightforward to show that the ith singular value of $\widetilde{C}$, denoted by $\widetilde{\lambda}_i$, is given by

$$\widetilde{\lambda}_i = \omega_i^2 + 1. \tag{66}$$

The results of Tavakoli and Reynolds [43] imply that the singular values of H_{D} govern the reduction in uncertainty in the state vector due to the assimilation of data with the Kalman filter. In addition, they show that small singular values of H_{D} have negligible influence on the reduction of uncertainty. From Eq. 66, we note that the singular values of $\widetilde{C}$ are defined by the singular values of H_{D}, so that truncating small values of $\widetilde{\lambda}_i$ corresponds to eliminating small singular values of H_{D} which have the smallest influence on the reduction of uncertainty. In this sense, the rescaling procedure presented in this section is optimal.

10.1 Subspace inversion

The inversion procedure presented in the previous section requires the SVD of a $N_n \times N_n$ matrix. However, when the number of data points is large, as in the case when assimilating seismic data, the computational cost of this SVD procedure is too high. Evensen [12] introduced a subspace inversion procedure, which is computationally more efficient in the case where $N_e \ll N_n$. Here, we briefly review the subspace inversion proposed by Evensen [12] and rewrite this procedure for the case in which rescaling is applied.

In the subspace inversion, we define $\widehat{C} = (N_e - 1)C$, where C is given by Eq. 54 and write $\widehat{C}$ in the following form:

$$\widehat{C} = \Delta D^f (\Delta D^f)^{\mathrm{T}} + (N_e - 1) C_{\mathrm{D}}. \tag{67}$$

Here, $\Delta D^f = D^f - \overline{D}^f$ where D^f denotes the $N_n \times N_e$ matrix of predicted data, i.e., the columns of D^f correspond to the vectors of predicted data obtained by the ensemble members. $\overline{D}^f$ is the $N_n \times N_e$ matrix with all columns equal to the mean ensemble prediction. Instead of computing the pseudo-inverse of $\widehat{C}$ directly by truncated SVD, we apply SVD to ΔD^f and truncate with the N_r largest singular values, i.e.,

$$\Delta D^f \approx U_r W_r V_r^{\mathrm{T}}, \tag{68}$$

where U_r is the $N_n \times N_r$ matrix containing as its columns the left singular vectors of ΔD^f corresponding to its first N_r largest singular values; V_r is the $N_e \times N_r$ corresponding matrix of right singular vectors and W_r is a diagonal matrix with the N_r largest singular values of ΔD^f as its diagonal entries. Because we can choose $N_r \leq \min\{N_n, N_e - 1\}$, we write Eq. 68 as an approximation; equality holds when we keep all nonzero singular values. Using Eq. 68 in Eq. 67, we obtain

$$\widehat{C} \approx U_r W_r \left[I_{N_r} + (N_e - 1) W_r^{-1} U_r^{\mathrm{T}} C_{\mathrm{D}} U_r W_r^{-1} \right] W_r U_r^{\mathrm{T}}. \tag{69}$$

In Eq. 69, we introduced the additional approximation that $U_r U_r^{\mathrm{T}} \approx I_{N_n}$. Defining the $N_r \times N_r$ symmetric matrix X as

$$X = (N_e - 1) W_r^{-1} U_r^{\mathrm{T}} C_{\mathrm{D}} U_r W_r^{-1}, \tag{70}$$

we can write Eq. 69 as

$$\widehat{C} \approx U_r W_r \left[I_{N_r} + X \right] W_r U_r^{\mathrm{T}}. \tag{71}$$

Because X is real symmetric positive semidefinite, the SVD of X is equivalent to a Schur decomposition,

$$X = Z_r \Gamma_r Z_r^{\mathrm{T}}, \tag{72}$$

where Γ_r is a $N_r \times N_r$ diagonal matrix with the ith diagonal entry equal to the ith eigenvalue of X, which is real and non-negative, and Z_r is a $N_r \times N_r$ orthogonal matrix with its ith column equal to the ith eigenvector of X. Using Eq. 72 in Eq. 71 and the fact that Z_r is an orthogonal matrix, we obtain

$$\widehat{C} \approx (U_r W_r Z_r)\left[I_{N_r} + \Gamma_r\right](U_r W_r Z_r)^{\mathrm{T}}. \tag{73}$$

Because W_r and Γ_r are diagonal matrices, it is trivial to compute their inverses and the pseudo-inverse of $\widehat{C}$ becomes

$$\widehat{C}^{+} = \left(U_r W_r^{-1} Z_r\right)\left[I_{N_r} + \Gamma_r\right]^{-1}\left(U_r W_r^{-1} Z_r\right)^{\mathrm{T}}, \tag{74}$$

which is the pseudo-inverse using the subspace inversion scheme of Evensen [12].

As before, we may still have scaling problems when truncating the small singular values of ΔD^f. Hence, we present a subspace inversion scheme for the case in which we rescale ΔD^f before applying truncated SVD. This rescaling can be done by rewriting Eq. 67 as

$$\widehat{C} = C_{\mathrm{D}}^{1/2}\left[C_{\mathrm{D}}^{-1/2}\Delta D^f (\Delta D^f)^{\mathrm{T}} C_{\mathrm{D}}^{-\mathrm{T}/2} + (N_e - 1) I_{N_n}\right] C_{\mathrm{D}}^{\mathrm{T}/2}. \tag{75}$$

Instead of using the truncated SVD of ΔD^f, we calculate the following truncated SVD:

$$C_{\mathrm{D}}^{-1/2}\Delta D^f \approx \widetilde{U}_r \widetilde{W}_r \widetilde{V}_r^{\mathrm{T}}. \tag{76}$$

Using the same procedure as before, we obtain

$$\widehat{C} \approx C_{\mathrm{D}}^{1/2}\widetilde{U}_r \widetilde{W}_r\left[I_{N_r} + (N_e - 1)\widetilde{W}_r^{-1}\widetilde{W}_r^{-1}\right]\widetilde{W}_r \widetilde{U}_r^{\mathrm{T}} C_{\mathrm{D}}^{\mathrm{T}/2}. \tag{77}$$

Finally, the pseudo-inverse of $\widehat{C}$ can be written as

$$\widehat{C}^{+} = \left(C_{\mathrm{D}}^{-\mathrm{T}/2}\widetilde{U}_r \widetilde{W}_r^{-1}\right)\left[I_{N_r} + \widetilde{\Gamma}_r\right]^{-1}\left(C_{\mathrm{D}}^{-\mathrm{T}/2}\widetilde{U}_r \widetilde{W}_r^{-1}\right)^{\mathrm{T}}. \tag{78}$$

where $\widetilde{\Gamma}_r$ is the $N_r \times N_r$ diagonal matrix given by

$$\widetilde{\Gamma}_r = (N_e - 1)\widetilde{W}_r^{-1}\widetilde{W}_r^{-1}. \tag{79}$$

Appendix B: Matrix manipulations

Derivation of Eq. 22

Equation 22 can be obtained using Eqs. 20, 21, and 15 as follows:

$$\begin{aligned}
\widetilde{C}_{\mathrm{Y}}^{a} &= C_{\mathrm{Y}}^{f} - \left[C_{\mathrm{Y}}^{f}H^{\mathrm{T}} \cdots C_{\mathrm{Y}}^{f}H^{\mathrm{T}}\right] \\
&\quad \times \begin{bmatrix} \frac{1}{N_a}C_{\mathrm{D}}^{-1} + B & B & \cdots & B \\ B & \frac{1}{N_a}C_{\mathrm{D}}^{-1} + B & \cdots & B \\ \vdots & & \ddots & \vdots \\ B & \cdots & & \frac{1}{N_a}C_{\mathrm{D}}^{-1} + B \end{bmatrix} \\
&\quad \times \begin{bmatrix} HC_{\mathrm{Y}}^{f} \\ \vdots \\ HC_{\mathrm{Y}}^{f} \end{bmatrix} \\
&= C_{\mathrm{Y}}^{f} - \left[C_{\mathrm{Y}}^{f}H^{\mathrm{T}} \cdots C_{\mathrm{Y}}^{f}H^{\mathrm{T}}\right] \begin{bmatrix} \left(\frac{1}{N_a}C_{\mathrm{D}}^{-1} + N_a B\right)HC_{\mathrm{Y}}^{f} \\ \vdots \\ \left(\frac{1}{N_a}C_{\mathrm{D}}^{-1} + N_a B\right)HC_{\mathrm{Y}}^{f} \end{bmatrix} \\
&= C_{\mathrm{Y}}^{f} - N_a C_{\mathrm{Y}}^{f}H^{\mathrm{T}}\left(\frac{1}{N_a}C_{\mathrm{D}}^{-1} + N_a B\right)HC_{\mathrm{Y}}^{f} \\
&= C_{\mathrm{Y}}^{f} - C_{\mathrm{Y}}^{f}H^{\mathrm{T}}\left[C_{\mathrm{D}}^{-1} - \left(C_{\mathrm{D}} + C_{\mathrm{DD}}^{f}\right)^{-1} C_{\mathrm{DD}}^{f} C_{\mathrm{D}}^{-1}\right] HC_{\mathrm{Y}}^{f} \\
&= C_{\mathrm{Y}}^{f} - C_{\mathrm{Y}}^{f}H^{\mathrm{T}}\left[I_{N_n} - \left(C_{\mathrm{D}} + C_{\mathrm{DD}}^{f}\right)^{-1} C_{\mathrm{DD}}^{f}\right] C_{\mathrm{D}}^{-1} HC_{\mathrm{Y}}^{f} \\
&= C_{\mathrm{Y}}^{f} - C_{\mathrm{Y}}^{f}H^{\mathrm{T}}\left[\left(C_{\mathrm{D}} + C_{\mathrm{DD}}^{f}\right)^{-1}\left(C_{\mathrm{D}} + C_{\mathrm{DD}}^{f}\right)\right. \\
&\quad \left. - \left(C_{\mathrm{D}} + C_{\mathrm{DD}}^{f}\right)^{-1} C_{\mathrm{DD}}^{f}\right] C_{\mathrm{D}}^{-1} HC_{\mathrm{Y}}^{f} \\
&= C_{\mathrm{Y}}^{f} - C_{\mathrm{Y}}^{f}H^{\mathrm{T}}\left(C_{\mathrm{D}} + C_{\mathrm{DD}}^{f}\right)^{-1}\left[C_{\mathrm{D}} + C_{\mathrm{DD}}^{f} - C_{\mathrm{DD}}^{f}\right] \\
&\quad \times C_{\mathrm{D}}^{-1} HC_{\mathrm{Y}}^{f} \\
&= C_{\mathrm{Y}}^{f} - C_{\mathrm{Y}}^{f}H^{\mathrm{T}}\left(C_{\mathrm{D}} + HC_{\mathrm{Y}}^{f}H^{\mathrm{T}}\right)^{-1} HC_{\mathrm{Y}}^{f} \\
&= C_{\mathrm{Y}}^{a}.
\end{aligned}$$

Derivation of Eq. 26

To derive Eq. 26, we first compute the gradient of $\widetilde{O}(y)$ with respect to y and set it equal to a zero vector, i.e.,

$$\nabla_y \widetilde{O}(y) = \left(C_{\mathrm{Y}}^{f}\right)^{-1}\left(y - y^f\right) + \widetilde{H}^{\mathrm{T}}\widetilde{C}_{\mathrm{D}}^{-1}\left(\widetilde{H}y - \widetilde{d}_{\mathrm{uc}}\right) = 0. \tag{80}$$

Rearranging Eq. 80, we obtain

$$\left[\left(C_{\mathrm{Y}}^{f}\right)^{-1}+\widetilde{H}^{\mathrm{T}}\widetilde{C}_{\mathrm{D}}^{-1}\widetilde{H}\right]y=\left(C_{\mathrm{Y}}^{f}\right)^{-1}y^{f}+\widetilde{H}^{\mathrm{T}}\widetilde{C}_{\mathrm{D}}^{-1}\widetilde{d}_{\mathrm{uc}}. \quad (81)$$

Solving Eq. 81 for y and denoting the result as $\widetilde{y}^{a}$, we obtain

$$\widetilde{y}^{a}=\left[\left(C_{\mathrm{Y}}^{f}\right)^{-1}+\widetilde{H}^{\mathrm{T}}\widetilde{C}_{\mathrm{D}}^{-1}\widetilde{H}\right]^{-1}\times\left(\left(C_{\mathrm{Y}}^{f}\right)^{-1}y^{f}+\widetilde{H}^{\mathrm{T}}\widetilde{C}_{\mathrm{D}}^{-1}\widetilde{d}_{\mathrm{uc}}\right). \quad (82)$$

Using of Eqs. 9 and 22, Eq. 82 can be written as

$$\widetilde{y}^{a}=C_{\mathrm{Y}}^{a}\left(\left(C_{\mathrm{Y}}^{f}\right)^{-1}y^{f}+\widetilde{H}^{\mathrm{T}}\widetilde{C}_{\mathrm{D}}^{-1}\widetilde{d}_{\mathrm{uc}}\right), \quad (83)$$

which is Eq. 26 of the text.

References

1. Aanonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C., Vallés, B.: Review of ensemble Kalman filter in petroleum engineering. SPE J **14**(3), 393–412 (2009)
2. Anderson, J.L.: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. Physica D: Nonlinear Phenomena **230**(1–2), 99–111 (2007)
3. Burgers, G., van Leeuwen, P., Evensen, G.: Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. **126**(6), 1719–1724 (1998)
4. Chen, C., Li, G., Reynolds, A.C.: Closed-loop reservoir management on the Brugge test case. Comput. Geosci. **14**(4), 691–703 (2010)
5. Chen, Y., Oliver, D.: Ensemble-based closed-loop optimization applied to Brugge field. In: Proceedings of the SPE Reservoir Simulation Symposium. The Woodlands, Texas, USA, 2–4 February, SPE 118926 (2009)
6. Cominelli, A., Dovera, L., Vimercati, S., Nævdal, G.: Benchmark study of ensemble Kalman filter methodology: history matching and uncertainty quantification for a deep-water oil reservoir. In: Proceedings of the International Petroleum Technology Conference. Doha, Qatar, 7–9 December, IPTC 13748 (2009)
7. Coutinho, E.J., Emerick, A.A., Li, G., Reynolds, A.C.: Conditioning multi-layered geologic models to well test and production logging data using the ensemble Kalman filter. In: Proceedings of the SPE Annual Technical Conference and Exhibition, Florence, Italy, 19–22 September, SPE 134542 (2010)
8. Emerick, A.A., Moraes, R.J., Rodrigues, J.R.P.: Calculating seismic attributes within a reservoir flow simulator. In: Proceedings of the Latin American & Caribbean Petroleum Engineering Conference, Buenos Aires, Argentina, 15–18 April, SPE 107001 (2007)
9. Emerick, A.A., Reynolds, A.C.: History matching a field case using the ensemble Kalman filter with covariance localization. SPE Reserv. Evalu. Eng. **14**(4), 423–432 (2011)
10. Emerick, A.A., Reynolds, A.C.: Combining the ensemble Kalman filter with Markov chain Monte Carlo for improved history matching and uncertainty characterization. SPE J. (2012, in press)
11. Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. **99**(C5), 10,143–10,162 (1994)
12. Evensen, G.: Sampling strategies and square root analysis schemes for the EnKF. Ocean Dyn. **54**(6), 539–560 (2004)
13. Evensen, G.: Data Assimilation: The Ensemble Kalman Filter. Springer, Berlin (2007)
14. Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation. IEEE Control Syst. Mag. 83–104 (2009)
15. Fahimuddin, A., Aanonsen, S.I., Skjervheim, J.-A.: 4D seismic history matching of a real field case with EnKF: use of local analysis for model updating. In: Proceedings of the SPE Annual Technical Conference and Exhibition. Florence, Italy, 19–22 September, SPE 134894 (2010)
16. Fahimuddin, A., Aanonsen, S.I., Skjervheim, J.-A.: Ensemble based 4D seismic history matching: Integration of different levels and types of seismic data. In: Proceedings of the SPE EUROPEC/EAGE Annual Conference and Exhibition, Barcelona, Spain, 14–17 June, SPE 131453 (2010)
17. Furrer, R., Bengtsson, T.: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. J. Multivar. Anal. **98**(2), 227–255 (2007)
18. Gao, G., Li, G., Reynolds, A.C.: A stochastic algorithm for automatic history matching. SPE J. **12**(2), 196–208 (2007)
19. Gao, G., Reynolds, A.C.: An improved implementation of the LBFGS algorithm for automatic history matching. SPE J. **11**(1), 5–17 (2006)
20. Gassmann, F.: Elastic waves through a packing of spheres. Geophysics **16**, 673–685 (1951)
21. Gosselin, O., Aanonsen, S., Aavatsmark, I., Cominelli, A., Gonard, R., Kolasinski, M., Ferdinandi, F., Kovacic, L., Neylon, K.: History matching using time-lapse seismic (HUTS). In: Proceedings of the SPE Annual Technical Conference and Exhibition. Denver, Colorado, 5–8 October, SPE 84464 (2003)
22. Haverl, M., Aga, M., Reiso, E.: Integrated workflow for quantitative use of time-lapse seismic data in history matching: a North Sea field case. In: Proceedings of the SPE EUROPEC/EAGE Annual Conference, Madrid, Spain, 13–16 June, SPE 94453 (2005)
23. Haverl, M., Skjervheim, J.-A.: 4D seismic modeling integrated with the ensemble Kalman filter method for history matching of reservoir simulation model. In: Proceedings of the 11th European Conference on the Mathematics of Oil Recovery, Bergen, Norway (2008)
24. Houtekamer, P.L., Mitchell, H.L.: Data assimilation using an ensemble Kalman filter technique. Mon. Weather Rev. **126**(3), 796–811 (1998)
25. Houtekamer, P.L., Mitchell, H.L.: A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. **129**(1), 123–137 (2001)
26. Kalman, R.E.: A new approach to linear filtering and prediction problems. Transactions of the ASME, Journal of Basic Engineering **82**, 35–45 (1960)
27. Kjelstadli, R., Lane, H., Johnson, D., Barkved, O., Buer, K., Kristiansen, T.: Quantitative history match of 4D seismic response and production data in the Valhall field. In: Proceedings of the Offshore Europe, Aberdeen, UK, 6–9 September, SPE 96317 (2005)
28. Li, G., Reynolds, A.C.: Iterative ensemble Kalman filters for data assimilation. SPE J. **14**(3), 496–505 (2009)
29. Li, R., Reynolds, A.C., Oliver, D.S.: History matching of three-phase flow production data. SPE J. **8**(4), 328–340 (2003)

30. Lorentzen, R.J., Shafieirad, A., Nævdal, G.: Closed loop reservoir management using the ensemble Kalman filter and sequential quadratic programming. In: Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, 2–4 February, SPE 119101 (2009)
31. Moré, J.J., Sorensen, D.: Computing a trust region step. SIAM J. Sci. Comput. **4**, 553–572 (1983)
32. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (1999)
33. Oliver, D.S., Reynolds, A.C., Liu, N.: Inverse Theory for Petroleum Reservoir Characterization and History Matching. Cambridge University Press, Cambridge (2008)
34. Peters, L., Arts, R., Brouwer, G., Geel, C., Cullick, S., Lorentzen, R., Chen, Y., Dunlop, K., Vossepoel, F., Xu, R., Sarma, P., Alhuthali, A., Reynolds, A.: Results of the Brugge benchmark study for flooding optimisation and history matching. SPE Reserv. Evalu. Eng. **13**(3), 391–405 (2010)
35. Reynolds, A.C., He, N., Oliver, D.S.: Reducing uncertainty in geostatistical description with well testing pressure data. In: Schatzinger, R.A., Jordan, J.F. (eds.) Reservoir Characterization—Recent Advances, pp. 149–162. American Association of Petroleum Geologists, Tulsa (1999)
36. Reynolds, A.C., Zafari, M., Li, G.: Iterative forms of the ensemble Kalman filter. In: Proceedings of 10th European Conference on the Mathematics of Oil Recovery, Amsterdam, 4–7 September (2006)
37. Rommelse, J.: Data Assimilation in Reservoir Management. Ph.D. thesis, Technical University of Delft, Delft, The Netherlands (2009)
38. Sakov, P., Bertino, L.: Relation between two common localisation methods for the EnKF. Comput. Geosci. **15**(2), 225–237 (2011)
39. Seiler, A., Evensen, G., Skjervheim, J.-A., Hove, J., Vabø, J.: Advanced reservoir management workflow using an EnKF based assisted history matching method. In: Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, 2–4 February, SPE 118906 (2009)
40. Skjervheim, J.-A., Evensen, G., Aanonsen, S.I., Ruud, B.O., Johansen, T.A.: Incorporating 4D seismic data in reservoir simulation models using ensemble Kalman filter. SPE J. **12**(3), 282–292 (2007)
41. Skjervheim, J.-A., Evensen, G., Hove, J., Vabø, J.G.: An ensemble smoother for assisted history matching. In: Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, USA, 21–23 February, SPE 141929 (2011)
42. Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM, Philadelphia (2005)
43. Tavakoli, R., Reynolds, A.C.: History matching with parameterization based on the SVD of a dimensionless sensitivity matrix. SPE J. **15**(12), 495–508 (2010)
44. Vallès, B.V., Nævdal, G.: Comparing different ensemble Kalman filter approaches. In: Proceedings of the 11th European Conference of the Mathematics of Oil Recovery, ECMOR XI, Bergen, Norway, B07 (2008)
45. van Leeuwen, P.J.: An ensemble smoother with error estimates. Mon. Weather Rev. **129**, 709–728 (2001)
46. van Leeuwen, P.J., Evensen, G.: Data assimilation and inverse methods in terms of a probabilistic formulation. Mon. Weather Rev. **124**, 2898–2913 (1996)
47. Waggoner, J., Cominelli, A., Seymour, R.: Improved reservoir modeling with time-lapse seismic in a Gulf of Mexico gas condensate reservoir. In: SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 29 September–2 October, SPE 77514 (2002)
48. Wang, Y., Li, G., Reynolds, A.C.: Estimation of depths of fluid contacts by history matching using iterative ensemble-Kalman smoothers. SPE J. **15**(2) (2010)
49. Whitaker, J.S., Hamill, T.M.: Ensemble data assimilation without perturbed observations. Mon. Weather Rev. **130**(7), 1913–1924 (2002)
50. Wu, Z., Reynolds, A.C., Oliver, D.S.: Conditioning geostatistical models to two-phase production data. SPE J. **3**(2), 142–155 (1999)
51. Zachariassen, E., Skjervheim, J., Vabø, J., Lunt, I., Hove, J., Evensen, G.: Integrated work flow for model update using geophysical monitoring data. In: Proceedings of the 73rd EAGE Conference & Exhibition, Vienna, Austria, 23–26 May (2011)
52. Zafari, M., Reynolds, A.C.: Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter, SPE J. **12**(3), 382–391 (2007)
53. Zhang, F., Reynolds, A.C.: Optimization algorithms for automatic history matching of production data. In: Proceedings of 8th European Conference on the Mathematics of Oil Recovery, Freiberg, Germany, 3–6 September (2002)
54. Zhang, F., Reynolds, A.C., Oliver, D.S.: Evaluation of the reduction in uncertainty obtained by conditioning a 3D stochastic channel to multiwell pressure data. Math. Geol. **34**(6), 713–740 (2002)