

Model-reduced gradient-based history matching

Małgorzata P. Kaleta · Remus G. Hanea ·
Arnold W. Heemink · Jan-Dirk Jansen

Received: 7 December 2009 / Accepted: 21 July 2010 / Published online: 5 August 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Gradient-based history matching algorithms can be used to adapt the uncertain parameters in a reservoir model using production data. They require, however, the implementation of an adjoint model to compute the gradients, which is usually an enormous programming effort. We propose a new approach to gradient-based history matching which is based on model reduction, where the original (nonlinear and high-order) forward model is replaced by a linear reduced-order forward model and, consequently, the adjoint of the tangent linear approximation of the original forward model is replaced by the adjoint of a linear reduced-order forward model. The reduced-order model is constructed with the aid of the proper orthogonal decomposition method. Due to the linear

character of the reduced model, the corresponding adjoint model is easily obtained. The gradient of the objective function is approximated, and the minimization problem is solved in the reduced space; the procedure is iterated with the updated estimate of the parameters if necessary. The proposed approach is adjoint-free and can be used with any reservoir simulator. The method was evaluated for a waterflood reservoir with channelized permeability field. A comparison with an adjoint-based history matching procedure shows that the model-reduced approach gives a comparable quality of history matches and predictions. The computational efficiency of the model-reduced approach is lower than of an adjoint-based approach, but higher than of an approach where the gradients are obtained with simple finite differences.

M. P. Kaleta · A. W. Heemink
Faculty of Electrical Engineering, Mathematics and
Computer Science, Delft Institute of Applied Mathematics,
Delft University of Technology, Mekelweg 4, 2628 CD,
Delft, The Netherlands

M. P. Kaleta
e-mail: m.p.kaleta@tudelft.nl

R. G. Hanea
TNO Built Environment and Geosciences,
Business Unit Geo Energy and Geo Information, TNO,
Princetonlaan 6, 3584 CB,
Utrecht, The Netherlands

J.-D. Jansen (✉)
Faculty of Civil Engineering and Geosciences,
Department of Geotechnology,
Delft University of Technology,
Stevinweg 1, 2628 CN,
Delft, The Netherlands
e-mail: j.d.jansen@tudelft.nl

Keywords Data assimilation · History matching ·
Model reduction · Proper orthogonal decomposition ·
Adjoint-free

1 Introduction

In typical reservoir simulation models, the uncertain parameters are, e.g., grid block permeabilities and porosities, relative permeabilities, or fault multipliers. A prior reservoir characterization is generally based on localized borehole and outcrop observations that are interpolated to give regional descriptions of uncertain geological properties. The interpolation process introduces uncertainty in those parameters that directly translates into an uncertainty about the reservoir behavior. The geological model can be improved by using surface and downhole production data (flow

rates and bottom hole pressures) in a process called *history matching*. History matching identifies the parameter values that minimize an *objective function* that represents the mismatch between modeled and observed production data. Usually the objective function is defined as a sum of weighted squared differences between observed and modeled data.

Due to the high computing cost of a reservoir model simulation, history matching procedures need to involve a reasonable number of model simulations to be usable. History matching has been investigated for the last few decades, and many algorithms have been developed, among them, *gradient-based history matching algorithms*. A particularly efficient class of optimization methods are those where a gradient of the objective function with respect to the model parameters is calculated by solving the adjoint, or co-state, problem as introduced by Courant and Hilbert [11]. In reservoir engineering, the adjoint method was used for the first time by Chen et al. [9] and later applied by, among others, Chavent et al. [8], Wasserman et al. [42], Watson et al. [43], Lee and Seinfeld [25], Yang et al. [44], Zhang and Reynolds [47], Li et al. [26], and Oliver et al. [31]. In the adjoint approach, the history matching problem is treated as an optimal control problem where the control variables are the unknown model parameters and where the objective function is minimized subject to the constraint that the state variables obey the prescribed reservoir model. Another use of adjoint-based methods in reservoir engineering is for *recovery optimization* or *production optimization*. In that case, the objective function is ultimate recovery or net present value, and the control variables are the well rates, well pressures, or well valve settings. Initially, this was done for the optimization of tertiary recovery processes, see Ramirez [32], later followed by water flooding optimization, see, e.g., [3, 5, 34, 36, 45] and [15]. In both the history matching and the recovery optimization problems, the necessary conditions for optimality lead to the gradients, which are now the total derivatives with respect to the controls of the objective function, modified to include the reservoir model constraint. These gradients can subsequently be used in gradient-based optimization methods. The adjoint approach is computationally very efficient because one gradient calculation requires just a single simulation of the forward model and a single simulation of the adjoint model backward in time, irrespective of the number of parameters. Usually, first-order gradient-based minimization algorithms, which avoid explicit computation of the Hessian, are used to perform the minimization process. In particular with the limited-memory Broyden–Fletcher–Goldfarb–Shanno or the Levenberg–Marquardt algorithm, the

history matching process can be performed efficiently (see for further details [31]). This approach, however, does have a few drawbacks: It can converge to a local minimum rather than the global one, and it requires the implementation of an adjoint model. Nevertheless, if correctly implemented, it is one of the most efficient approaches existing today to solve the history matching problem. Usually, in reservoir models, the Jacobian matrices of the system are available because they are used in Newton–Raphson iteration during the forward simulation. Even so, the implementation of the adjoint equations is an immense programming effort which, moreover, requires access to the simulation code. This implies that there is a need for gradient-based, but adjoint-free optimization methods, a requirement that becomes even more pressing if reservoir simulation is combined with another forward simulation, e.g., of geomechanics or rock physics, with a code for which no Jacobians are available. We note that the naive solution of computing the gradients through finite differences, i.e., by perturbing the parameters one by one and computing the associated changes in the objective function, is not a realistic option for reservoir models which may have up to millions of parameters and which may require several hours for a single forward simulation. A somewhat more feasible finite-difference approach is one where reparameterization is used to reduce the number of parameters, e.g., with the aid of a zonation presented by Jahns [23] or with the aid of a limited number of spatial basis functions that are linear combinations of the original parameters presented by Dadashpour [13]. However, also in this case, the number of forward simulations required for the finite-difference approach may be too large for realistic applications.

In reservoir engineering, other gradient-based, adjoint-free methods have been proposed in recent years. Gao et al. [19] used the simultaneous perturbation stochastic approximation method of Spall [35] where an approximate gradient is calculated using a stochastic perturbation of all parameters together. The same method was applied by Wang et al. [41] for the recovery optimization problem. Another method uses an ensemble of control variables and computes the cross-covariance between the control variables and the objective function which can then be used as an approximate gradient. This “ensemble optimization” technique was proposed for recovery optimization by Lorentzen et al. [28] and thereafter refined by Chen et al. [10]. For history matching, a similar approximate gradient-based technique has gained enormous popularity over the past years in the form of the ensemble Kalman filter (EnKF); see [16] and [1] for recent

overviews. Although somewhat disguised, because the EnKF assimilates data sequentially rather than all together, the basic idea of replacing an exact gradient by an ensemble-based correlation is also present in this method.

Another way to avoid the implementation of the adjoint model for the original reservoir model is to replace it with simplified model, for which the adjoint derivation is simpler. This approach has been proposed in weather prediction modeling by Courtier et al. [12] as a solution to the computationally expensive large-scale inverse problem. They derived an incremental approach, in which the forward solution of a high-resolution nonlinear model is replaced by the solution of a lower resolution, approximate linear model. The disadvantage of this method is the need to acquire a linear model, which is a simplified approximation of the original model (with simplified physical parameterizations). Another way to obtain the approximated model is by the use of a reduced order modeling technique called the proper orthogonal decomposition (POD) method, also known as the Karhunen–Loève method, principal component analysis or the method of empirical orthogonal functions. It is a data-driven projection-based method. The origin of this method goes back to 1946, when it was introduced independently by Karhunen [24] and Loève [27] as a statistical tool to analyze random process data. The method was called for the first time the POD by Lumley [29], when it was used for the study of turbulent flow. In the POD method, a low-order projection subspace is determined by processing data obtained from numerical simulations of the high-dimensional model, which are expected to provide relevant information about the dynamic behavior of the system. The high-dimensional equations are projected on the low-dimensional subspace resulting in a low-dimension model. The application of the POD method reduces the CPU time of the model simulation but does not change the complexity of the problem and, consequently, does not solve the implementation problem of the adjoint model. The POD method has been successfully applied to increase the computational efficiency of reservoir simulation by Heijn et al. [21], Markovinovic and Jansen [30], and Cardoso et al. [7]. Use of the POD method in combination with adjoint-based optimization has been investigated by, e.g., Cao et al. [6] and Fang et al. [17] in oceanography and by Van Doren et al. [38] in reservoir simulation.

Vermeulen and Heemink [40] proposed an approach based on model reduction that avoids an implementation of the adjoint model of the tangent linear approximation of the original nonlinear model and that shifts

the minimization into lower dimensional space. First, they construct the reduced-order tangent linear approximation of the original forward model and, thereafter, thanks to the linear character of the reduced-order model, they easily derive the adjoint model. In this approach, the POD method is used to obtain an approximate low-dimensional version of the tangent-linear model. This is a somewhat different procedure than the conventional application of the POD method for state-space reduction. Afterward, the adjoint of the reduced-order linear model can be easily implemented and due to the reduced model size the minimization problem is solved efficiently. Vermeulen and Heemink [40] applied this method to groundwater flow problems where it proved to be computationally very efficient. Later, it was applied by Altaf et al. [2] for water depth estimation in a coastal engineering problem and also there it proved to be very efficient. In this paper, we will apply this method to the reservoir model history matching problem. This new algorithm is not better in terms of efficiency or robustness than the classical adjoint-based approach, but it does not require the implementation of the adjoint code, it operates in a low-dimensional space, which makes it much more efficient than the naive finite-difference method, and it can easily be implemented with any simulator without having access to simulation source code.

The paper is arranged as follows: The history matching problem is defined in Section 2, the classical adjoint-based history matching formulation is presented in Section 3, and the proposed gradient-based history matching formulation using model reduction is described in detail in Section 4. Section 5 contains results from two different twin data assimilation experiments using the adjoint-based approach, a finite-difference-based approach, and the model reduction-based approach.

2 History matching as inverse modeling

The discrete model for a single simulation step of the reservoir system from time t_{i-1} to time t_i can be described by an equation of the form

$$\mathbf{x}(t_i) = \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}], i = 1, \dots, N, \quad (1)$$

where $\mathbf{x}(t_i) \in X \subset \mathbb{R}^n$ denotes the state vector of pressures and saturations at time t_i , $\boldsymbol{\theta}$ denotes the vector of uncertain parameters, and N denotes the total number of simulation time steps. The dynamic operator $\mathbf{f}_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$ represents the reservoir simulator and is nonlinear and deterministic. See, e.g., [4] for further details. The relationship between measured production

data $\mathbf{y}(t_i)$ and state variables $\mathbf{x}(t_i)$ can be described by a nonlinear operator $\mathbf{h}_i : \mathbb{R}^n \rightarrow \mathbb{R}^m$, which represents the so-called *well model* (see, e.g., [4] for further details). If we assume that observations are imperfect, then the simulated measurements are described by

$$\mathbf{y}(t_i) = \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}] + \mathbf{v}_i, i = 1, \dots, N_o, \quad (2)$$

where \mathbf{v}_i is the observation error and N_o is the number of time steps where the observations are taken. We note that although Eqs. 1 and 2 are presented in explicit form, they will usually be implicit for realistic implementations. However, this does not influence the validity of our theory, and we use explicit formulations to simplify the notation.

The uncertain parameters in the model can be estimated by minimizing an objective function that measures the difference between simulated and observed data \mathbf{d} . In the case of assimilation of production data, the data are sparse and it is not possible to correctly estimate all parameters from this information. The parameter-estimation problem is ill-posed and can therefore result in nonunique parameter estimates. Moreover, such estimates are often geologically unrealistic. One way to make the history matching problem well-posed is to rely on some background information in the form of a prior estimate of the model parameters [20]. The objective function then consists of two terms—a background (prior) term and an observation term:

$$\begin{aligned} J(\boldsymbol{\theta}, \mathbf{x}(t_i)) &= \frac{1}{2}(\boldsymbol{\theta}_{\text{init}}^b - \boldsymbol{\theta})^T \mathbf{R}_b^{-1}(\boldsymbol{\theta}_{\text{init}}^b - \boldsymbol{\theta}) \\ &+ \frac{1}{2} \sum_{i=1}^{N_o} [\mathbf{d}(t_i) - \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]]^T \\ &\times \mathbf{R}_i^{-1} [\mathbf{d}(t_i) - \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]], \end{aligned} \quad (3)$$

where \mathbf{R}_i is the covariance matrix of the observation errors at time t_i , $\boldsymbol{\theta}_{\text{init}}^b$ represents the prior parameters, and \mathbf{R}_b represents the covariance matrix of the prior parameter errors and thus models the uncertainty associated with the prior information. By minimizing the objective function (Eq. 3), we find a model that is close to the prior model, while it simultaneously minimizes the misfit between the data and the model results.

3 Adjoint-based history matching algorithms

The parameter estimation problem can be interpreted as a constrained minimization problem, where the con-

straints are formed by the system Eq. 1, which can be solved using iterative gradient-based minimization methods. The minimization methods consist of the following iterative steps:

- Determine a direction \mathbf{s}_i in parameters space, which leads to a lower objective function
- Find the size of the step length α_i along that direction, such that it minimizes $J(\boldsymbol{\theta}_i + \alpha_i \mathbf{s}_i)$
- Set the new estimate $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + \alpha_i \mathbf{s}_i$

The algorithm proceeds until the minimum of the objective function is found or certain stopping criteria are satisfied. To determine the direction, we need to calculate the gradient of the objective function with respect to the parameters. The gradient can be obtained by reformulating the minimization problem as an unconstrained one, through inclusion of the constraints with the aid of Lagrange multipliers; see, e.g., Oliver et al. [31]. This leads to:

$$\begin{aligned} \left(\frac{dJ}{d\boldsymbol{\theta}}\right)^T &= - \sum_{i=1}^N \left(\frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}}\right)^T \boldsymbol{\lambda}(t_i) - \mathbf{R}_b^{-1}(\boldsymbol{\theta}_{\text{init}}^b - \boldsymbol{\theta}) \\ &- \sum_{i=1}^{N_o} \left(\frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}}\right)^T \mathbf{R}_i^{-1} [\mathbf{d}(t_i) - \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]], \end{aligned} \quad (4)$$

where $\boldsymbol{\lambda}$ represents a vector of Lagrangian multipliers (or adjoint states) which satisfies the following equation:

$$\begin{aligned} \boldsymbol{\lambda}(t_i) &= \left(\frac{\partial \mathbf{f}_{i+1}[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_i)}\right)^T \boldsymbol{\lambda}(t_{i+1}) \\ &+ \left(\frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_i)}\right)^T \mathbf{R}_i^{-1} [\mathbf{d}(t_i) - \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]] \end{aligned} \quad (5)$$

for $i = N, \dots, 1$ with an end condition $\boldsymbol{\lambda}(t_{N+1}) = \mathbf{0}$.

4 Gradient-based history matching using model reduction

Classical gradient-based history matching formulated for a general model represents a nonlinear constrained optimization problem that is very difficult to solve. It can be greatly simplified with the hypothesis that the objective function can be made quadratic by assuming that the \mathbf{f}_i and \mathbf{h}_i operators can be linearized. The quadratic problem has theoretical advantages since it

involves solving only linear equations and it has a unique solution. Therefore, *the tangent linear model of the original model* is written using the first-order Taylor formula in the vicinity of the background parameter θ^b ; this results in the following equation:

$$\Delta \bar{\mathbf{x}}(t_i) = \frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_{i-1})} \Delta \bar{\mathbf{x}}(t_{i-1}) + \sum_{j=1}^p \frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \theta_j} \Delta \theta_j, \tag{6}$$

where p is the size of the parameter space, $\bar{\mathbf{x}}$ is the linearized state vector, and $\Delta \bar{\mathbf{x}}$ is a deviation of the model from the background trajectory. The partial derivatives in Eq. 6 are evaluated at $(\mathbf{x}^b(t_{i-1}), \theta^b)$, where \mathbf{x}^b is the nonlinear state vector with a parameter value θ^b , that is, $\mathbf{x}^b(t_i) = \mathbf{f}_i[\mathbf{x}^b(t_{i-1}), \theta^b]$. The Eq. 6 can be rewritten as

$$\begin{bmatrix} \Delta \bar{\mathbf{x}}(t_i) \\ \Delta \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_{i-1})} & \frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \Delta \bar{\mathbf{x}}(t_{i-1}) \\ \Delta \boldsymbol{\theta} \end{bmatrix}, \tag{7}$$

which is a linear model in terms of variations.

The minimization process is now performed by the following iterative steps:

- Linearize the model around the background geological parameters θ_k^b , where k denotes the iteration number and $\theta_1^b = \theta_{\text{init}}^b$
- Find $\Delta \theta_k$, such that it minimizes the quadratic objective function
- Update $\theta_{k+1}^b = \theta_k^b + \Delta \theta_k$

The history matching procedure turns into a scheme that consists of two loops: an inner loop, which finds the minimum of the quadratic objective function, and an outer loop, where the original model is used to redefine the model trajectory and to calculate the original objective function. It iterates until some predefined convergence criteria are met. In practice, in order to make the computation more efficient, the tangent linear model is replaced by a simpler low-resolution linear model. This is called the incremental approach. It has been described in Courtier et al. [12] and is used in operational systems for weather forecasting. The methodology presented in this paper is based on this two-loop approach. The difference is that we do not need to acquire a simpler linear model because we construct it by using a reduced-order modeling technique and information from the perturbations of the original model. Because the reduced-order model is designed to capture the

dominant dynamics of the original system, we expect that we can successfully use it for a fast reduction of the objective function in the inner loop. We note, however, that there is no guarantee that the reduced-order model will always lead to the same response to parameter variations as the full-order model.

4.1 Parameter reduction

Since the vector $\boldsymbol{\theta}$ may consists of properties in each grid block of the model, e.g., permeabilities or porosities, the number of uncertain parameters p can be very large. In that case, it is necessary to reparameterize the vector $\boldsymbol{\theta}$. In our case, we used the POD method (KL expansion) to reparameterize the parameter space and, consequently, to reduce the number of parameters to be estimated. In particular, we will consider reparameterization of the permeability field as preparation to the numerical example to be discussed below. The permeability vector $\boldsymbol{\theta}$ is approximated by

$$\boldsymbol{\theta} \approx \boldsymbol{\theta}^b + \Phi_\eta \boldsymbol{\eta} \iff \Delta \boldsymbol{\theta} = \Phi_\eta \boldsymbol{\eta}, \tag{8}$$

where the permeability patterns creating Φ_η are the first p_{red} dominant eigenvectors of a low-rank approximation of the covariance matrix of the permeability field, represented by an ensemble of prior permeability fields. Alternatively, other reparameterization methods such as the discrete cosine transform [22] or a discrete wavelet transform [33] may be chosen, as long as they lead to a significant reduction in the number of parameters.

4.2 Initialization

The initial parameter vector θ_{init}^b is chosen as a first guess of the uncertain parameters, $\theta_1^b = \theta_{\text{init}}^b$. To keep the notation simple, we skip the index indicating the outer iteration and we denote the reference parameters as θ^b . Next, the high-order reservoir model is simulated with parameters values θ^b , and the initial objective function is calculated.

4.3 Collection of snapshots and pattern selection

The POD method is used to obtain an approximate low-order version of the tangent-linear model. This is a somewhat different procedure than the conventional application of the POD method for state-space reduction. In the conventional method, a data matrix is built from selected state vectors of a forward simulation of the large-scale numerical model, called *snapshots*. Spatial basis functions are obtained by computing a

low-rank approximation of the covariance matrix of the data and by selecting the leading eigenvectors. Next, the (tangent-) linear model equations are projected on the low-order basis formed by the basis functions. It results in a reduced-order model, in terms of coefficients multiplying the basis functions, which is still able to reproduce the dominant dynamic behavior of the original model, see, e.g., [7] or [38] for further details. If the original model is nonlinear, the matrix coefficients of the tangent-linear high-order model need to be recomputed every time step before applying the reduction, which significantly reduces the computational advantage of the POD method as compared to the linear case. Moreover, the conventional POD method requires the availability of the high-order tangent model, i.e., of the Jacobians with respect to the states. In our approach, we aim at obtaining a reduced-order approximation of the tangent linear model directly by computing approximate derivatives of the reduced-order model.

Since the reduced-order model is used for parameter estimation, the snapshots should be able to represent the behavior of the system for modified parameter values. Therefore, snapshots are created in the following way:

$$\mathbf{x}_j(t_i) = \mathbf{f}_i[\mathbf{x}^b(t_{i-1}), \boldsymbol{\theta}^b + \boldsymbol{\phi}_\eta^j \eta_j] - \mathbf{f}_i[\mathbf{x}^b(t_{i-1}), \boldsymbol{\theta}^b], \quad (9)$$

where $\boldsymbol{\phi}_\eta^j$ is the j th pattern in matrix Φ_η and η_j is the size of the perturbation. Then snapshots are put as columns in matrix

$$\mathbf{X} = \{\mathbf{x}_1(t_1), \dots, \mathbf{x}_1(t_N), \dots, \mathbf{x}_{p_{\text{red}}}(t_1), \dots, \mathbf{x}_{p_{\text{red}}}(t_N)\} \quad (10)$$

of size $n \times s$, where s is the total number of snapshots. The given set of snapshots spans a subspace on which the reduced-order model is built. To find a basis of this subspace, the eigenvalue problem

$$(\mathbf{X}\mathbf{X}^T)\boldsymbol{\phi}_i = \lambda_i\boldsymbol{\phi}_i; \quad i \in \{1, \dots, s\} \quad (11)$$

should be solved. The eigenvectors corresponding to the largest eigenvalues of matrix $\mathbf{X}\mathbf{X}^T$ represent the dominant directions present in the snapshots, called *patterns*. For each pattern, its relative importance α_i , defined as the percentage of the eigenvalues in the total sum of eigenvalues, is calculated, that is, $\alpha_i = \frac{\lambda_i}{\sum_{j=1}^s \lambda_j}$. By choosing the n_{red} patterns with the largest relative importance an orthonormal projection matrix Φ is created by taking vectors $\boldsymbol{\phi}_i$ as its columns, for $i \in \{1, \dots, n_{\text{red}}\}$, where $n_{\text{red}} \leq s$. The snapshots reconstructed from only n_{red} patterns approximate the elements in the set \mathbf{X} optimally (in some average sense). It means that if

$\hat{\mathbf{x}}_j$ is a reconstruction of $\mathbf{x}_j \in \mathbf{X}$ defined as a linear combination of n_{red} patterns, then the average error $\frac{1}{s} \sum_{j=1}^s \|\mathbf{x}_j - \hat{\mathbf{x}}_j\|$ is minimal among any other linear reconstruction of the dimension n_{red} .

Since typically $s \ll n$, instead of solving the eigenvalue problem given by Eq. 11, the reduced eigenvalue problem

$$(\mathbf{X}^T\mathbf{X})\mathbf{w}_i = \lambda_i\mathbf{w}_i; \quad i \in \{1, \dots, s\} \quad (12)$$

is solved. The eigenvalues of those problems are the same, and the following relationship holds between the eigenvectors: $\boldsymbol{\phi}_i = \lambda_i^{-\frac{1}{2}}\mathbf{X}\mathbf{w}_i$, for $i \in \{1, \dots, s\}$. Alternatively, instead of solving the eigenvalue problem involving the product of the matrices $(\mathbf{X}\mathbf{X}^T$ or $\mathbf{X}^T\mathbf{X})$, a singular value decomposition directly on the snapshot matrix \mathbf{X} can be performed. Then the matrix \mathbf{X} is decomposed as $\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, where $\mathbf{U} \in \mathbb{R}^{n \times n}$, $\mathbf{V} \in \mathbb{R}^{s \times s}$ are orthogonal matrices and $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times s}$ is a pseudodiagonal having on its diagonal nonnegative elements arranged in decreasing order. It can be shown that the eigenvectors of the matrix $\mathbf{X}\mathbf{X}^T$ can be computed as the left singular vectors, i.e., elements of \mathbf{U} , and their sorted eigenvalues are equal to squared elements of the diagonal of $\boldsymbol{\Sigma}$.

In order to select the number n_{red} , we use the relative importance α_i . n_{red} is chosen as the smallest number, such that $\sum_{j=1}^{n_{\text{red}}} \alpha_j \geq \alpha^e$, where α^e is some percentage of the total relative importance. The accuracy of a reduced model increases with an increase in the number of patterns in the matrices Φ . For a given number of snapshots, the highest accuracy is achieved by taking α^e equal to one. On the other hand, the use of more patterns means lower computational efficiency. Therefore, there is a trade-off between efficiency and accuracy. One can notice that if the last eigenvector which is included in the set of patterns Φ has the same eigenvalue as the next eigenvector which is not selected because the required level of relative importance is satisfied, then this procedure is not uniquely defined and depends on the sorting routine used to order the eigenvalues. Hence, one may check the next eigenvalue and include a corresponding eigenvector if the eigenvalues are equal.

4.4 Building the reduced-order linear model

A model can be reduced by reconstructing the incremental state $\Delta\bar{\mathbf{x}}(t_i)$ according to

$$\Delta\bar{\mathbf{x}}(t_i) = \Phi\mathbf{z}(t_i), \quad (13)$$

where Φ is the projection matrix derived in the previous subsection and $\mathbf{z} \in \mathbb{R}^{n_{\text{red}}}$ is the reduced state vector. Combining Eqs. 13, 6, and 8 results in

$$\mathbf{z}(t_i) = \mathbf{N}_i \mathbf{z}(t_{i-1}) + \mathbf{N}_i^\theta \boldsymbol{\eta}, \tag{14}$$

where

$$\mathbf{N}_i = \Phi^T \frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_{i-1})} \Phi, \tag{15}$$

$$\mathbf{N}_i^\theta = \Phi^T \frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \Phi_\eta, \tag{16}$$

and where \mathbf{z} represents the response of the original model to a perturbation of parameters, such that we can write:

$$\begin{bmatrix} \mathbf{z}(t_i) \\ \boldsymbol{\eta} \end{bmatrix} = \begin{bmatrix} \mathbf{N}_i & \mathbf{N}_i^\theta \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{z}(t_{i-1}) \\ \boldsymbol{\eta} \end{bmatrix}. \tag{17}$$

The partial derivatives in formulas 15 and 16 can not be easily computed explicitly and should be approximated by

$$\begin{aligned} & \left[\frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \right]_j \\ & \approx \frac{\mathbf{f}_i[\mathbf{x}^b(t_{i-1}), \boldsymbol{\theta}^b + \epsilon_\theta \mathbf{i}_j^\theta] - \mathbf{f}_i[\mathbf{x}^b(t_{i-1}), \boldsymbol{\theta}^b]}{\epsilon_\theta}, \end{aligned} \tag{18}$$

and

$$\begin{aligned} & \left[\frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_{i-1})} \right]_j \\ & \approx \frac{\mathbf{f}_i[\mathbf{x}^b(t_{i-1}) + \epsilon \mathbf{i}_j, \boldsymbol{\theta}^b] - \mathbf{f}_i[\mathbf{x}^b(t_{i-1}), \boldsymbol{\theta}^b]}{\epsilon} \end{aligned} \tag{19}$$

where ϵ and ϵ_θ are the intervals in which the partial differentials are linearized and \mathbf{i}_j and \mathbf{i}_j^θ are the j th columns of identity matrices with dimension, $n \times n$ and $p \times p$, respectively. Of course, those perturbations are too expensive for large-scale models. From Eq. 15, however, we see that there is no need to calculate those partial derivatives explicitly, and instead of approximating terms $\frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_{i-1})}$, we approximate the terms $\frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_{i-1})} \Phi$ as

$$\begin{aligned} & \frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_{i-1})} \boldsymbol{\phi}_j \\ & \approx \frac{\mathbf{f}_i[\mathbf{x}^b(t_{i-1}) + \epsilon \boldsymbol{\phi}_j, \boldsymbol{\theta}^b] - \mathbf{f}_i[\mathbf{x}^b(t_{i-1}), \boldsymbol{\theta}^b]}{\epsilon}, \end{aligned} \tag{20}$$

where $\boldsymbol{\phi}_j$ is the j th column of Φ . In the same manner, we approximate the terms $\frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \Phi_\eta$ as

$$\begin{aligned} & \frac{\partial \mathbf{f}_i[\mathbf{x}(t_{i-1}), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \boldsymbol{\phi}_\eta^j \\ & \approx \frac{\mathbf{f}_i[\mathbf{x}^b(t_{i-1}), \boldsymbol{\theta}^b + \epsilon_\theta \boldsymbol{\phi}_\eta^j] - \mathbf{f}_i[\mathbf{x}^b(t_{i-1}), \boldsymbol{\theta}^b]}{\epsilon_\theta}, \end{aligned} \tag{21}$$

where $\boldsymbol{\phi}_\eta^j$ is the j th column of Φ_η . It results in a reduced-order model which is linearized along the patterns. The dimension of the reduced-order model depends on the number of patterns n_{red} and on the number of variables p_{red} to be estimated, which are both expected to be small (see Section 4.1). Moreover, the matrices in the reduced-order model are known explicitly and the reduced-order model is linear. Therefore, the CPU time of a single reduced-order model simulation is negligible. However, the computational overhead to construct the reduced-order model may be significant.

We described how to build the reduced-order linear model, which is a low-order approximation to the high-order tangent linear approximation of the original reservoir model. We presented the formulation of the method using an explicit form of the reservoir equations motivated by the resulting simplification in the description of the method. An explicit method calculates the state of a system at a current time step from the parameters and the state of the system at the previous time step, while an implicit method finds a solution by solving an equation involving both the current state of the system and the previous one. Since, the tangent linear model in both cases can be rewritten such that it calculates the approximate state at a current time step from the approximate state at the previous time step, and from the parameters, the choice of an explicit or an implicit method is irrelevant.

4.5 Improving parameters in reduced space

When the reduced-order approximate linear model is available, the adjoint model is easy to implement (due to its linear character) and is governed by the following equation

$$\begin{aligned} \underline{\boldsymbol{\lambda}}(t_i) = & \mathbf{N}_{i+1}^T \underline{\boldsymbol{\lambda}}(t_{i+1}) + \left(\frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_i)} \Phi \right)^T \\ & \times \mathbf{R}_i^{-1} \left[\mathbf{d}(t_i) - \mathbf{h}_i[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b] - \frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_i)} \Phi \mathbf{z}(t_i) \right. \\ & \left. - \frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \Phi_\eta \boldsymbol{\eta} \right], \end{aligned} \tag{22}$$

where $\underline{\lambda}$ represents a vector of the reduced adjoint state and $\underline{\lambda}(t_{N+1}) = \mathbf{0}$. The objective function for the reduced-order model is

$$\begin{aligned} \hat{J}(\boldsymbol{\eta}, \mathbf{z}(t_i)) &= \frac{1}{2} \sum_{i=1}^{N_o} \left[\mathbf{d}(t_i) - \mathbf{h}_i[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b] - \frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_i)} \Phi \mathbf{z}(t_i) \right. \\ &\quad \left. - \frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \Phi_\eta \boldsymbol{\eta} \right]^T \\ &\times \mathbf{R}_i^{-1} \left[\mathbf{d}(t_i) - \mathbf{h}_i[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b] - \frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_i)} \Phi \mathbf{z}(t_i) \right. \\ &\quad \left. - \frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \Phi_\eta \boldsymbol{\eta} \right] \\ &+ \frac{1}{2} [\boldsymbol{\theta}_{\text{init}}^b - \boldsymbol{\theta}^b - \Phi_\eta \boldsymbol{\eta}]^T \mathbf{R}_b^{-1} [\boldsymbol{\theta}_{\text{init}}^b - \boldsymbol{\theta}^b - \Phi_\eta \boldsymbol{\eta}], \end{aligned} \tag{23}$$

and the gradient is

$$\begin{aligned} \left(\frac{d\hat{J}}{d\boldsymbol{\eta}} \right)^T &= - \sum_{i=1}^N (\mathbf{N}_i^\theta)^T \underline{\lambda}(t_i) - \Phi_\eta^T \mathbf{R}_b^{-1} [\boldsymbol{\theta}_{\text{init}}^b - \boldsymbol{\theta}^b - \Phi_\eta \boldsymbol{\eta}] \\ &\quad - \sum_{i=1}^{N_o} \left(\frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \Phi_\eta \right)^T \\ &\quad - \mathbf{R}_i^{-1} \left[\mathbf{d}(t_i) - \mathbf{h}_i[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b] - \frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_i)} \Phi \mathbf{z}(t_i) \right. \\ &\quad \left. - \frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \Phi_\eta \boldsymbol{\eta} \right]. \end{aligned} \tag{24}$$

Here, we can approximate the partial derivatives of the observation model in the same manner as we did for the forward model, namely

$$\frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_i)} \boldsymbol{\phi}_j \approx \frac{\mathbf{h}_i[\mathbf{x}^b(t_i) + \epsilon \boldsymbol{\phi}_j, \boldsymbol{\theta}^b] - \mathbf{h}_i[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b]}{\epsilon} \tag{25}$$

and

$$\frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \boldsymbol{\phi}_\eta^j \approx \frac{\mathbf{h}_i[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b + \epsilon_\theta \boldsymbol{\phi}_\eta^j] - \mathbf{h}_i[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b]}{\epsilon_\theta}. \tag{26}$$

The minimization is performed using a quasi-Newton optimization where the Hessian of the objective function is updated using the BFGS method; see [18]. No scaling of the parameters in BFGS is used. Since each iteration of the line search algorithm requires a run of a simulator, we need to minimize that number of

iterations. Therefore, instead of using an exact line search, we use a backtracking search routine, which is best suited to be used with quasi-Newton optimization algorithms; see [14]. The minimization algorithm requires convergence criteria to terminate. We decided that the algorithm terminates when

- The objective function almost ceases to change, i.e.,

$$\frac{|\hat{J}(\boldsymbol{\eta}_{k+1}) - \hat{J}(\boldsymbol{\eta}_k)|}{\max\{|\hat{J}(\boldsymbol{\eta}_{k+1})|, 1\}} < \epsilon_J, \tag{27}$$

- The estimates almost stop to change, i.e.,

$$\frac{|\boldsymbol{\eta}_{k+1} - \boldsymbol{\eta}_k|}{\max\{|\boldsymbol{\eta}_{k+1}|, 1\}} < \epsilon_p, \tag{28}$$

where $\epsilon_J = 10^{-4}$ and $\epsilon_p = 10^{-4}$.

4.6 Improving parameters in high-order space

The solution of the inner loop is an optimum for the reduced-order linearized system but not necessarily for the original one. Therefore, we check the value of the original objective function for the new parameters $\boldsymbol{\theta}_{i+1}^b = \boldsymbol{\theta}^b + \Phi_\eta \boldsymbol{\eta}_{opt}$, where $\boldsymbol{\eta}_{opt}$ is the solution of minimization process. The process is repeated as long as the objective function satisfies

$$mN_o - 2\sqrt{2mN_o} \leq 2J(\boldsymbol{\theta}_{i+1}^b) \leq mN_o + 2\sqrt{2mN_o}. \tag{29}$$

This convergence criterion is a consequence of the assumption that the model is linear and $\boldsymbol{\theta}$ has a normal distribution. Moreover, model errors are not considered. In this case, the minimum of objective function J has a χ^2 distribution with mN_o degrees of freedom. It is reasonable to assume that this fact also approximately applies in the nonlinear case. The constraint (Eq. 29) tells us that the objective function should fall in the interval of two standard deviations ($2\sqrt{2mN_o}$) distance from the mean expected value (mN_o). If J has a χ^2 distribution with mN_o degrees of freedom, then as mN_o tends to infinity, the distribution of J tends to a standard normal distribution. This property justifies the interval of two standard deviations, which implies that if we would repeat the numerical experiment with different random measurement errors drawn from the same distribution, about 95% of the results would be within two standard deviations away from the mean. If the objective function does not obey the stopping criterion, then the process is iterated with a new background parameter $\boldsymbol{\theta}_{i+1}^b$. Since a relationship exists between patterns and model parameterization, whenever the background parameter is changed, a new background state and a new set of patterns need to be identified. Thereafter, a new reduced-order model

is built and the inner loop is performed again. The advantage of this criterion compared to the alternative criteria is that it works even if only a single outer loop is performed. The history matching algorithm described in this section has been summarized by the flow chart in Fig.1.

4.7 Computational complexity

The overall computational costs of the model-reduced approach consists of two parts:

- (a) Preprocessing costs of constructing the reduced-order model:
 - Cost of generating representative snapshots which span a large portion of the variability of the permeability field
The generation of “good” snapshots is an important part of the POD method. Our strategy is as follows: First, we check the range of the permeability values that are present in the ensemble of prior realizations. Then, we choose a number of permeability patterns,

which we change during one simulation. We try to reduce the CPU time by changing a few patterns per simulation during initial outer loops. The size of the perturbation depends on the number of outer iteration as well, namely we perturb with the maximum range in first loop and we reduce it in subsequent loops. We have not yet established a fully systematic approach which guarantees the best compromise between efficiency and representativeness of the snapshots, but from our experience, the described approach works well.

- Cost of solving the reduced eigenvalue problem
This cost is equivalent to the cost of a singular value decomposition of matrix \mathbf{X} , which is $O(n \times s^2)$ [37]. Since in our case s is a low number, this cost is low. Moreover, we may use a Lanczos procedure to only compute the n_{red} largest singular values.
- Cost of approximating partial derivatives along patterns
This is computationally the most expensive part of the proposed procedure and its CPU time is comparable to the time of $p_{red} + n_{red}$ simulations of the high-order model.

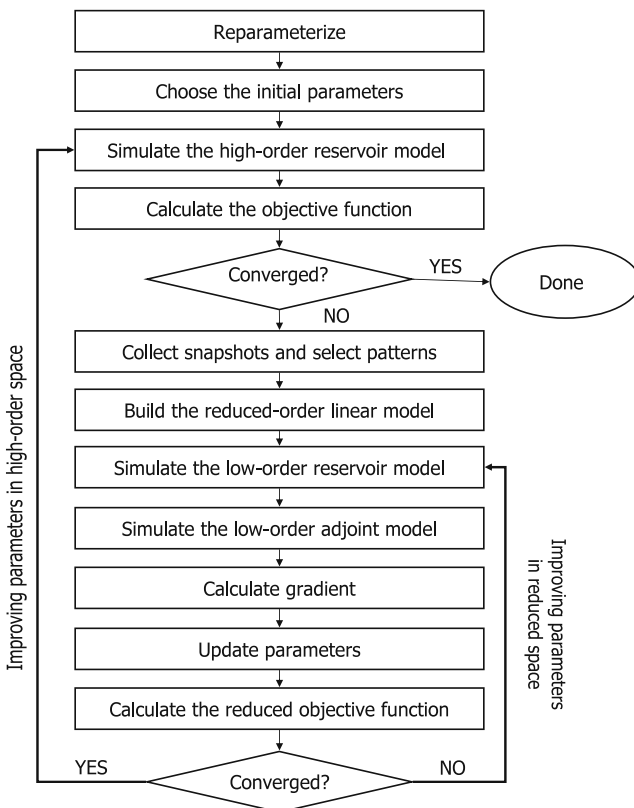


Fig. 1 Model-reduced gradient-based history matching algorithm

- (b) Cost of solving the reduced system and simplified history matching problem
 - Cost of solving the reduced system
The cost of solving a system of n_{red} linear equations is equal to $O((n_{red})^2)$, which can be neglected in our case.
 - Cost of the optimization procedure
In case of a quadratic objective function, the quasi-Newton routine iterates approximately $p_{red} + 1$ times; see [18].

4.8 Additional issues

Pressure and saturation variables have a totally different physical behavior, a different order of numerical values, and generally a different variability in their values. The POD procedure favors variables which values show greater variability; this can have negative consequences on the pattern identification. If all variables are of the same kind, then the data can be preprocessed before the POD procedure by normalizing them to have unit variance. Therefore, we decided to collect snapshots for pressures and saturations separately, following Van Doren et al. [38]. Instead of solving one eigenvalue problem to obtain the POD matrix Φ , we solve two separate eigenvalue problems (one for the

matrix of pressure snapshots $\mathbf{X}_p \in \mathbb{R}^{\frac{n}{2} \times s}$ and one for the matrix of saturation snapshots $\mathbf{X}_s \in \mathbb{R}^{\frac{n}{2} \times s}$). It yields two different POD matrices, $\Phi_p \in \mathbb{R}^{\frac{n}{2} \times n_p}$ and $\Phi_s \in \mathbb{R}^{\frac{n}{2} \times n_s}$ for pressure, and saturation, respectively, and the final POD matrix Φ is formed as

$$\Phi = \begin{bmatrix} \Phi_p & \mathbf{0}_{\frac{n}{2} \times n_s} \\ \mathbf{0}_{\frac{n}{2} \times n_p} & \Phi_s \end{bmatrix} \in \mathbb{R}^{n \times n_{\text{red}}},$$

where n_p and n_s are number of patterns for pressure and saturation, respectively, and $n_{\text{red}} = n_p + n_s$.

5 Numerical experiments

5.1 Model settings

We performed a numerical “twin experiment” in which we used the model-reduced gradient-based approach to estimate the uncertain permeability field of a 3D reservoir model from synthetic noisy production data. The model, originally introduced by Van Essen et al. [39], describes iso-thermal slightly compressible two-phase (oil–water) flow, in a channelized reservoir with eight injection and four production wells; see Fig. 2. The model is based on the assumptions that the effects of capillary pressure can be neglected, that the permeability is isotropic, and that the porosity, permeability, viscosity, and density are pressure independent.

– Reservoir geometry

The reservoir model is a part of a quadrilateral of the size $480 \times 480 \times 28$ m, divided into $60 \times 60 \times 7$ uniform Cartesian grid blocks, of which there are 18,553 active, forming an egg-shaped domain with no-flow boundaries.

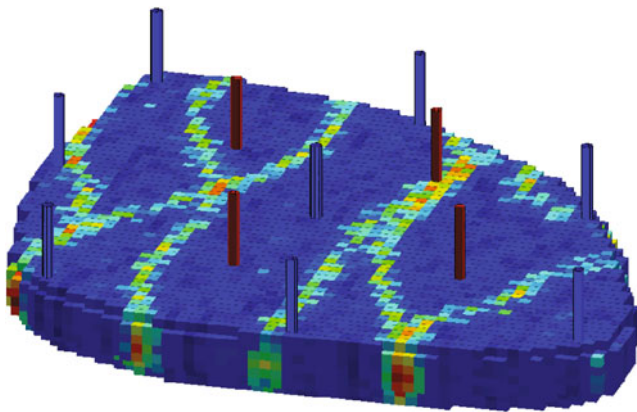


Fig. 2 The “true” reservoir permeability field with well locations. Producers are indicated in red, injectors in blue

– Reservoir properties

The porosity is assumed to be uniform and equals 0.2. The “true” permeability field on a log scale is shown in Fig. 3. The permeability contrast between the channels and the background fill is relatively mild (about a factor of 10). The rock compressibility $c_r = 0 \times 1/\text{Pa}$.

– Initial conditions

The initial reservoir pressures is 40 MPa and the initial water saturation is taken as connate water saturation $S_{wc} = 0.2$.

– Well locations and constraints

Water is injected at a rate of $32 \text{ m}^3/\text{day}$ in the eight vertical injectors, while the four vertical producers are operated at constant bottom-hole pressures of 39.5 MPa. All wells are perforated at each layer of the reservoir.

– Fluid properties

The oil and water viscosities are identical: $\mu_o = \mu_w = 10^{-3} \text{ Pa s}$ and so are the compressibilities: $c_w = c_o = 10^{-10} 1/\text{Pa}$. The densities of oil and water are given by $\rho_o = 900 \text{ kg/m}^3$ and $\rho_w = 1,000 \text{ kg/m}^3$.

– Relative permeabilities

We assumed Corey-type relative permeabilities with Corey exponents for oil and water: $n_o = 4$ and $n_w = 3$. The residual oil saturation was taken as $S_{or} = 0.15$ and the connate water saturation as $S_{wc} = 0.2$.

5.2 History matching settings

The assimilation period was chosen to be 3 years during which observations were taken from four

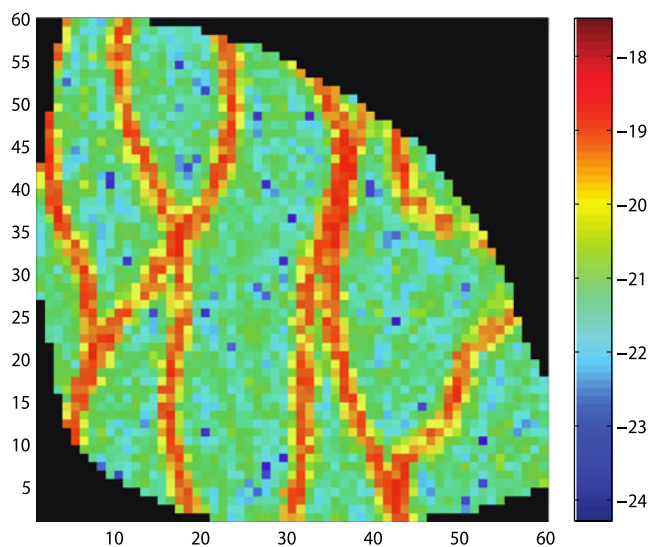


Fig. 3 Log of true permeability field in the fourth layer [m^2]

production and eight injection wells every 60 days, resulting in 18 time instances. During this assimilation period, water breakthrough occurred in one well, after about 500 days. Noisy observations were generated from the model with the “true” permeability field and consisted of bottom-hole pressures in the injectors and water and oil rates in the producers, with a random normally distributed noise of 5% of their values. It resulted in 72 oil rates and five water rates measured in the producers and 144 bottom-hole pressures measured in the injectors, which gives in total 221 data points. In the assimilation procedure, the errors statistics were assumed to be known, i.e., errors were represented by a normal distribution with zero mean and a standard deviation equal to 5% of the observed data. Moreover, it was assumed, correctly, that distinct observations were affected by physically independent noise, resulting in a diagonal covariance matrix for the observation errors. The ensemble of reservoir models was created using in-house geostatistical software with an anisotropic variogram. We used an ensemble of 100 realizations which are Gaussian distributed random fields in each layer, conditioned to the well data corrupted with an error in the order of 10^{-4} of the true value. All realizations were conditioned to the same hard data. Each layer had a randomly oriented ellipsoidal covariance with randomly chosen correlation length between four to eight grid blocks for each layer at each realization. No vertical correlation was assumed. An ensemble of 100 permeability fields was used to create the background permeability θ_{init}^b , taken as the ensemble average, and to estimate the background error covariance matrix. The permeability field was reparameterized using a KL-expansion described above, resulting in 22 permeability patterns with 22 corresponding parameters. Such a small number of parameters to represent the 18,553 grid block permeability values is motivated by the fact that the available data are very sparse, which implies that the identifiability of the parameter space is very low [46].

5.3 Reduced model settings

Additional numerical parameter settings are required for the model-reduced approach, in particular, the

Table 1 Experiment 1: model-reduced gradient-based history matching

Outer iter	J	n_{red}	p_{red}	n_{sim}
0	338	—	22	1
1	113	65	22	98(= 11 + 54 + 11 + 22)

p_{red} the number of parameters, n_{red} the number of patterns, n_{sim} the number of simulations

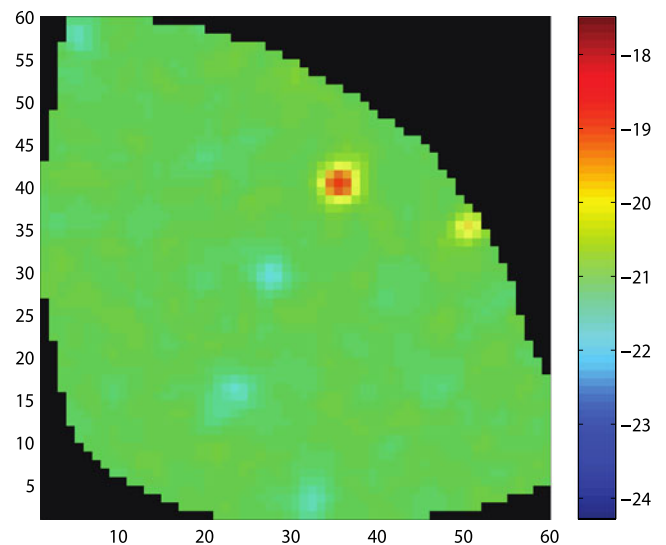


Fig. 4 Experiment 1. Log of prior permeability field in the fourth layer [m²]

number of snapshots and the number of patterns used to build the reduced-order model. In order to cover the dynamic behavior of the system, snapshots should be collected during all the assimilation times for different ranges of the permeability fields. However, to get a rough idea of the direction of the parameter update, we started with a lower number of snapshots in the first outer iteration (reducing also the dimension of the eigenvalue problem) and we increased it in the successive ones. We selected a snapshot every 60 days from 11 simulations, where each simulation was based on a different permeability field created by perturbing two different parameters per simulation. We

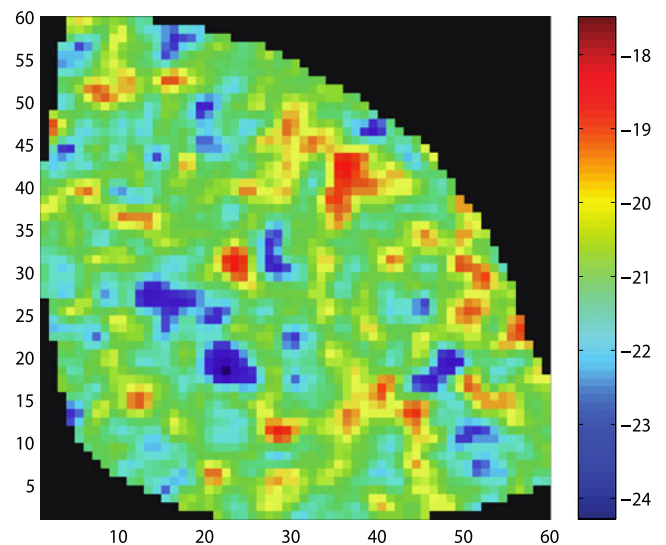


Fig. 5 Experiment 1. Log of permeability field from the model-reduced approach in the fourth layer [m²]

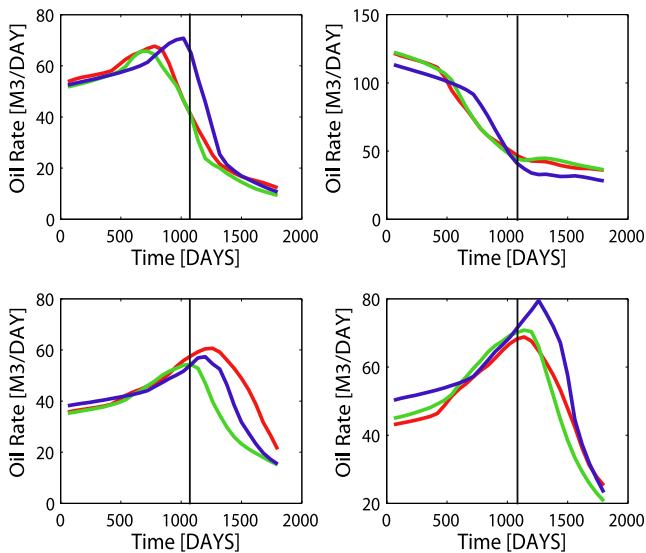


Fig. 6 Experiment 1. Oil rate in the production wells, obtained with the prior permeability field (*blue*), the true permeability field (*red*), and the estimated permeability field (*green*) using the model-reduced approach. The *black vertical lines* indicate the end of the assimilation period

collected snapshots for pressures and saturations separately, which resulted in 200 snapshots for pressures and 200 snapshots for saturations, and two eigenvalue problems to be solved. This separation allowed us to choose different relative importance levels for pressure and for saturation, which varied over the iterations. We started with a lower relative importance in the first outer iteration and increased it in the successive ones. The relative importance for the pressures was initiated at 85% for the first iteration and was set to increase 5% in each successive iteration up to 95%. Because the saturation field displays a much larger spatial variability than the pressure field, it results in a higher number of patterns for the same level of relative importance. We fixed the relative importance for saturation to 80% in the first iteration, increasing it in the successive iterations up to 90%, in steps of 5%.

Table 2 Experiment 1: model-reduced gradient-based history matching with fewer patterns

Outer iter	J	n_{red}	p_{red}	n_{sim}
0	338	—	22	1
1	114	35	22	68(= 11 + 29 + 6 + 22)

p_{red} the number of parameters, n_{red} the number of patterns, n_{sim} the number of simulations

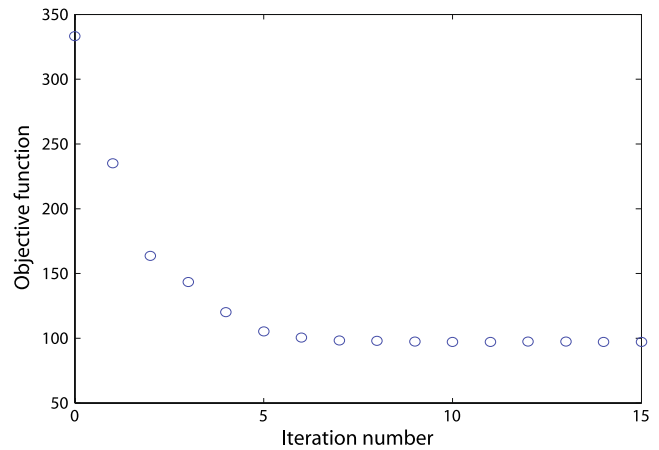


Fig. 7 Experiment 1. Successive iterations of the objective function value

5.4 History matching results

5.4.1 Experiment 1

It was noticed in [21, 38] and [7] that the pressure behavior can be well represented by a few patterns but that the situation is worse for the saturation behavior, which is caused by the moving fluid interface. A significantly larger number of patterns is therefore required to satisfy the chosen accuracy level. We selected 54 + 11 patterns for the first iteration (54 for saturation and 11 for pressure), which means that the reduced-order model operates in dimension 65 + 22. The results of the first data assimilation experiment are summarized in Table 1.

The initial value of the objective function was 338 and resulted from a simulation with the prior parameters, as depicted in Fig. 4. The first outer iteration terminated with parameters for which the original objective function decreased to 113. The stopping criterion for the outer loop (Eq. 29) was satisfied ($179 \leq 2J(\theta_1^b) \leq 263$), and the procedure terminated. The computational cost of the model-reduced approach was about equal to the time required to simulate 98 high-order model evaluations. More precisely, 98 is a summation of 11 simulations during which snapshots are collected, 54 + 11 simulations required to approximate the partial derivatives along saturation and pressure patterns,

Table 3 Experiment 1: classical adjoint-based history matching

Iterations	J	p_{red}	n_{sim}
0	338	22	1
15	97	22	79(= 15 × 2 + 49)

p_{red} the number of parameters, n_{sim} the number of simulations

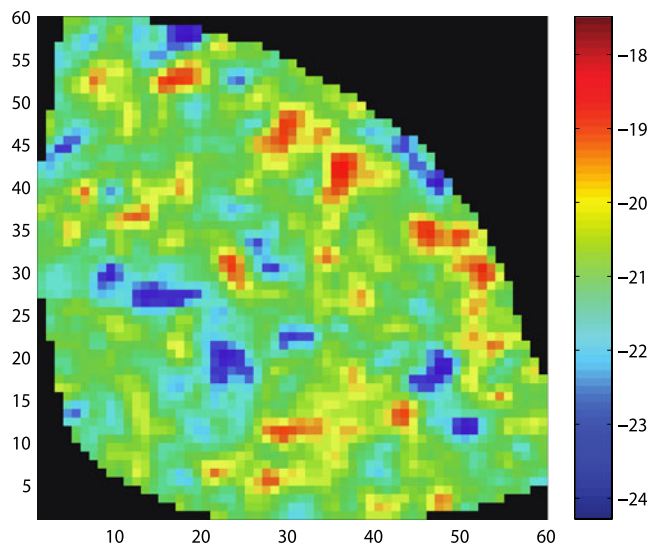


Fig. 8 Experiment 1. Log of permeability field from the adjoint-based approach in the fourth layer [m²]

and 22 simulations required to approximate the partial derivatives with respect to the parameters.

The final estimate is presented in Fig. 5. From a geological point of view, the obtained estimate does not represent the true reservoir (see Figs. 3 and 5). Clearly, the information coming from sparse well data is not sufficient to capture the complicated geological structure, a well-known issue in petroleum reservoir history matching. However, if we consider the match

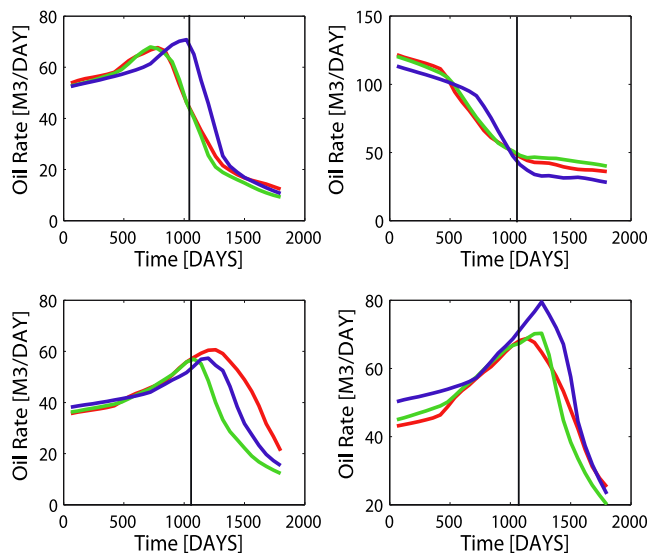


Fig. 9 Experiment 1. Oil rate in the production wells, obtained with the prior permeability field (blue), the true permeability field (red), and the estimated permeability field (green) using the adjoint-based approach. The black vertical lines indicate the end of the assimilation period

Table 4 Experiment 1: finite-difference based history matching

Iterations	J	p_{red}	n_{sim}
0	338	22	1
6	111	22	141 (= 6 × 23 + 3)

p_{red} the number of parameters, n_{sim} the number of simulations

between observed data and predictions in Fig. 6, we see that the history matching exercise resulted in an improved match for past data in all four producers and improved predictions in three out of the four producers. In an attempt to further reduce the computing time, we repeated the model-reduced history matching exercise with an even smaller number of patterns to build the reduced-order model. We obtained similar results in approximately 68 iterations; see Table 2 for further details. Next, we compared the results with those of the classical approach using the adjoint of the tangent linear model. Because the convergence of any minimization problem strongly depends on the number of estimated parameters, we used the same reparameterization for the adjoint-based case as for the model-reduced case. It was stopped according to the criterion defined by Eqs. 27 and 28. The adjoint-based history matching exercise converged after 15 iterations (see Fig. 7 and Table 3). The minimization was performed using a quasi-Newton optimization where the Hessian of the objective function was updated using the BFGS method; see [18]. It required 15 evaluations of the adjoint model and 49 evaluations of the forward model (additional evaluations of forward model were caused by the line-search procedure) and obtained a

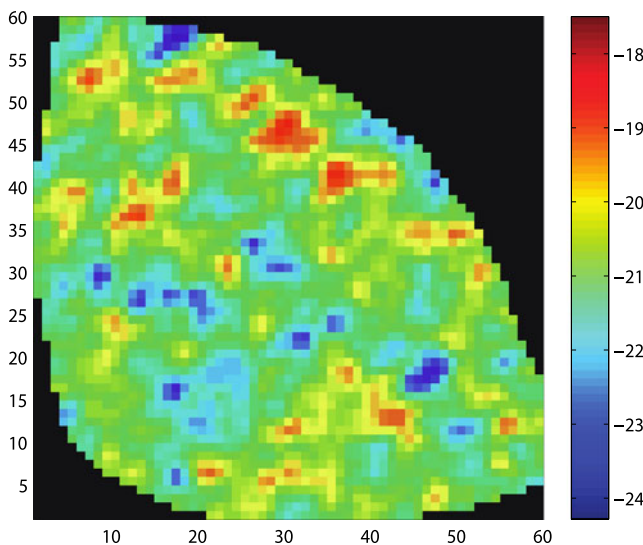


Fig. 10 Experiment 1. Log of permeability field from the finite-difference-based approach in the fourth layer [m²]

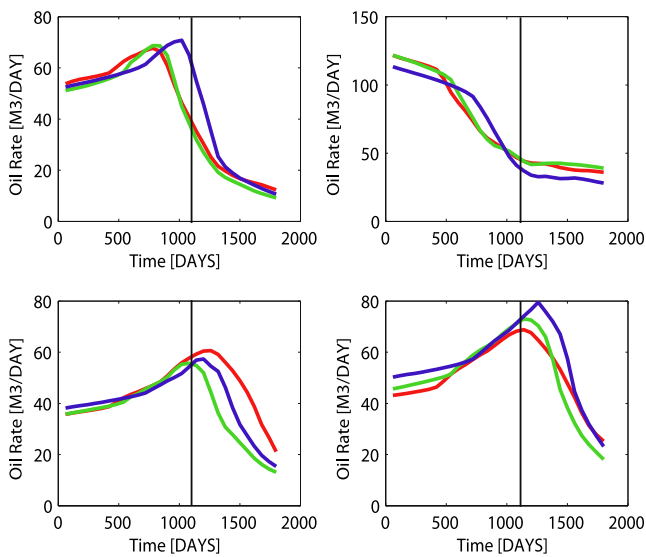


Fig. 11 Experiment 1. Oil rate in the production wells, obtained with the prior permeability field (*blue*), the true permeability field (*red*), and the estimated permeability field (*green*) using the finite-difference-based approach. The *black vertical lines* indicate the end of the assimilation period

minimal value of the objective function of 97. The estimated parameters and the liquid rates are presented in Figs. 8 and 9, respectively. The results closely resemble those of the model-reduced approach. Both methods converged to solutions that give good matches for all producers and good predictions for three out of four producers. In both cases, the difficulty in the prediction is related to the same well, located in the region that is swept least, and for which the observed production

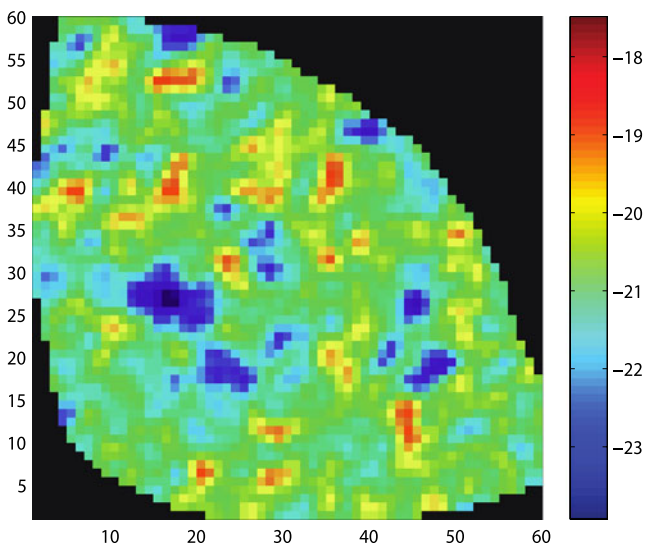


Fig. 12 Experiment 1. Log of permeability field from the model-reduced approach with fewer patterns in the fourth layer [m^2]

Table 5 Experiment 2: model-reduced gradient-based history matching

Outer iter	J	n_{red}	p_{red}	n_{sim}
0	838	—	22	1
1	100	51	22	84(= 11 + 47 + 4 + 22)

p_{red} the number of parameters, n_{red} the number of patterns, n_{sim} the number of simulations

rates are the lowest. Summarizing, for the example considered, the approximate derivatives as obtained from the model-reduced approach resulted in a similar behavior of the optimizer as the “exact” derivatives obtained from the adjoint-based approach.

In order to compare the computational effort of the model-reduced approach, we performed an additional computation using a finite-difference approach. We iterated the optimization until we obtained a value of the objective function comparable to the result of the model-reduced approach. Table 4, Figs. 10, and 11 represent those results. As expected, the numerical performance of the adjoint-based approach is better than that of the model-reduced approach. In this example, we required 79 and 98 simulations for each of the methods, respectively. In fact, we could have used less iterations in the adjoint-based case to achieve a similar reduction in the objective function as for the model-reduced case: See Fig. 7 which indicates that we could have used just seven iterations. In that case, the adjoint-based approach would have required only about seven forward, seven backward, and an estimated five-line search iterations bringing the total to 19, i.e., the adjoint-based approach would then be about five

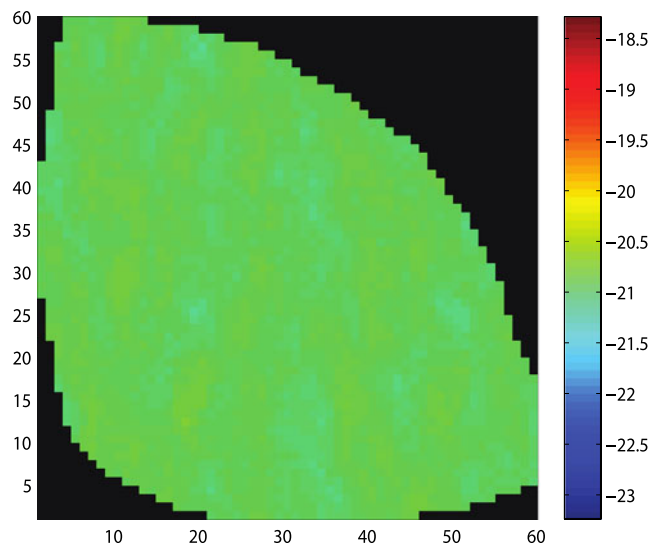


Fig. 13 Experiment 2. Log of prior permeability field in the fourth layer [m^2]

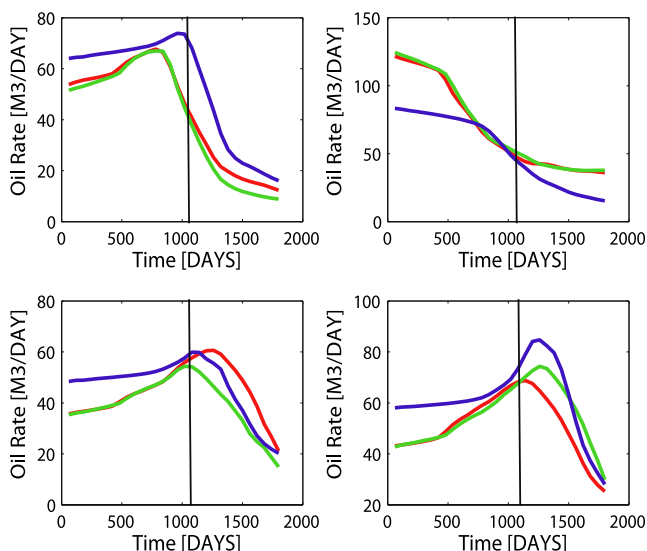


Fig. 14 Experiment 2. Oil rate in the production wells, obtained with the prior permeability field (blue), the true permeability field (red), and the estimated permeability field (green) using the model-reduced approach. The black vertical lines indicate the end of the assimilation period

times as efficient as the model-reduced approach. At the other hand, there is also scope to improve the efficiency of the model-reduced approach: As indicated by Table 2, an almost identical objective function value could have been obtained with 35(= 29 + 6) patterns, which reduces the number of iterations from 98 to 68. The obtained estimate is presented in Fig. 12.

More importantly, the model-reduced approach is more effective than a finite-difference approach, even when using the reparameterization as applied in our

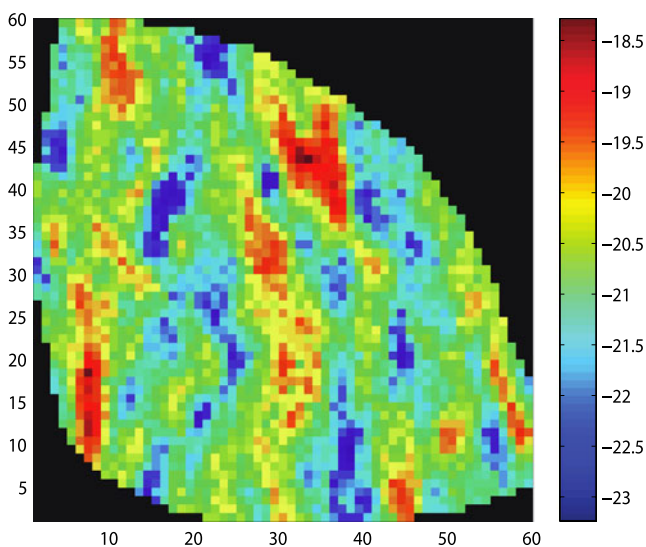


Fig. 15 Experiment 2. Log of permeability field from the model-reduced approach in the fourth layer [m²]

Table 6 Experiment 2: classical adjoint-based history matching

Iterations	J	p_{red}	n_{sim}
0	838	22	1
13	92	22	59(= 13 × 2 + 33)

p_{red} the number of parameters, n_{sim} the number of simulations

example. A finite-difference approach required six iterations and one additional function evaluation for the line search routine, which gives $6 \times (22 + 1) + 3 = 141$ simulations which makes the model-reduced approach with 35 patterns more than twice as efficient as a finite-difference approach.

5.4.2 Experiment 2

In the second experiment, we repeated the history matching procedure with a different initial guess for the permeability field. All other settings for model reduction and history matching stayed the same. For this case, we assumed, following Van Essen et al. [39], that the main direction of the channels was known, e.g., from seismic measurements, but that no specific knowledge about the channel configuration was available. The set of 100 realizations of the reservoir were sketched by hand based on the geological insight, with a strong vertical correlation, and each realization displayed an alternative channel configuration. The generated ensemble of permeability fields was used to create the background permeability θ_{init}^b , taken as the ensemble average, and to estimate the background error covariance matrix. The results of this experiment are summarized in Table 5. We selected 51 (= 47 + 4) patterns for the first outer iteration (47 for saturation and four for pressure), which means that the reduced model operates in dimension 51 + 22. We observed again that the pressure behavior is well represented by a few patterns, only four, but that the saturation requires significantly more patterns, namely 47. The initial value of the objective function was 838, and it resulted from the prior parameter field presented in Fig. 13. The first outer iteration terminated with parameters for which the objective function decreased to 100. The stopping criterion for the outer loop (Eq. 29) was satisfied ($179 \leq 2J(\theta_1^b) \leq 263$), and the procedure was terminated. In this example, the computational cost

Table 7 Experiment 2: finite-difference-based history matching

Iterations	J	p_{red}	n_{sim}
0	838	–	22
11	101	–	254(= 11 × 23 + 1)

p_{red} the number of parameters, n_{sim} the number of simulations

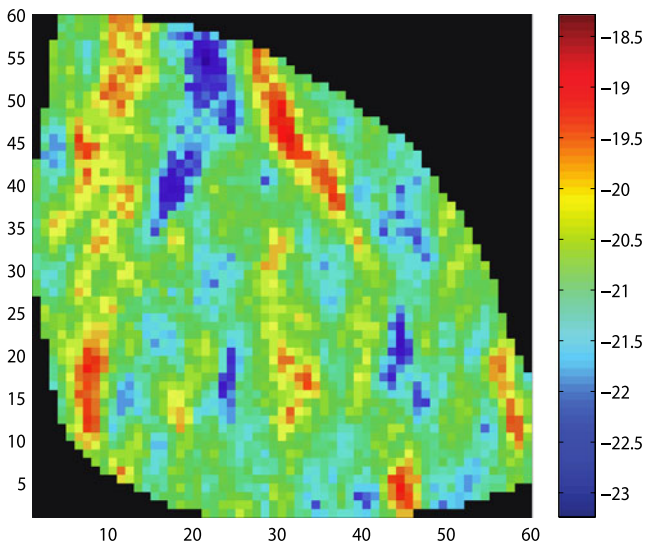


Fig. 16 Experiment 2. Log of permeability field from the adjoint-based approach in the fourth layer [m²]

of the model-reduced approach was about the time required to simulate 84 high-order model evaluations (11 simulations during which snapshots are collected, about 47 + 4 simulations required to approximate the partial derivatives along saturation and pressure patterns and 22 simulations required to approximate the partial derivatives with respect to parameters). In Fig. 14, the comparison of oil production curves resulting from different permeability fields is depicted. The estimate is able to reproduce the past data from all four producers, while the prediction is nearly perfect for two wells (top

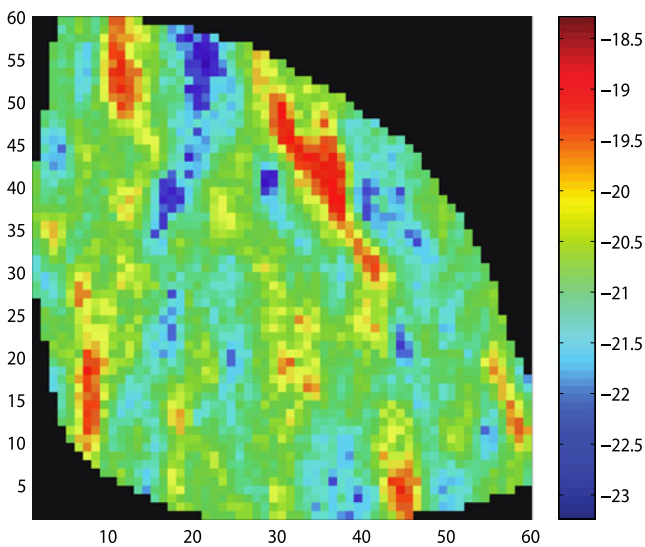


Fig. 17 Experiment 2. Log of permeability field from the finite-difference-based approach in the fourth layer [m²]

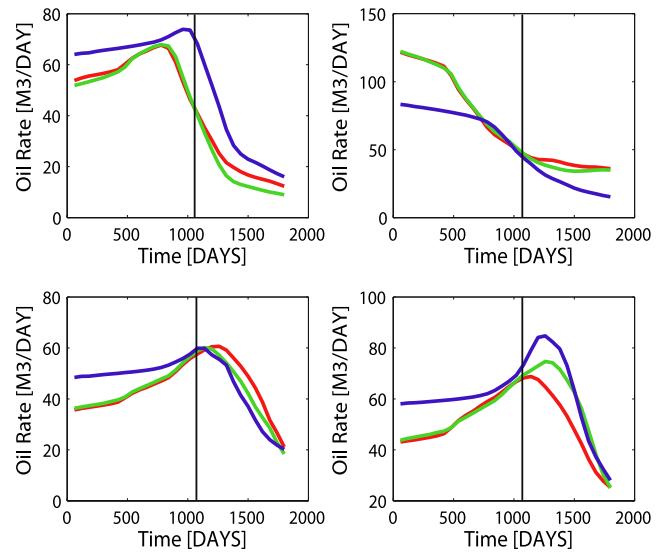


Fig. 18 Experiment 2. Oil rate in the production wells, obtained with the prior permeability field (*blue*), the true permeability field (*red*), and the estimated permeability field (*green*) using the adjoint-based approach. The *black vertical lines* indicate the end of the assimilation period

left and top right plots) and improved for one well (bottom right plot). Similar to the first experiment, the worst predictions were obtained for the well depicted on the bottom left plot. The final estimated parameters are presented in Fig. 15. Also for this second example, we checked the performance of the model-reduced

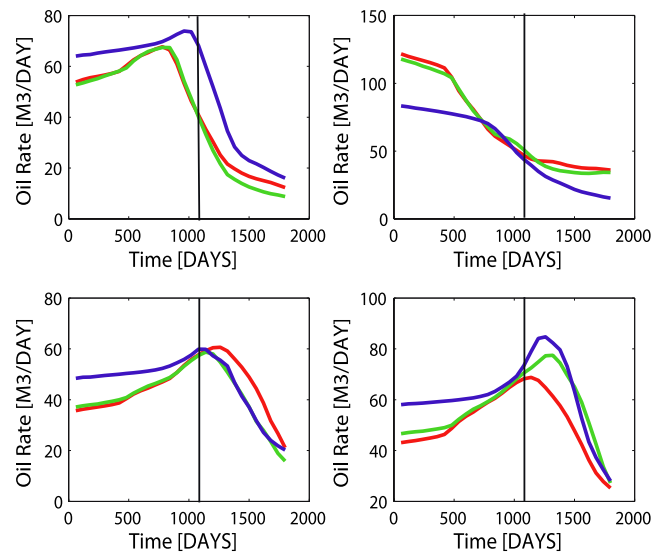


Fig. 19 Experiment 2. Oil rate in the production wells, obtained with the prior permeability field (*blue*), the true permeability field (*red*), the estimated permeability field (*green*) using the finite-difference-based approach. The *black vertical lines* indicate the end of the assimilation period

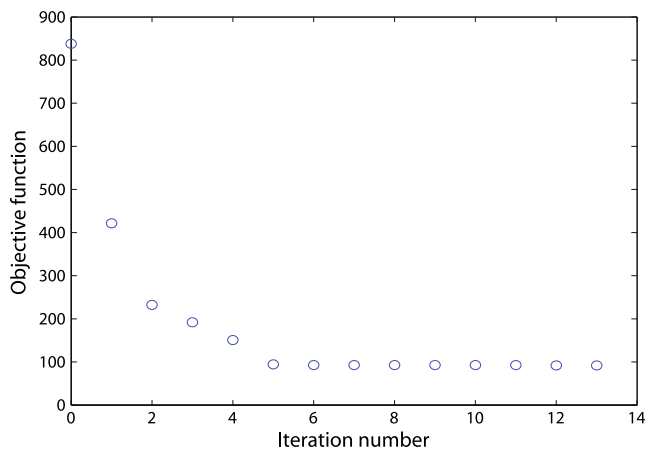


Fig. 20 Experiment 2. Successive iterations of the objective function value

approach in comparison to the classical adjoint-based approach and a finite-difference approach (see Tables 5, 6, and 7), respectively. Comparison of Figs. 15, 16, and 17 and comparison of Figs. 14, 18, and 19, respectively, show that in the second example the results of the model-reduced approach and of a finite-difference-based approach come very close to those of the adjoint-based approach.

The adjoint case converged after 13 iterations (see Fig. 20). It required 13 evaluations of the forward model, another 13 of the adjoint model, and 33 line searches, bringing the total to 59 iterations to reach the minimal objective function value of 92. Just like in example one, we could have used fewer iterations, five in this case, to obtain a similar value of the objective function, which would have brought the total number of iterations for the adjoint-based case to about 15. In other words, for this second example, the adjoint-based approach is up to about six times as efficient as the model-reduced approach. However, also in this case, there is scope to improve the efficiency of the model reduction by further reducing the number of patterns. Moreover, the model-reduced approach is now about three times more efficient than the finite-difference approach.

6 Conclusion

We presented a model-reduced gradient-based history matching procedure based on the POD method. The reduction technique identifies a low-dimensional subspace of the state space, in which the dynamics of the reservoir state are well represented. Subsequently, linearized reservoir equations are constructed and re-

peatedly solved in a low-dimensional space during the optimization procedure to minimize the mismatch between observed and simulated output.

The method was used to adapt the permeability field in a two-phase (oil–water) reservoir simulation model containing 18,553 grid blocks. The results from two numerical experiments with different prior permeability fields showed that the model-reduced approach gives results comparable to those of a classical adjoint-based history matching procedure. In the examples considered, the adjoint-based approach was about five to six times as efficient as the model-reduced approach, but the latter was about two to three times as efficient as a finite-difference approach. Moreover, in case of a larger number of wells, allowing for the estimation of a larger number of (reparameterized) model parameters, it is expected that the efficiency compared to a finite-difference approach improves.

The classical gradient-based history matching procedure is a very efficient one, but for complex models, it is very difficult to implement. The model-reduced approach is computationally more expensive, but it does not require access to the simulation code and is easy to implement. The classical adjoint-based approach gives the exact gradients of the objective function with respect to the parameters. The model-reduced history matching procedure uses a simplified reduced-order forward model and a corresponding reduced-order adjoint model, and therefore, the calculated gradients are not exact, but they proved to be accurate enough to decrease the objective function in all iteration steps.

Acknowledgements The authors acknowledge financial support from the Integrated Systems Approach to Petroleum Production (ISAPP) program, which is jointly sponsored by Shell International E&P, TNO, and Delft University of Technology.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Aanonsen, S.I., Naevdal, G., Oliver, D.S., Reynolds, A.C., Valles, B.: The ensemble Kalman filter in reservoir engineering—a review. *SPE J.* **14**(3), 393–412 (2009)
2. Altaf, M.U., Heemink, A.W., Verlaan, M.: Inverse shallow-water flow modeling using model reduction. *Int. J. Multiscale Comp. Eng.* **7**(6), 577–594 (2009)
3. Asheim, H.: Maximization of water sweep efficiency by controlling production and injection rates. In: Paper SPE 18365 Presented at the SPE European Petroleum Conference, London, UK (1988)

4. Aziz, K., Settari, A.: Petroleum Reservoir Simulation. Applied Science, London (1979)
5. Brouwer, D.R., Jansen, J.D.: Dynamic optimization of water flooding with smart wells using optimal control theory. *SPE J.* **9**(4), 391–402 (2004)
6. Cao, Y., Zhu, J., Luo, Z., Navon, I.M.: Reduced order modeling of the upper tropical Pacific ocean model using proper orthogonal decomposition. *Comput. Math. Appl.* **52**(8–9), 1373–1386 (2006)
7. Cardoso, M.A., Durlafsky, L., Sarma, P.: Development and application of reduced-order modeling procedures for subsurface flow simulation. *Int. J. Numer. Methods Eng.* **77**(9), 1322–1350 (2009)
8. Chavent, G., Dupuy, M., Lemonnier, P.: History matching by use of optimal theory. *SPE J.* **15**(1), 74–86 (1975)
9. Chen, W.H., Gavalas, G.R., Seinfeld, J.H., Wasserman, M.L.: A new algorithm for automatic history matching. *SPE J.* **14**(4), 593–608 (1974)
10. Chen, Y., Oliver, D.S., Zhang, D.: Efficient ensemble-based closed-loop production optimization. *SPE J.* **14**(4), 634–645 (2009)
11. Courant, R., Hilbert, D.: *Methods of Mathematical Physics*. Wiley Interscience, USA (1953)
12. Courtier, P., Thépaut, J.-N., Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**, 1367–1387 (1994)
13. Dadashpour, M., Echeverria Ciaurri, D., Mukerji, T., Kleppe, J., Landrø, M.: Simple zonation and principle component analysis for speeding up porosity and permeability estimation from 4D seismic and production data. In: *Proceeding of 71st EAGE Conference and Exhibition, Amsterdam, The Netherlands* (2009)
14. Dennis, J.E., Schnabel, R.B.: *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs (1983)
15. Doublet, D.C., Aanonsen, S.I., Tai, X.-C.: An efficient method for smart well production optimisation. *J. Pet. Sci. Eng.* **69**(1–2), 25–39 (2009)
16. Evensen, G.: *Data Assimilation—the Ensemble Kalman Filter*, 2nd edn. Springer, Berlin (2009)
17. Fang, F., Pain, C.C., Navon, I.M., Piggot, M.D., Gorman, G.J., Farrell, P.E., Allison, P.A., Goddard, J.H.: A POD reduced-order 4D-Var adaptive mesh ocean modeling approach. *Int. J. Numer. Methods Fluids* **60**(7), 709–732 (2008)
18. Fletcher, R.: *Practical Methods of Optimization*. Wiley-Interscience, Padstow (1991)
19. Gao, G., Li, G., Reynolds, A.C.: A stochastic optimization algorithm for automatic history matching. *SPE J.* **12**(2), 196–208 (2007)
20. Gavalas, G.R., Shah, P.C., Seinfeld, J.H.: Reservoir history matching by Bayesian estimation. *SPE J.* **16**(6), 337–350 (1976)
21. Heijn, T., Markovinic, R., Jansen, J.D.: Generation of low-order reservoir models using system-theoretical concepts. *SPE J.* **9**(2), 202–218 (2004)
22. Jafarpour, B., McLaughlin, D.B.: Reservoir characterization with the discrete cosine parameterization. Part I, parameterization. *SPE J.* **14**(1), 182–188 (2009)
23. Jahns, H.O.: A rapid method for obtaining a two-dimensional reservoir description from well pressure response data. *SPE J.* **6**(12), 315–327 (1966)
24. Karhunen, K.: Zur spektral theorie stochastischer prozesse. *Ann. Acad. Sci. Fenn. A1* **34**, 1–7 (1946)
25. Lee, T.Y., Seinfeld, J.H.: Estimation of two-phase petroleum reservoir properties by regularization. *J. Comput. Phys.* **69**, 397–419 (1987)
26. Li, R., Reynolds, A.C., Oliver, D.S.: History matching of three-phase flow production data. *SPE J.* **8**(4), 328–340 (2003)
27. Loève, M.: Fonctions aleatoires de second ordre. *Rev. Sci.* **84**, 195–206 (1946)
28. Lorentzen, R.J., Berg, A.M., Naevald, G., Vefring, E.H.: A new approach for dynamic optimization of waterflooding problems. In: *Paper SPE 99690 Presented at the SPE Intelligent Energy Conference and Exhibition, Amsterdam, The Netherlands* (2006)
29. Lumley, J.L.: The structure of inhomogeneous turbulence flows. In: Yaglom, A.M., Tatarsky, V.I. (eds.) *Atmospheric Turbulence and Radio Wave Propagation*, pp. 166–178. Nauka, Moskou (1967)
30. Markovinic, R., Jansen, J.D.: Accelerating iterative solution methods using reduced-order models as solution predictors. *Int. J. Numer. Methods Eng.* **68**(5), 525–541 (2006)
31. Oliver, D.S., Reynolds, A.C., Liu, N.: *Inverse theory for petroleum reservoir characterization and history matching*. Cambridge University Press, Cambridge (2008)
32. Ramirez, W.F.: *Application of Optimal Control Theory to Enhanced Oil Recovery*. Elsevier, Amsterdam (1987)
33. Sahni, I., Horne, R.N.: Multiresolution wavelet analysis for improved reservoir description. *SPE Reserv. Evalu. Eng.* **8**(1), 53–69 (2005)
34. Sarma, P., Chen, W.H., Durlafsky, L.J., Aziz, K.: Production optimization with adjoint models under nonlinear control-state path inequality constraints. *SPE Reserv. Evalu. Eng.* **11**(2), 326–339 (2008)
35. Spall, J.C.: Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Trans. Aerosp. Electron. Syst.* **34**(3), 817–823 (1998)
36. Sudaryanto, B., Yortsos, Y.C.: Optimization of fluid front dynamics in porous media using rate control. *Phys. Fluids* **12**(7), 1656–1670 (2000)
37. Trefethen, L.N., Bau III, D.: *Numerical Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia (1997)
38. Van Doren, J.F.M., Markovinic, R., Jansen, J.D.: Reduced-order optimal control of water flooding using POD. *Comput. Geosci.* **10**(1), 137–158 (2006)
39. Van Essen, G.M., Zandvliet, M.J., Van den Hof, P.M.J., Bosgra, O.H., Jansen, J.D.: Robust waterflooding optimization of multiple geological scenarios. *SPE J.* **14**(1), 202–210 (2009)
40. Vermeulen, P.T.M., Heemink, A.W.: Model-reduced variational data assimilation. *Mon. Weather Rev.* **134**(10), 2888–2899 (2006)
41. Wang, C., Li, G., Reynolds, A.C.: Production optimization in closed-loop reservoir management. *SPE J.* **14**(3), 506–523 (2009)
42. Wasserman, M.L., Emanuel, A.S., Seinfeld, J.H.: Practical applications of optimal-control theory to history-matching multiphase simulator models. *SPE J.* **15**(4), 347–355 (1975)
43. Watson, A.T., Seinfeld, J.H., Gavalas, G.R., Woo, P.T.: History matching in two-phase petroleum reservoirs. *SPE J.* **20**(6), 521–532 (1980)
44. Yang, P., Watson, A.T., Armasu, R.V.: Automatic history matching with variable-metric methods. *SPE Reserv. Evalu. Eng.* **3**(3), 995–1001 (1988)

45. Zandvliet, M.J., Bosgra, O.H., Jansen, J.D., Van den Hof, P.M.J., Kraaijevanger, J.F.B.M.: Bang-bang control and singular arcs in reservoir flooding. *J. Pet. Sci. Eng.* **58**(1 & 2), 186–200 (2007)
46. Zandvliet, M.J., van Doren, J.F.M., Bosgra, O.H., Jansen, J.D., van den Hof, P.M.J.: Controllability, observability and identifiability in single-phase porous media flow. *Comput. Geosci* **12**(4), 605–622 (2008)
47. Zhang, F., Reynolds, A.C.: Optimization algorithms for automatic history matching of production data. In: *Proceeding of the 8th European Conference on the Mathematics of Oil Recovery*, Freiberg, Germany (2002)