

Comprehensive framework for gradient-based optimization in closed-loop reservoir management

Vladislav Bukshtynov · Oleg Volkov · Louis J. Durlafsky · Khalid Aziz

Received: date / Accepted: date

Abstract An efficient, robust and flexible adjoint-based computational framework for performing closed-loop reservoir management is developed and applied. The methodology includes gradient-based production optimization and data assimilation (history matching). Flexibility is achieved through use of automatic differentiation (AD) within the reservoir simulation, production optimization and history matching modules. The use of AD will also facilitate the application of closed-loop reservoir management to physical models of higher complexity. A fast sequential convex programming (SCP) solver based on the method of moving asymptotes (MMA) is applied for the production optimization component of the closed-loop. This technique is shown to outperform the sequential quadratic programming (SQP) method, which is commonly used for production optimization computations. The history matching component of the workflow integrates both production data and proxy seismic measurements into a unified adjoint-based data assimilation framework. The effect of noisy data, and data of different types, on the accuracy of the history matching component is assessed. The overall closed-loop reservoir management methodology is tested using the well-documented Brugge model. Results demonstrate the efficient performance of the

individual closed-loop components and the improvement in net present value that is achieved using these procedures.

Keywords Closed-loop reservoir management · Production optimization · History matching · Seismic data · Noisy data · Gradient-based optimization · Adjoint formulation · Automatic differentiation · PCA-based parameterization

1 Introduction

The reservoir modeling and optimization capabilities that have been developed in recent years hold great promise for improving reservoir performance. Closed-loop reservoir management (CLRM) is one such strategy. CLRM, which has been widely studied [3, 7, 15, 19–21, 27, 37, 44], essentially involves periodic history matching and production optimization computations. Various minimization procedures have been applied for these computations, though our emphasis here will be on gradient-based methods, with gradients computed using adjoints for both history matching and production optimization.

CLRM is illustrated schematically in Figure 1. The production system includes the field and facilities, in which production and seismic data are collected periodically. These data are then used for the model update step, in which the geological model or models are adjusted such that the predicted reservoir response closely matches the actual data. Time-varying optimal well settings (flow rates or bottom-hole pressures) are then determined based on this updated geological model. Hydrocarbon production using these controls proceeds until the next closed-loop step, at which time the geological model is again updated and the optimal controls recomputed. The loop is repeated for the duration of the production time frame. Because the geological model may be highly uncertain at early stages of production, the model

V. Bukshtynov (✉) · O. Volkov · L. J. Durlafsky · K. Aziz
Department of Energy Resources Engineering,
Stanford University, Stanford, CA 94305-2220, USA
E-mail: bukshtu@stanford.edu

O. Volkov
E-mail: ovolkov@stanford.edu

L. J. Durlafsky
E-mail: lou@stanford.edu

K. Aziz
E-mail: aziz@stanford.edu

and controls can change significantly during the course of CLRM.

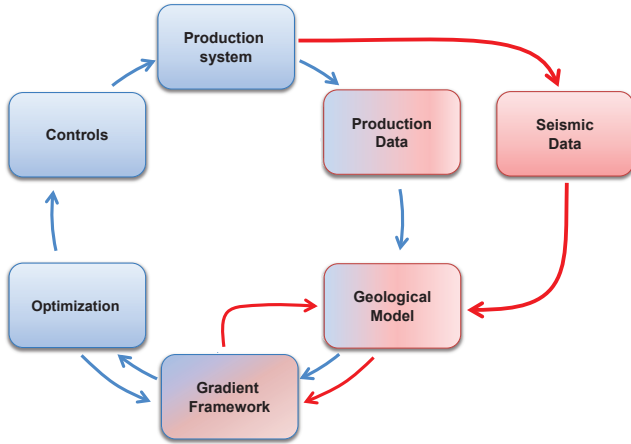


Fig. 1 Closed-loop reservoir management scheme.

The two areas most relevant to CLRM are production optimization and data assimilation. The literature associated with both of these fields is very extensive, so our focus here will be on papers that are directly relevant to this study. A recent review of history matching procedures can be found in [29]. As noted above, our interest here is in gradient-based history matching, discussed in, e.g., [25,30,37]. These methods are very efficient if adjoint procedures are used to construct gradients. In this paper, our adjoint-based history matching scheme utilizes both production and seismic data. Many previous researchers (e.g., [8–10, 12, 18, 31, 38, 41]) have considered the combined use of production and seismic data, but little previous work has involved the application of a general adjoint-gradient procedure for both production and seismic data, as is presented here. We also introduce a simple procedure for generating proxy seismic data with spatial filtering and varying degrees of noise, and the impact of noise on the quality of the history matched solution is assessed.

Gradient-based production optimization has also been addressed in a number of recent papers; e.g., [2,3,24,27,35,37,43,44]. See [20] for a comprehensive description of this general area and an extensive literature review. An important aspect of our work is the choice of an efficient nonlinear programming solver to ensure computational efficiency in the production optimization component of the closed-loop scheme. Specifically, we compare a state-of-the-art SQP solver with an SCP solver based on MMA [39]. The former is commonly used in production optimization [24,37], though the latter does not appear to have been previously applied in this application area.

The automatic differentiation procedure used in this work enables us to obtain adjoint-based gradients for both

production optimization and history matching. The AD technique has been implemented into Stanford’s Automatic Differentiation-based General Purpose Research Simulator, AD-GPRS, using the Automatic Differentiation Expression Templates Library (ADETL). This procedure was originally developed in [45,46] and later extended in [47,48].

This paper proceeds as follows. In the next section we give the mathematical details of the CLRM scheme. The production optimization and history matching components are cast as optimization problems that are solved using an adjoint gradient algorithm. Computational aspects of the methodology, including the optimization software, are described in Section 3. In Section 4 we present a range of computational results for the synthetic Brugge model. These results entail the use of both production and 4D (time-lapse) seismic data. Concluding remarks are provided in Section 5.

2 Mathematical description

2.1 Reservoir simulation model

Although our formulation is quite general, we describe our model here in terms of an isothermal oil-water flow problem. The general set of governing partial differential equations (PDEs) can be expressed as

$$\begin{aligned} \mathbf{g}(\mathbf{x}, \mathbf{u}) &= \mathbf{0}, \\ \mathbf{x}(t_0) &= \mathbf{x}_0, \\ t &\in [t_0, t_f]. \end{aligned} \quad (1)$$

Here t is time, \mathbf{x} denotes the state variables (phase pressure p_p and saturation s_p in this case), and \mathbf{u} represents model parameters (for history matching) or well control parameters (for production optimization). The first equation in (1) is a standard mass balance equation for component/phase p , with $p = o, w$, which designate oil and water (see, e.g., [1])

$$\frac{\partial}{\partial t} (\phi \rho_p s_p) + \text{div}(\rho_p \mathbf{v}_p) + \rho_p q_p = 0, \quad p = o, w. \quad (2)$$

Here the phase flux \mathbf{v}_p is given by Darcy’s law

$$\mathbf{v}_p = -\mathbf{k} \frac{k_{rp}}{\mu_p} (\nabla p_p - \gamma_p \nabla D). \quad (3)$$

The phase dependent parameters appearing in (2)–(3) are molar density ρ_p , relative permeability k_{rp} , viscosity μ_p , vertical pressure gradient $\gamma_p = \rho_p g$, and source term q_p . The space dependent parameters are porosity ϕ , absolute permeability \mathbf{k} , and vertical depth D . The system is closed by the relation $s_o + s_w = 1$ and a capillary pressure relationship (if capillary effects are included; otherwise $p_o = p_w$). We reiterate that, although the above equations are for an oil-water system, the implementation can handle general compositional problems.

In the discretized model, the total volumetric phase flow rate in well j ($q_{j,p}$), and the phase flow rate in completion i in well j ($q_{j,p}^i$), are given by

$$q_{j,p} = \sum_i q_{j,p}^i, \quad j = 1, \dots, N_{\text{well}},$$

$$q_{j,p}^i = \left(T^w \frac{k_{rp}}{\mu_p} \right)_i (p^i - p_j^w - \rho_j \Delta z). \quad (4)$$

Here T^w represents the well transmissibility factor (well index), p^i is the well-block pressure, and $p_j^w + \rho_j \Delta z$ is the wellbore pressure at the corresponding location. The latter involves the well bottom-hole pressure (BHP) p_j^w , which is defined at a reference depth z_0 , along with the adjustment $\rho_j \Delta z$, which accounts for the difference in wellbore pressure between the reference and cell-center depth (here ρ_j is the average density of the fluid in the wellbore). Note that, for simplicity, frictional and other pressure losses in the wellbore are neglected.

2.2 Optimization problems

CLRM entails solving two types of optimization problems sequentially. We defer the detailed description of these problems to Section 2.8, and the adjoint gradient solution algorithm to Sections 2.4 and 2.5. Here, we consider an optimization problem in a general form, with the objective function defined as

$$\mathcal{J}(\mathbf{x}, \mathbf{u}) = \int_{t_0}^{t_f} \mathcal{F}(t, \mathbf{x}, \mathbf{u}) dt. \quad (5)$$

Here $\mathcal{F}(t, \mathbf{x}, \mathbf{u})$ is a nonlinear function whose form, for production optimization and history matching problems respectively, will be specified in Sections 2.8.1 and 2.8.2. The optimal solution $\hat{\mathbf{u}}$ is obtained through minimization or maximization of the objective function $\mathcal{J}(\mathbf{x}, \mathbf{u})$ as follows:

$$\text{history matching:} \quad \hat{\mathbf{u}} = \underset{\mathbf{u}}{\operatorname{argmin}} \mathcal{J}(\mathbf{x}, \mathbf{u}), \quad (6a)$$

$$\text{production optimization:} \quad \hat{\mathbf{u}} = \underset{\mathbf{u}}{\operatorname{argmax}} \mathcal{J}(\mathbf{x}, \mathbf{u}), \quad (6b)$$

subject to the reservoir flow equations (1) and any applicable constraints that define a feasibility region for \mathbf{u} . The details of the constraint treatment are provided in Section 2.7. Note that in the optimization problems (6a) and (6b) the state variables \mathbf{x} are considered to be dependent on \mathbf{u} through an implicit relationship defined by (1). A proper treatment of this relationship is discussed in Section 2.4.

Solutions to the problem (6) in general are located either at an extremum or on the boundary of the feasibility region. We postpone the feasibility discussion until Section 2.7 and consider the algorithm for finding extrema. These points are characterized by the first-order optimality condition, which

requires the Gâteaux differential of the objective function (5), defined as

$$d\mathcal{J}(\mathbf{u}; \delta\mathbf{u}) \triangleq \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{J}(\mathbf{u} + \varepsilon \cdot \delta\mathbf{u}) - \mathcal{J}(\mathbf{u})}{\varepsilon}, \quad (7)$$

to vanish for all perturbations $\delta\mathbf{u}$ [26], i.e.,

$$d\mathcal{J}(\hat{\mathbf{u}}; \delta\mathbf{u}) = 0, \quad \forall \delta\mathbf{u}. \quad (8)$$

Due to its complexity the equality (8) is solved iteratively $\hat{\mathbf{u}} = \lim_{m \rightarrow \infty} \mathbf{u}^{(m)}$, where

$$\begin{cases} \mathbf{u}^{(m+1)} = \mathbf{u}^{(m)} + \tau^{(m)} \delta\mathbf{u}^{(m)}, & m = 1, \dots, \\ \mathbf{u}^{(1)} = \mathbf{u}_0. \end{cases} \quad (9)$$

In (9), \mathbf{u}_0 is the initial guess corresponding to an approximation of \mathbf{u} , and $\tau^{(m)}$ is the length of the step in the search direction at iteration m . In the simplest case, the search direction

$$\delta\mathbf{u}^{(m)} = \underset{\mathbf{v}}{\operatorname{argmin}} d\mathcal{J}(\mathbf{u}^{(m)}; \mathbf{v}) \quad (10)$$

represents the steepest descent direction. However, a more efficient search is achieved through use of advanced minimization/maximization techniques. In this work, we employ two dedicated SCP algorithms, discussed in Section 2.3, to handle the $\tau^{(m)} \delta\mathbf{u}^{(m)}$ term in (9). Note that, since the optimization problem (6) is in general nonconvex, condition (8) characterizes only a local, rather than the global, optimum $\hat{\mathbf{u}}$. However, because SCP solvers accumulate search direction information from a number of iterations, the trust region is larger, and the optimal solution “less” local, than in steepest descent. We now discuss SCP in more detail.

2.3 Sequential convex programming (SCP)

An SCP algorithm finds an optimum $\hat{\mathbf{u}}$ by solving a sequence of convex programming (CP) problems where an objective function is replaced by its convex approximation. The CP problems are solved at each iteration m to determine the update $\tau^{(m)} \delta\mathbf{u}^{(m)}$ in (9) within the feasibility region of the optimization variable space. When the optimization problem is known to have a nonunique solution, the success of SCP relies on the characteristics of the trust region and the ability to define and solve individual CP problems with minimal computational effort.

Commonly used approximations are those obtained with Taylor expansions. Approximations of order higher than two are not used in production optimization and history matching because of their prohibitive computational cost. The linear approximation is not applied directly because, due to the nonlinearity of these problems, it leads to a small trust region. However, it often serves as a building block for other

convex approximations that depend on the exact sensitivity of the objective function with respect to the optimization variables. In the next section we provide a detailed description of the adjoint gradients, which provide the most accurate estimation of this sensitivity.

The quadratic expansion within the sequential quadratic programming algorithm has proven to be highly efficient for production optimization and history matching [24, 37]. The computational efficiency of SQP is achieved by using the adjoint gradients to construct both linear and quadratic terms in the approximation, along with a dedicated optimization algorithm to solve a quadratic problem. The quadratic term accumulates the gradient information from multiple optimization iterations. Therefore, the quadratic approximation can have an arbitrarily large trust region and can thus avoid some local minima. Depending on the form of the objective function and the initial guess, this capability could have positive or negative consequences. A common observation is that SQP, though quite general, requires some amount of tuning for a particular optimization problem.

Dedicated SCP algorithms that use convex approximations consistent with the particular nonlinear form of the objective function have been developed in many application areas. Among these, the reconstruction of the geometrical scaling factors in structural optimization resembles the mathematical formulation of the production optimization problem with BHP controls. This similarity is due to the separability and quasi-linearity of the objective function and PDE constraints with respect to individual optimization variables (i.e., partial Hessians $\mathcal{J}_{\mathbf{uu}}$ and $g_{\mathbf{uu}}$ are diagonal or zero in both problems). In this situation, the use of convex approximations based on the linearization of functions with respect to reciprocal optimization variables enhances the convergence properties of the optimization procedure [14].

A successful modification of this approach, called method of moving asymptotes (MMA) [39], uses an approximation of \mathcal{J} in the vicinity of $\bar{\mathbf{u}}$ in the form

$$\begin{aligned} \tilde{\mathcal{J}}(\mathbf{u}) = \mathcal{J}(\bar{\mathbf{u}}) + \sum_{i \in I_+} \nabla \mathcal{J}_i \left[\frac{(U_i - \bar{u}_i)^2}{U_i - u_i} - (U_i - \bar{u}_i) \right] \\ - \sum_{i \in I_-} \nabla \mathcal{J}_i \left[\frac{(\bar{u}_i - L_i)^2}{(u_i - L_i)} - (\bar{u}_i - L_i) \right], \end{aligned} \quad (11)$$

where u_i designates a component of the vector \mathbf{u} , $\nabla \mathcal{J}_i$ is a component of the gradient of the objective function \mathcal{J} , U_i and L_i are tunable parameters, $I_+ = \{i : \nabla \mathcal{J}_i > 0\}$, and $I_- = \{i : \nabla \mathcal{J}_i < 0\}$. In (11), L_i and U_i are adjusted such that $\tilde{\mathcal{J}} > \mathcal{J}$ at the minimum of \mathcal{J} , and $\tilde{\mathcal{J}} < \mathcal{J}$ at the maximum of $\tilde{\mathcal{J}}$, for minimization and maximization problems respectively. This treatment is applied to stabilize and accelerate the convergence of the optimization process [14]. In this method, L_i and U_i act as asymptotes that change the curvature of the approximation $\tilde{\mathcal{J}}(\mathbf{u})$. Their initial values are set

to the lower and upper bounds of the optimization variables $\bar{\mathbf{u}}$. In the case of a quasi-linear objective function, the ability to adjust the curvature from the beginning of the optimization process plays an important role. If L_i and U_i are chosen “far away” from $\bar{\mathbf{u}}$, then $\tilde{\mathcal{J}}(\mathbf{u})$ has small curvature in a large neighborhood of $\bar{\mathbf{u}}$, which allows the optimizer to take large steps at the beginning of the optimization process.

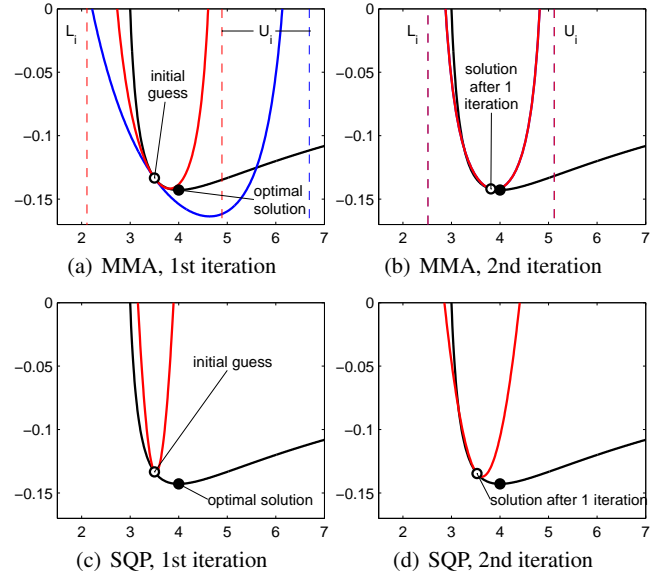


Fig. 2 (a, b) illustrate the adjustment of asymptotes in MMA. The first two iterations of MMA are compared with a trust region method (SQP) in (c, d). All figures show the objective function (black line), and the convex approximation before (thick blue line) and after (thick red line) adjustment used by the optimizer. In (a), MMA asymptotes before (vertical blue dashed line) and after (vertical red dashed lines) adjustment are shown. In (b), neither the convex approximation nor the asymptote requires adjustment.

The advantages of MMA relative to SQP can be illustrated as follows. In SQP, the curvature of the initial approximation of \mathcal{J} is assigned with a default constant value, since at this point we only have a unique instance of the gradient, which does not allow us to approximate the Hessian. The optimization process requires a certain number of iterations before a working approximation of the Hessian is constructed. In MMA, by contrast, the curvature of the approximation of \mathcal{J} at the point $\bar{\mathbf{u}}$ is directly controlled by the values of L_i and U_i . Therefore a suitable curvature is available at the first optimization iteration.

Figure 2 illustrates in a simple fashion this adjustment of the curvature. The example involves a minimization problem in one variable. It entails two iterations of MMA, which are compared with two iterations of SQP (namely the trust region SQP). The MMA optimization is started with an initial approximation (thick blue line) shown in Figure 2(a). At the first iteration, the minimum of the approximation function is below the objective function. The asymptotes (verti-

cal dashed lines) are then shifted to produce a new approximation (thick red line). At the second MMA iteration, the minimum of the approximation (thick red line) is above the objective function, so no shift in the asymptotes is needed.

Note that MMA is a trust region method, i.e., it does not perform a line search. However it is proven in [49] that MMA converges from any initial guess. MMA is also often more efficient computationally than SQP. This is because a solution of the convex programming problem in MMA is readily available. In addition, MMA requires fewer resources than SQP to store the approximation.

A major disadvantage of MMA relative to SQP, however, is that it relies on a particular form of the objective function and regularization (if any). For example, in the history matching problem, the curvature of the objective function is less prone to changes because of the need to maintain geological consistency in the reservoir parameters. For this reason MMA, which significantly alters the curvature of the approximation to fit the objective function, demonstrates poor performance for the history matching problem, in contrast to the production optimization problem.

2.4 Adjoint-based gradients

For PDE-constrained optimization problems, we must account for the implicit relationship (1) between state and control variables. This is accomplished by estimating sensitivities of the objective functions with respect to the PDE constraints, called adjoint variables [28]. One of the most efficient ways to obtain these sensitivities is through an augmented objective function called the Lagrangian \mathcal{L} , which shares the same extrema as the original objective function $\mathcal{J}(\mathbf{x}, \mathbf{u})$

$$\mathcal{L}(\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}) = \mathcal{J}(\mathbf{x}, \mathbf{u}) + \langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}, \mathbf{u}) \rangle, \quad (12)$$

where $\boldsymbol{\lambda}$ are the Lagrange multipliers for $\mathbf{g}(\mathbf{x}, \mathbf{u})$ and $\langle \cdot, \cdot \rangle$ denotes the inner product. At this stage, we eliminate the PDE constraint (1) at the cost of increasing the space of the optimization variables with $\boldsymbol{\lambda}$. After requiring that the partial derivatives of the Lagrangian (12) with respect to $\boldsymbol{\lambda}$, \mathbf{x} and \mathbf{u} vanish, we obtain the first optimality conditions, which are, respectively

- state equation (the same as (1))

$$\mathbf{g}(\mathbf{x}, \mathbf{u}) = \mathbf{0}, \quad (13)$$

- adjoint equation

$$\boldsymbol{\lambda} \mathbf{g}_{\mathbf{x}}(\mathbf{x}, \mathbf{u}) = -\mathcal{J}_{\mathbf{x}}(\mathbf{x}, \mathbf{u}), \quad (14)$$

- optimization equation

$$\boldsymbol{\lambda} \mathbf{g}_{\mathbf{u}}(\mathbf{x}, \mathbf{u}) = -\mathcal{J}_{\mathbf{u}}(\mathbf{x}, \mathbf{u}), \quad (15)$$

where $\mathbf{g}_{\mathbf{x}}$ and $\mathbf{g}_{\mathbf{u}}$ correspond to the linearization of the state equation (1) with respect to \mathbf{x} and \mathbf{u} . Equations (13)–(15) are solved sequentially in the following manner. The state variables \mathbf{x} are computed first from (13), after which the adjoint variables $\boldsymbol{\lambda}$ are found from adjoint equation (14). Equation (15) is then used to update control variables \mathbf{u} . This equation is not solved explicitly. Instead, it defines a linear relation between a perturbation $\delta \mathbf{u}$ and the total variation of \mathcal{L} , also known as the Fréchet derivative

$$d\mathcal{L} = \langle \mathcal{J}_{\mathbf{u}}(\mathbf{x}, \mathbf{u}) + \boldsymbol{\lambda} \mathbf{g}_{\mathbf{u}}(\mathbf{x}, \mathbf{u}), \delta \mathbf{u} \rangle. \quad (16)$$

The first term of the inner product in (16) is called the adjoint gradient. It is independent of $\delta \mathbf{u}$ and points in the direction of the steepest ascent of \mathcal{L} . This property is also utilized to choose between the nearest maximum or minimum. Algorithm 1 summarizes the steps in the adjoint gradient approach for a general optimization problem.

Algorithm 1 Adjoint gradient workflow for a general optimization problem

```

m ← 1
u(1) ← initial guess u0
repeat
  given estimate of u(m), solve state equations (13) for x(m)
  given x(m) and u(m), solve adjoint equations (14) for λ(m)
  update u(m+1) using (9) with a descent/ascent direction derived
  from the gradient
  ∇uℳ ≜ ℳu(x(m), u(m)) + λ(m) gu(x(m), u(m))
  m ← m + 1
until the termination criteria are satisfied to a given tolerance

```

2.5 Discrete formulation

Let us assume that, as is typical in reservoir simulation, (1) is discretized in space using a finite volume procedure and in time using the backward Euler (fully implicit) scheme. This results in a nonlinear algebraic system of equations

$$\mathbf{g}_n(\mathbf{x}_n, \mathbf{x}_{n-1}, \mathbf{u}_n) = \mathbf{0}, \quad (18)$$

$$\mathbf{x}_0 = \mathbf{x}(t_0),$$

where $\mathbf{x}_n = \mathbf{x}(t_n)$, $n = 1, \dots, N$, is the discretized state variable at the end of time step n (which is of size Δt_n). The AD-GPRS solution procedure for solving (18), for multiphase, multicomponent flow, is described in [47].

The continuous objective function (5) is approximated by a first-order scheme

$$\mathcal{J}(\mathbf{x}, \mathbf{u}) = \sum_{n=1}^N \Delta t_n \mathcal{F}_n(t_n, \mathbf{x}_n, \mathbf{u}_n) + \mathcal{O}(\max\{\Delta t_n\}). \quad (19)$$

Following the steps described in Section 2.4, the discrete adjoint gradient with respect to control variables \mathbf{u}_n is given by

$$\nabla_{\mathbf{u}} \mathcal{J} = \sum_{n=1}^N \left(\Delta t_n \frac{\partial \mathcal{F}_n}{\partial \mathbf{u}_n} + \boldsymbol{\lambda}_n^T \frac{\partial \mathbf{g}_n}{\partial \mathbf{u}_n} \right), \quad (20)$$

where the discrete adjoint variables $\boldsymbol{\lambda}_n$ satisfy

$$\begin{aligned} \boldsymbol{\lambda}_n^T \frac{\partial \mathbf{g}_n}{\partial \mathbf{x}_n} &= - \left(\boldsymbol{\lambda}_{n+1}^T \frac{\partial \mathbf{g}_{n+1}}{\partial \mathbf{x}_n} + \Delta t_n \frac{\partial \mathcal{F}_n}{\partial \mathbf{x}_n} \right), \\ \boldsymbol{\lambda}_N^T \frac{\partial \mathbf{g}_N}{\partial \mathbf{x}_N} &= - \Delta t_N \frac{\partial \mathcal{F}_N}{\partial \mathbf{x}_N}, \quad n = 1, 2, \dots, N-1. \end{aligned} \quad (21)$$

Here and throughout this paper, $(\cdot)^T$ denotes transpose. More details on the derivation of (20) and (21) can be found in [34].

2.6 Automatic differentiation

As should be clear from our discussion above, the use of discrete adjoint gradients (20) is the key component of our CLRM procedure. This approach is appropriate for CLRM as it (1) provides all partial derivatives in (20)–(21) based on the numerical solution of the reservoir and adjoint equations (13)–(14), and (2) does not lead to excessive overhead or restrictions on the solution of the reservoir equations themselves. It is important to note that the use of adjoints is much more efficient than obtaining numerical partial derivatives of the functions \mathcal{F}_n and \mathbf{g}_n with respect to state and control variables using a finite difference scheme, which becomes very expensive as the number of optimization variables increases.

Analytical differentiation has traditionally been used within reservoir engineering to construct the adjoint equations. This requires access to source code and detailed calculations and coding, which are prone to error. In addition, when updating or enhancing the capabilities of the forward simulation model, the detailed adjoint treatment must be revisited.

Due in part to its ability to overcome the problems mentioned above, automatic differentiation is gaining popularity in scientific and high performance computing. On the one hand, it allows analytical, i.e., exact and fast, evaluation of the partial derivatives of all nonlinear functions and quantities involved in the formulation of the reservoir equations. On the other hand, any new physics added to the existing reservoir model is automatically taken into account. The novelty of our computational approach applied to CLRM involves utilizing AD in obtaining gradients for both the production optimization and history matching components. We note that AD was used in previous work for the computation of gradients for production optimization in compositional simulation [24].

The reservoir equations are in general PDEs of parabolic type. Therefore, their adjoints are also parabolic, and are solved backward in time as seen in (21). Thus, the reservoir equations and their adjoints must be solved sequentially. One approach would be to store all partial derivatives employed in (21), which amounts approximately to $N_{\text{var}} \times N_{\text{var}} \times N$ floating point numbers, where N_{var} is the total number of state variables obtained after spatial discretization and N is the number of time steps in the discrete reservoir model (18). In addition, since the terms $\partial \mathcal{F}_n / \partial \mathbf{x}_n$, $\partial \mathcal{F}_n / \partial \mathbf{u}_n$, $\partial \mathbf{g}_n / \partial \mathbf{u}_n$ are not required in the forward simulation, computing and storing them will unavoidably slow down the process. Our AD-based reservoir simulator AD-GPRS adopts a different approach by storing only the state variables, enumerated to approximately $N_{\text{var}} \times N$, and re-assembling the residual \mathbf{g}_n and the constitutive parts of \mathcal{J} as required. The optimization module in AD-GPRS and the AD framework itself ensure the fidelity of this reassembly. In addition, the framework is customized to adapt to changes in the optimization variables caused, for example, by changes in well controls during the simulation (e.g., switches from BHP to rate control).

2.7 Handling bounds and constraints

Most of the optimization parameters in CLRM are subject to limitations. For example, BHPs and flow rates are restrained by technical limits on the wells and facilities. The geological model parameters such as permeability and porosity are constrained to fall within physical ranges. In the optimization problems these limitations are translated into bounds on the optimization variables \mathbf{u} , which are satisfied by the optimizer. If a particular limitation is not applied directly to an optimization variable, it enters as a nonlinear output constraint. This is the case, for example, for constraints on phase flow rates when the well control is BHP. In this situation, constraints can be satisfied either during the simulation or within the optimization [24, 35]. Numerical studies [24, 43] have shown that, when using an SQP solver, it is often advantageous to enforce the constraints during the simulation (even though this requires the use of some heuristics) rather than in the optimization.

In this work, we apply an approach related to that in [24] and proceed as follows. For purposes of illustration, assume we have a production well operating under BHP control (with maximum and minimum BHP bounds specified) that is subject to a maximum water rate constraint. The production optimization module will have provided the target BHP for the time step under consideration. If the use of this BHP does not lead to a violation of the water rate constraint at the current Newton iteration, then this BHP acts as the control. If, however, this target BHP does lead to a violation of the water rate constraint at the current Newton iteration,

then we perform an approximate test to assess whether or not the well can operate under water rate control, with the target given by the maximum water rate constraint, while not having the BHP fall below the value determined by the optimizer. If the test indicates that the rate control specification is feasible (with respect to BHP), then the well is operated under rate control. If rate control is infeasible, then the well is shut in for this Newton iteration. Analogous procedures are applied for other production and injection well constraints. It is important to note that wells can also be shut in during an AD-GPRS run when back flow occurs; i.e., when a zone intersected by the injector actually produces into the well or a zone intersected by a producer actually injects fluid into the reservoir.

2.8 Objective functions

2.8.1 Production optimization

The objective function used in this work for production optimization is net present value (NPV). The discretized cash-flow appearing in the objective function (19) as the term \mathcal{F}_n is expressed as

$$\mathcal{F}_n(t_n, \mathbf{x}_n, \mathbf{u}_n) = \sum_{j=1}^{N_{\text{well}}} \sum_{p=1}^{N_p} C_{j,p}(t_n) q_{j,p}(\mathbf{x}_n, \mathbf{u}_n), \quad (22)$$

where $q_{j,p}$ and $C_{j,p}$ denote respectively the volumetric production/injection flow rate and discounted price of phase p in well j , and N_p designates the number of phases. Note that in production optimization the objective function (19) is implemented in the problem (6b).

2.8.2 History matching

When solving the history matching problem (6a), the model parameters \mathbf{u} are grid-block or interface properties of the discrete reservoir, e.g., porosity, permeability, or interblock transmissibility. It is common to define an objective function in a least-square data mismatch fashion, e.g., [30]:

$$\mathcal{J}(\mathbf{x}, \mathbf{u}) = (G(\mathbf{x}, \mathbf{u}) - \mathbf{d}_{\text{obs}})^T C_D^{-1} (G(\mathbf{x}, \mathbf{u}) - \mathbf{d}_{\text{obs}}) + \mathcal{R}, \quad (23)$$

where \mathbf{d}_{obs} are observed production and/or seismic data (the latter are treated here as approximate phase saturation data), $G(\mathbf{x}, \mathbf{u})$ is the model response given by (1), matrix C_D represents the measurement error, and \mathcal{R} is an appropriate regularization term, discussed in Section 2.9.

Assuming that measurement k in the sequence of N_{obs} observations, $k = 1, \dots, N_{\text{obs}}$, is collected at time $\tau_k = t_n$, the function $\mathcal{F}_n(t_n, \mathbf{x}_n, \mathbf{u}_n)$ in (19) may be rewritten as

$$\mathcal{F}_n(t_n, \mathbf{x}_n, \mathbf{u}) = f_k(\mathbf{x}_n, \mathbf{u}) \delta(t_n - \tau_k), \quad (24)$$

where $\delta(t_n - \tau_k) = 1 / \Delta t_n$ is the first-order approximation of the Dirac delta function. The particular form of f_k depends on the type of measured data, namely

- well production rate data

$$f_k(\mathbf{x}_n, \mathbf{u}) = \sum_{j=1}^{N_{\text{well}}^q} \sum_{p=1}^{N_p} C_{j,p} (q_{j,p} - \tilde{q}_{j,p})^2, \quad (25)$$

- well BHP data

$$f_k(\mathbf{x}_n, \mathbf{u}) = \sum_{j=1}^{N_{\text{well}}^{\text{BHP}}} C_j (p_j^{\text{BHP}} - \tilde{p}_j^{\text{BHP}})^2, \quad (26)$$

- time-lapse seismic data, e.g., phase saturation s_p , given by a seismic inversion or a proxy of seismic observations

$$f_k(\mathbf{x}_n, \mathbf{u}) = \sum_{j=1}^{N_{\text{block}}} \sum_{p=1}^{N_p-1} C_p (s_{j,p} - \tilde{s}_{j,p})^2, \quad (27)$$

where coefficients $C_{j,p}$, C_j , C_p are diagonal elements of matrix C_D^{-1} , $q_{j,p}$, p_j^{BHP} and $s_{j,p}$ designate simulated phase production rates, well BHPs and phase saturation, and $\tilde{q}_{j,p}$, \tilde{p}_j^{BHP} and $\tilde{s}_{j,p}$ indicate the corresponding observed quantities. Here N_{well}^q and $N_{\text{well}}^{\text{BHP}}$ are the numbers of wells providing phase production rate and BHP data, respectively, while N_{block} is the number of reservoir grid blocks where seismic measurements are available.

Well flow rates for different phases can differ substantially. We address this by changing volumetric rates to mass rates for production wells and by scaling the various contributions to $\mathcal{J}(\mathbf{x}_n, \mathbf{u})$. Specifically, (25) is replaced by

$$f_k(\mathbf{x}_n, \mathbf{u}) = \sum_{j=1}^{N_{\text{well}}^q} \sum_{p=1}^{N_p} C_{j,p} D_p^2 (q_{j,p} - \tilde{q}_{j,p})^2, \quad (28)$$

where D_p is a phase relative density defined as follows

$$D_p = \frac{\rho_p}{\max\{\rho_o, \rho_w\}}, \quad (29)$$

where ρ_p designates phase $p = o, w$ density at standard conditions.

Scaling is required because the sensitivity of the full objective function $\mathcal{J}(\mathbf{x}_n, \mathbf{u})$ with respect to \mathbf{u} may vary by orders of magnitude for the different contributions $f_k(\mathbf{x}_n, \mathbf{u})$ in (25)–(27). Let us denote by \mathcal{J}_i , $i = 1, 2, 3$, the terms of the objective function corresponding to the three types of data as defined in (25)–(27). To balance the contributions \mathcal{J}_i , we weight each data type using the initial values of \mathcal{J}_i (designated $\mathcal{J}_i^{(1)}$). We thus introduce weighting factors α_i , which appear in the calculation of \mathcal{J} as follows

$$\mathcal{J} = \sum_{i=1}^3 \alpha_i \mathcal{J}_i, \quad \alpha_i = \frac{\min\{\mathcal{J}_i^{(1)}\}}{\mathcal{J}_i^{(1)}}. \quad (30)$$

2.9 PCA-based parameterization for history matching

From a mathematical viewpoint, the history matching problem as stated above is usually over-parameterized for large models. This is the case because the amount of observed data is small compared to the number of unknown model parameters. The added complication is that the underlying optimization problem is nonconvex. This results in large uncertainty regarding the optimal solution; i.e., the optimal solution is nonunique. From the geological perspective, the solution of the history matching problem should honor available prior information. This information can be quantified from a set of prior geological realizations, as described below.

From both viewpoints, history matching requires regularization, which can be incorporated by augmenting the objective function \mathcal{J} in (23) with an additional model mismatch term \mathcal{R} . This term can be represented as a weighted difference between the solution for model parameters \mathbf{u} and a reference set of parameters \mathbf{u}_{ref}

$$\mathcal{R}(\mathbf{u}) = \alpha_{\mathcal{R}}(\mathbf{u} - \mathbf{u}_{\text{ref}})^T \mathcal{C}_{\mathcal{R}}^{-1}(\mathbf{u} - \mathbf{u}_{\text{ref}}), \quad (31)$$

where $\alpha_{\mathcal{R}}$ is a scaling coefficient. The choice of matrix $\mathcal{C}_{\mathcal{R}}$ corresponds to different types of regularization: finite difference matrices for Tikhonov-type [11], or a covariance matrix, denoted by $\mathcal{C}_{\mathcal{M}}$, to maintain spatial correlation in history matching problems [30]. The former is not considered in this paper, whereas the latter is incorporated through an appropriate control space re-parameterization technique.

In general, the control space can be re-parameterized based on various types of information, e.g., structure of the gradient, geological or spatial features, geostatistics, etc. Here we consider only re-parameterization based on Principal Component Analysis (PCA), which is also known as Proper Orthogonal Decomposition or Karhunen–Loève Expansion.

We use the PCA representation to convert a set of prior realizations characterized by correlated variables (parameters \mathbf{u}) into a set of linearly uncorrelated variables (principal components ξ) through application of an orthogonal transformation. Below we provide a brief description of the general approach for mapping model parameters \mathbf{u} to new control variables ξ by performing PCA based on Singular Value Decomposition (SVD). For further details refer to [36, 37, 42].

Without loss of generality, we consider a geological model which contains N_u model parameters. We assume the existence of a set of N_r realizations \mathbf{u}_j , $j = 1, \dots, N_r$, each of size N_u , that may or may not be conditioned to hard data (these realizations would typically be generated using geostatistical modeling software). For simplicity, we could also assume a Gaussian (normal) distribution for the model parameters, i.e., $\mathbf{u} \sim \mathbb{N}(\bar{\mathbf{u}}, \mathcal{C}_{\mathcal{M}})$, where $\bar{\mathbf{u}} = (1/N_r) \sum_{j=1}^{N_r} \mathbf{u}_j$.

The aforementioned covariance matrix $\mathcal{C}_{\mathcal{M}}$ may be approximated by

$$\mathcal{C}_{\mathcal{M}} \approx \frac{\mathcal{X}\mathcal{X}^T}{N_r - 1}, \quad \mathcal{X}^{N_u \times N_r} = [\mathbf{u}_1 - \bar{\mathbf{u}} \ \dots \ \mathbf{u}_{N_r} - \bar{\mathbf{u}}]. \quad (32)$$

It is more efficient to perform SVD on the matrix $\mathcal{Y} = \mathcal{X}/\sqrt{N_r - 1}$ of size $N_u \times N_r$ rather than on the covariance matrix $\mathcal{C}_{\mathcal{M}}$ of size $N_u \times N_u$, as $N_r \ll N_u$. The SVD factorization with truncation is then applied to matrix \mathcal{Y}

$$\mathcal{Y} \approx \tilde{\mathcal{U}}_{N_{\xi}} \tilde{\Sigma}_{N_{\xi}} \tilde{\mathcal{V}}_{N_{\xi}}^T, \quad (33)$$

where diagonal matrix $\tilde{\Sigma}_{N_{\xi}}$ contains the singular values of \mathcal{Y} , and matrices $\tilde{\mathcal{U}}_{N_{\xi}}$ and $\tilde{\mathcal{V}}_{N_{\xi}}^T$ contain the left and right singular vectors of \mathcal{Y} . The matrix $\tilde{\Sigma}_{N_{\xi}}$ is truncated to keep only N_{ξ} singular values σ_k , $k = 1, \dots, N_{\xi}$ (analogous truncations are applied to $\tilde{\mathcal{U}}_{N_{\xi}}$ and $\tilde{\mathcal{V}}_{N_{\xi}}^T$).

We define a linear transformation

$$\Phi^{N_u \times N_{\xi}} = \tilde{\mathcal{U}}_{N_{\xi}} \tilde{\Sigma}_{N_{\xi}}, \quad N_{\xi} \leq N_{\min} = \min\{N_u, N_r\}, \quad (34)$$

to project the initial control space, in which the model parameters \mathbf{u} were defined, onto the reduced-dimension ξ -space, containing only the N_{ξ} largest principal components ξ , by means of the unique mapping

$$\mathbf{u} = \Phi \xi + \bar{\mathbf{u}}. \quad (35)$$

For large models, the SVD required to construct Φ can be quite time consuming. More efficient algorithms, for example the kernel PCA approach applied by [36, 42], can be used in such cases.

Various approaches can be used to determine the size of the ξ -space; i.e., the N_{ξ} value. Options include the Kaiser criterion $\sigma^2 \geq 1$ [23], the scree test [6], and the inclusion of a prescribed portion of the variance (energy) contained in eigenvalues $\lambda_1 = \sigma_1^2$ through $\lambda_k = \sigma_k^2$. With this last approach, given the (prescribed) parameter r_{opt} , N_{ξ} is determined such that the following condition is satisfied

$$\frac{\sum_{k=1}^{N_{\xi}} \sigma_k^2}{\sum_{k=1}^{N_{\min}} \sigma_k^2} \geq r_{\text{opt}}. \quad (36)$$

When constructing the backward mapping, the simplest approach is to approximate the inverse of the matrix Φ (which cannot be inverted due to its size $N_u \times N_{\xi}$) using a pseudo-inverse matrix $\hat{\Phi}^{-1}$. Then

$$\xi = \Phi^{-1}(\mathbf{u} - \bar{\mathbf{u}}) \approx \hat{\Phi}^{-1}(\mathbf{u} - \bar{\mathbf{u}}), \quad (37)$$

where

$$\hat{\Phi}^{-1} = \tilde{\Sigma}_{N_{\xi}}^{-1} \tilde{\mathcal{U}}_{N_{\xi}}^T. \quad (38)$$

The matrix $\hat{\Phi}^{-1}$ is derived using (34)–(35) and the orthonormality of the matrix $\tilde{\mathcal{U}}_{N_{\xi}}$.

The optimization problem (6a) can now be restated in terms of the new model parameters ξ as follows

$$\hat{\xi} = \underset{\xi}{\operatorname{argmin}} \mathcal{J}(\mathbf{x}_n, \xi), \quad (39)$$

subject to the discretized reservoir model (18), and using the mappings given by (35) and (37). By applying (35) and the chain rule for derivatives, the gradient $\nabla_{\xi} \mathcal{J}$ of the objective function \mathcal{J} with respect to the control variables ξ can be expressed as

$$\nabla_{\xi} \mathcal{J} = \Phi^T \cdot \nabla_{\mathbf{u}} \mathcal{J} = \tilde{\Sigma}_{N_{\xi}} \tilde{U}_{N_{\xi}}^T \cdot \nabla_{\mathbf{u}} \mathcal{J}. \quad (40)$$

This expression is derived by projecting the gradient $\nabla_{\mathbf{u}} \mathcal{J}$ defined in (20) from the initial space onto the reduced-dimension ξ -space.

In the results that follow, we apply the PCA-based re-parameterization described above, which constrains the geological model to honor hard data and “resemble” the a priori realizations in terms of spatial correlation structure. We do not, however, directly include the regularization term $\mathcal{R}(\mathbf{u})$ in the objective function. For PCA-based re-parameterization procedures, it has been shown in, e.g., [42], that a regularization of the form $\xi^T \xi$ appears in the computation of the maximum a posteriori (MAP) estimate in ξ -space. This term could be easily incorporated into our treatment as required.

A summary of the complete optimization scheme to solve the minimization problem (6a), utilizing a solution of the problem (39), is provided in Algorithm 2.

Algorithm 2 Optimization workflow utilizing PCA-based control space re-parameterization

```

m ← 1
u(1) ← initial guess u0
construct Φ and Φ-1 by (34) and (38)
ξ(1) ← u(1) using (37)
repeat
  given estimate of u(m), solve state equations (18) for x(m)
  given x(m) and u(m), solve adjoint equations (21) for λ(m)
  ∇u J(u(m)) ← u(m), x(m), λ(m) by (20)
  ∇ξ J(ξ(m)) ← ∇u J(u(m)) by (40)
  update ξ(m+1) with a descent direction derived from (9) by
  ξ(m+1) = ξ(m) - τ(m) ∇ξ J(ξ(m))
  u(m+1) ← ξ(m+1) by (35)
  m ← m + 1
until the termination criteria are satisfied to a given tolerance

```

3 Computational algorithm for CLRM

3.1 General description

In this work the computational model used for CLRM, as shown in Figure 1, consists of three main parts:

- production optimization (PO), in which problem (6b) is solved,
- history matching (HM), in which problem (6a) is solved,
- data handling (production and/or seismic) generated from the true or emulated production system.

All three parts are also represented by the blue rectangles within the computational flow chart shown in Figure 3.

We introduce the following general notation in the detailed description of the closed-loop model used in this paper, which is presented in Figure 3 and Algorithm 3:

- The parameter N_{ℓ} is the number of cycles (iterations) in the CLRM. The subscript index $\ell = 1, \dots, N_{\ell}$ indicates the CLRM cycle.
- \mathcal{S} defines the reservoir facility (well) schedule. Thus \mathcal{S}_0 is the initial well schedule, while $\mathcal{S}_{\text{opt}} = \mathcal{S}_{N_{\ell}}$ designates the optimal schedule obtained as a result of solving the full CLRM optimization problem.
- \mathcal{R} denotes a set of geological parameters. \mathcal{R}_0 and $\mathcal{R}_{\text{opt}} = \mathcal{R}_{N_{\ell}-1}$ describe correspondingly the initial and optimal representations of the geological parameters.
- \mathcal{H} denotes all of the data available at the current time t_{curr} .
- Δt_{ℓ} defines the duration of one CLRM cycle, e.g., one year.

In the CLRM description in Algorithm 3 and Figure 3, we assume that initial reservoir parameters \mathcal{R}_0 are available based on prior geological knowledge. This enables us to first solve the production optimization problem. In fact, for any CLRM cycle ℓ , we solve the PO problem first (at $t = t_{\text{curr}}$), and the HM problem second (after accumulating new data; i.e., at $t = t_{\text{curr}} + \Delta t_{\ell}$). If a prior model is not available, we would reverse the order of the CLRM solutions (solve HM first, PO second). This will not introduce any additional complications into the CLRM procedure.

3.2 Representation of observed data

Following the notation introduced in Section 2.8.2, measurements may include phase production rates \tilde{q}_p , phase saturation \tilde{s}_p , and BHPs \tilde{p}^{BHP} . If these measurements are obtained from a true production system, noise is present in the observed data \mathbf{d}_{obs} in (23). Here, however, we use the reservoir model (18) to generate historical data, assuming that the true geological properties $\mathcal{R}_{\text{true}}$ are known. In this case synthetic noise could be introduced as follows

$$\mathbf{d}_{\text{obs}} = \mathbf{d}_{\text{true}} + \Sigma \mathbf{y}, \quad (42)$$

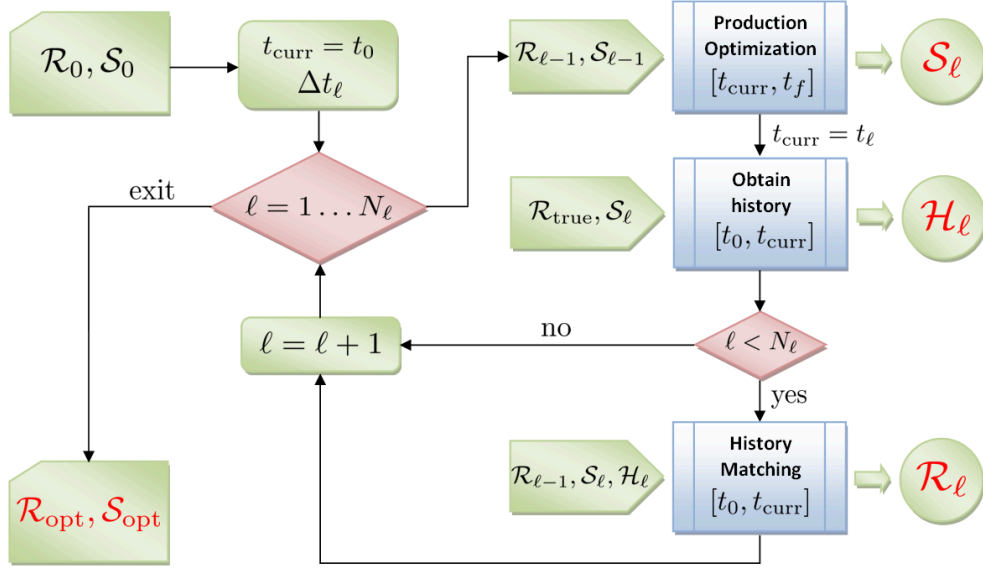


Fig. 3 Computational flow chart for closed-loop reservoir management.

Algorithm 3 Workflow for CLRM

$\mathcal{R}_0 \leftarrow$ initial reservoir characteristics
 $\mathcal{S}_0 \leftarrow$ initial well schedule
 $\Delta t_\ell \leftarrow \frac{t_f - t_0}{N_\ell}$
 $t_{\text{curr}} \leftarrow t_0$
for $\ell = 1$ **to** N_ℓ **do**
 if $\ell > 1$ **then**
 simulate $(\mathcal{R}_{\ell-1}, \mathcal{S}_{\ell-1})$ for $t \in [t_0, t_{\text{curr}}]$
 end if
 Algorithm 1 for PO $(\mathcal{R}_{\ell-1}, \mathcal{S}_{\ell-1})$ for $t \in [t_{\text{curr}}, t_f]$
 $\mathcal{S}_\ell \leftarrow \mathcal{S}_{\ell-1}$ optimized for $t \in [t_{\text{curr}}, t_f]$
 $t_{\text{curr}} \leftarrow t_{\text{curr}} + \Delta t_\ell$
 $\mathcal{H}_\ell \leftarrow \mathcal{R}_{\text{true}}$ and \mathcal{S}_ℓ for $t \in [t_0, t_{\text{curr}}]$
 if $\ell < N_\ell$ **then**
 Algorithm 2 for HM $(\mathcal{R}_{\ell-1}, \mathcal{S}_\ell, \mathcal{H}_\ell)$ for $t \in [t_0, t_{\text{curr}}]$
 $\mathcal{R}_\ell \leftarrow \mathcal{R}_{\ell-1}$ optimized
 end if
end for

where \mathbf{d}_{true} represents data without noise, \mathbf{y} is a vector of random numbers drawn from a normal distribution $\mathbb{N}(0, 1)$, and Σ is a diagonal matrix. The components σ_i of matrix Σ play the role of standard deviations, defined for BHP data as

$$\sigma_{\text{BHP}} = \gamma_{\text{BHP}}, \quad (43)$$

and for phase production rate q_p data as

$$\sigma_q = \begin{cases} \sigma_{q,\min}, & \gamma_q \cdot q_p < \sigma_{q,\min}, \\ \gamma_q \cdot q_p, & \sigma_{q,\min} \leq \gamma_q \cdot q_p \leq \sigma_{q,\max}, \\ \sigma_{q,\max}, & \gamma_q \cdot q_p > \sigma_{q,\max}, \end{cases} \quad (44)$$

where γ_{BHP} , γ_q , $\sigma_{q,\min}$ and $\sigma_{q,\max}$ are prescribed (constant) “noise” parameters.

3.3 Modeling proxy seismic data

Time-lapse (4D) seismic observations can be used as an additional data source during reservoir production. These results are usually considered in the form of a solution of a separate inverse problem, e.g., full waveform inversion (FWI). In this problem, geophysical model parameters, which are coefficients in the seismic wave equation, are estimated by matching, for example, calculated seismograms and observed data [13, 40]. Performing FWI requires the solution of the seismic wave equation as a forward problem. For current purposes, however, the process of obtaining the field property data can be emulated using a synthetic model, as described in, e.g., [8, 12, 31]. A degree of realism can be achieved by adding noise to the synthetic data, which in our case are phase saturations. As the level of noise increases, however, the information content of these data will decrease. In addition, the seismic wavelength, which typically differs from the dimensions of reservoir grid blocks, can be incorporated by applying spatial filtering.

In this work, as indicated in (27), time-lapse seismic data enters in the form of estimates for the phase saturation \tilde{s}_p . We use the simulated saturation data, in some cases supplemented with noise, as a proxy. Figure 4 illustrates schematically the general process of constructing proxy seismic data $\mathbf{d}_{\text{proxy}}$. Three steps are involved:

1. Generate “true” data \mathbf{d}_{true} from the solution of the reservoir equation (18) performed with the true field $\mathcal{R}_{\text{true}}$.
2. Add noise to the simulated data \mathbf{d}_{true} . This noise may be described by (42), where standard deviation σ_s is given by

$$\sigma_s = \gamma_s(1 - \phi), \quad (45)$$

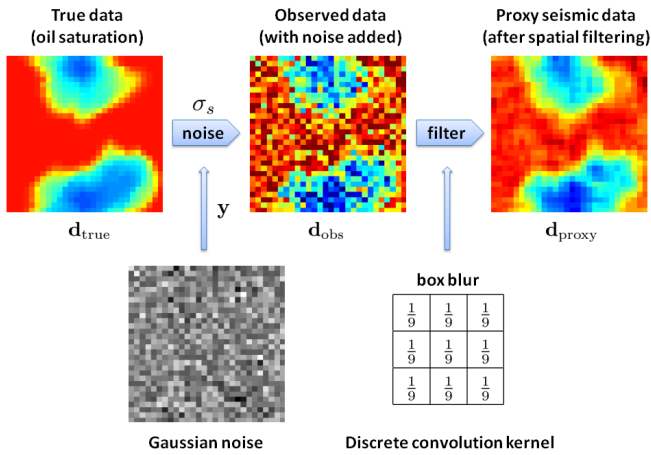


Fig. 4 Schematic showing the procedure for obtaining proxy seismic data for a 2D example.

where γ_s is a constant noise parameter. This simple form accounts qualitatively for the fact that the seismic measurement is more precise when porosity is higher. When a gas phase is present, an additional multiplier of the form $(1 - s_g^{\text{true}})$, where s_g^{true} is the true gas saturation, would appear in (45).

3. Apply spatial filtering to the simulated data with added noise; i.e., average observed data \mathbf{d}_{obs} over the half wave-length.

Incorporating noise of a specified percentage η entails adjustment of the noise parameter γ_s in (45).

Step 3 can be accomplished by employing a simple spatial filtering operation, e.g., a discrete convolution with a rectangular kernel, which also has a smoothing effect. As an example for the 2D case, Figure 4 shows a 3×3 “box blur” discrete kernel (convolution matrix). Filtering with this template replaces values for a given pixel in the original image by the average of the target pixel and its eight direct neighbors. Kernels of this class are also called low pass filters, as they remove high image frequencies and also contribute to noise reduction. Details on other filtering techniques, as well as algorithms for their practical implementation, are available in the image processing literature; see, e.g., [17, 33].

4 Computational results

In this section we present computational results for the closed-loop reservoir modeling described in Sections 2 and 3. We first overview the reservoir simulation and optimization packages utilized in this work. Then, after a brief discussion of the model and optimization parameters, we assess the CLRM framework and the performance of the production optimization and history matching components. Our

intent in these computations is to demonstrate the general capabilities of our adjoint gradient-based CLRM.

Prior to running the CLRM, we performed a diagnostic test commonly employed to verify the correctness of the gradient computation [4,5]. This test entails computing the total variation of the objective function in an arbitrary direction $\delta \mathbf{u}$ using (16). Here $\delta \mathbf{u} = \varepsilon \cdot \mathbf{1}$, where $\mathbf{1}$ denotes the vector of ones of the same dimension as $\delta \mathbf{u}$, and ε is a parameter (which we vary). The variation (16) computed from the AD implementation was then compared with its finite difference approximation for a range of ε using different values of the maximum time step Δt_{max} . We observed close agreement between the AD and finite difference approximations over a several order-of-magnitude range in ε , which confirms the accuracy of the AD gradients used in our CLRM framework.

4.1 Optimization software

All computational results presented in Section 4 are obtained using AD-GPRS. The simulator is used both for forward simulations and for constructing adjoint gradients, which are provided to an SCP solver. The SCP packages used in this work are SNOPT and NLOPT, which provide robust implementations of SQP and MMA, respectively. Details on the use of these packages can be found in [16] and [22].

The termination criteria available in the SCP solvers are a maximum number of gradient evaluations (for both solvers), a tolerance on the relative change of the objective function (for NLOPT), and a tolerance on the violation of the optimality conditions (for SNOPT).

To improve the performance of the SQP method provided by SNOPT, all objective function values for production optimization are scaled by the constant coefficient $1 \cdot 10^{-7}$. The resulting PO objective function values are then of order $10^0 - 10^2$. For history matching, the production data component of the objective function is scaled as shown in (30), with α set to $1.18 \cdot 10^{-7}$. The initial values of the HM objective are then of order $10^0 - 10^1$. We note finally, that because SNOPT constructs a gradient of large dimension for both the PO and HM problems, we use the limited memory L-BFGS approximation in place of the full Hessian approximation [16].

4.2 Model parameters

The optimization module in AD-GPRS is able to perform optimization for different types of model parameters. The adjoint gradient framework is only suitable for continuous-valued variables. The available control parameters for production optimization are well BHPs and phase flow rates, which can be used simultaneously. For history matching

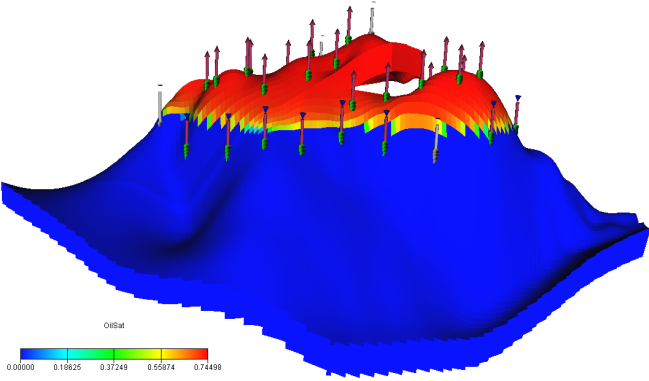


Fig. 5 Brugge model (reservoir structure and initial oil saturation).

the optimization parameters are defined as block-by-block multipliers either to grid-block permeabilities or interblock transmissibilities.

The CLRM results presented below are obtained using well BHPs $\mathbf{p}_n^{\text{BHP}}$ and multipliers $\boldsymbol{\mu}_T$ to interblock transmissibilities as the optimization parameters. As the transmissibility field in the model described in Section 4.3 is of Gaussian type, history matching parameters are subject to log-scale parameterization, i.e.,

$$\begin{aligned} \text{PO: } \mathbf{u}_n &= \mathbf{p}_n^{\text{BHP}}, \\ \text{HM: } \mathbf{u} &= \ln \boldsymbol{\mu}_T, \quad \nabla_{\ln \boldsymbol{\mu}_T} \mathcal{J} = \boldsymbol{\mu}_T \nabla_{\boldsymbol{\mu}_T} \mathcal{J}. \end{aligned} \quad (46)$$

Available prior information is incorporated by applying the PCA-based re-parameterization technique discussed in Section 2.9.

4.3 Brugge test case model

The model used in this work is based on the Brugge benchmark case [8, 31, 32]. We preserve the geometry of the Brugge model, shown in Figure 5, including the grid structure and the distribution of active cells. A total of 104 geological realizations exist for this case, with each realization containing information on grid-block depths and volumes, net to gross ratio, porosity and permeability distributions, and initial pressure and saturations. To apply the PCA-based re-parameterization, the original realizations have been converted into realizations for interblock transmissibilities. Realization #73 is removed from this set and is designated to be the “true” transmissibility field. This model is then used to generate synthetic production and oil saturation data, both of which provide observed data (the treatments discussed in Sections 3.2 and 3.3 can additionally be applied). The remaining 103 realizations are used to construct a PCA-based representation of the interblock transmissibility. Realization #101 is used as the initial guess for history matching, and the initial well schedule S_0 is taken from the original Brugge test case model. The model contains only oil and water.

Continuous-valued adjoint history matching imposes certain restrictions on the reservoir model. For example, only transmissibilities in layers 3 through 8 have consistent sets of active grid blocks across realizations. This problem could be treated by making all blocks active and introducing fictitious transmissibilities, though for now we simply limit our model to include only layers 3 – 8. Except where otherwise stated, the computational results presented in Section 4 are obtained without adding noise to the production and seismic data. Other model and optimization parameters are presented in Table 1. The original Brugge model parameters are also shown for comparison. Additional specifications for this model can be found in [31]. In all computations in this paper, we use $\Delta t_{\max} = 10$ days.

4.4 Closed-loop performance

Evaluation of CLRM performance involves consideration of the two distinct optimization problems within the overall procedure. We assess the CLRM quality in terms of the production optimization objective function NPV, computed over the full time frame (20 years), as a function of \mathcal{S} and \mathcal{R} . We compute NPV after each closed-loop iteration ℓ . Two different NPVs, denoted $\text{NPV}(\mathcal{S}_\ell, \mathcal{R}_\ell)$ and $\text{NPV}(\mathcal{S}_\ell, \mathcal{R}_{\text{true}})$, are computed. These values correspond to expected and observed results, respectively, and will be referred to as “model response” and “reservoir response.”

Figure 6 shows the evolution of $\text{NPV}(\mathcal{S}_\ell, \mathcal{R}_\ell)$ toward $\text{NPV}(\mathcal{S}_\ell, \mathcal{R}_{\text{true}})$ as ℓ increases from 1 to 20. Results are presented for two PO solvers — SQP and MMA. The HM solver is always SQP (see Section 2.3). The reservoir response (red points) converges quickly to a constant value. Similar behavior is observed in the model response (blue points), and the model and reservoir NPV values are very close for $\ell \geq 8$ for SQP and $\ell \geq 5$ for MMA. The “convergence” of the model NPV to the true reservoir NPV cannot be expected to be observed in general. It occurs in this case because the 104 Brugge model realizations cluster into a few sets, and within a particular set the level of variation is relatively low, and because we use “true” saturation data as proxy seismic data. As a result, the history matched model closely resembles the true model after a few CLRM iterations (as illustrated later). Nonetheless, the results in Figure 6 demonstrate that our CLRM procedure is indeed operating in the desired manner. Results in which noise and filtering are added to the proxy seismic data will be presented below.

4.5 Performance of production optimization (PO)

To assess the PO performance by itself, we introduce a reference solution S^* , which is obtained by performing a single

Table 1 Parameters for the computational model.

Parameter	Units	Notation	Computational model	Original Brugge model
Grid size			$139 \times 48 \times 6$	$139 \times 48 \times 9$
Wells		N_{well}	30 standard wells	30 smart wells
Production life	years	N_{ℓ}	20	30
Closed-loop schedule			20×1 -year cycles	2×10 -year cycles
Geological realizations		N_r	103 (transmissibility)	104 (permeability, porosity, NTG)
True reservoir properties		$\mathcal{R}_{\text{true}}$	realization #73	not given
Rock and fluid properties			same	given
Initial pressure and saturation			same	given
Initial facility (well) schedule		\mathcal{S}_0	same	given
Initial reservoir properties		\mathcal{R}_0	realization #101	
NPV, oil price/water cost	US\$/bbl		80/5	80/5
NPV discount			10%	10%
BHP bounds, lower/upper	bar		50/180	50/180
Maximum fluid rate, injector/producer	bbl/day		4000/3000	4000/3000
BHP production data		\bar{p}^{BHP}	not used	monthly for the first 10 years
Rate production data		\tilde{q}_p	monthly for 20 years	daily for the first 10 years
4D-seismic data (proxy saturation)		\tilde{s}_p	yearly, synthetic	at time 0 and 10 years, “inverted”

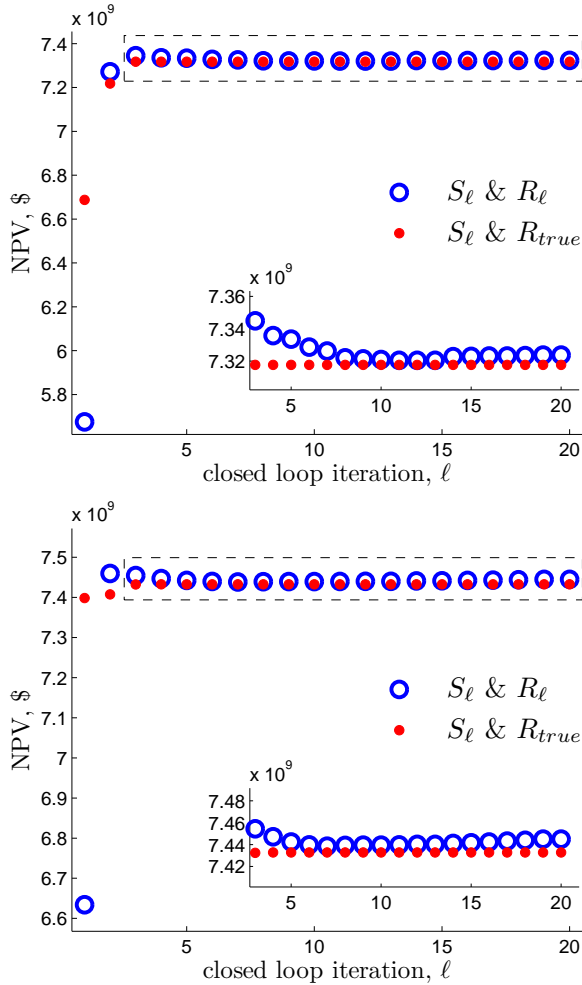


Fig. 6 NPV at the end of the reservoir life (20 years) as a function of closed-loop iteration ℓ for (top) SQP and (bottom) MMA. The red dots represent the reservoir response $\text{NPV}(S_\ell, \mathcal{R}_{\text{true}})$, while the blue circles show model response $\text{NPV}(S_\ell, \mathcal{R}_\ell)$ (notation defined in Section 4.4). Magnification of the regions indicated by the dashed rectangles is shown in the subplots.

production optimization with SQP over the 20-year reservoir lifespan, starting with the initial guess \mathcal{S}_0 (as noted earlier, \mathcal{S}_0 is taken from the original Brugge test case description). For this computation, we use the true reservoir model; i.e., $\mathcal{R} = \mathcal{R}_{\text{true}}$. Figure 7 shows the comparison of NPV as a function of time for the initial guess \mathcal{S}_0 (thick dashed line), the optimal solution \mathcal{S}_{opt} (blue line), and the reference solution \mathcal{S}^* (red line). We would not expect there to be much difference between \mathcal{S}_{opt} and \mathcal{S}^* since the reservoir model is specified to be $\mathcal{R}_{\text{true}}$. This is in fact observed — the two NPVs after 20 years are $7.32 \cdot 10^9$ USD (\mathcal{S}_{opt}) and $7.29 \cdot 10^9$ USD (\mathcal{S}^*). The similarity between the two solutions is also evident in the BHP controls for the two cases, shown in Figures 8(c,d) and 8(g,h).

Note that, from the results in Figure 8, it is evident that many of the wells experience shut in over significant portions of the simulation period. This is due to backflow in the forward simulation, and not to our treatment of the nonlinear constraints (which can also lead to well shut in). By limiting the ranges of allowable BHPs for injection and production wells, we observed that nearly all of these shut ins can be eliminated. This treatment, however, leads to lower NPVs than are achieved when we use the BHP bounds given in Table 1.

We next generate a pool of reference solutions, designated \mathcal{S}_i^* , $i = 1, \dots, 15$, by performing production optimization starting with different initial guesses $\mathcal{S}_{0,i}$ (we again take $\mathcal{R} = \mathcal{R}_{\text{true}}$ and use SQP as the optimizer). Figure 7 shows the NPV trajectories computed for \mathcal{S}^{**} (in gray), which are seen to fall within a narrow range. Specifically, the NPVs at 20 years lie within the interval $[7.16 \cdot 10^9, 7.39 \cdot 10^9]$. From these results we can conclude that the difference between \mathcal{S}_{opt} and \mathcal{S}^{**} is essentially insignificant because the NPV for \mathcal{S}_{opt} is contained within the envelop of NPVs for \mathcal{S}^{**} .

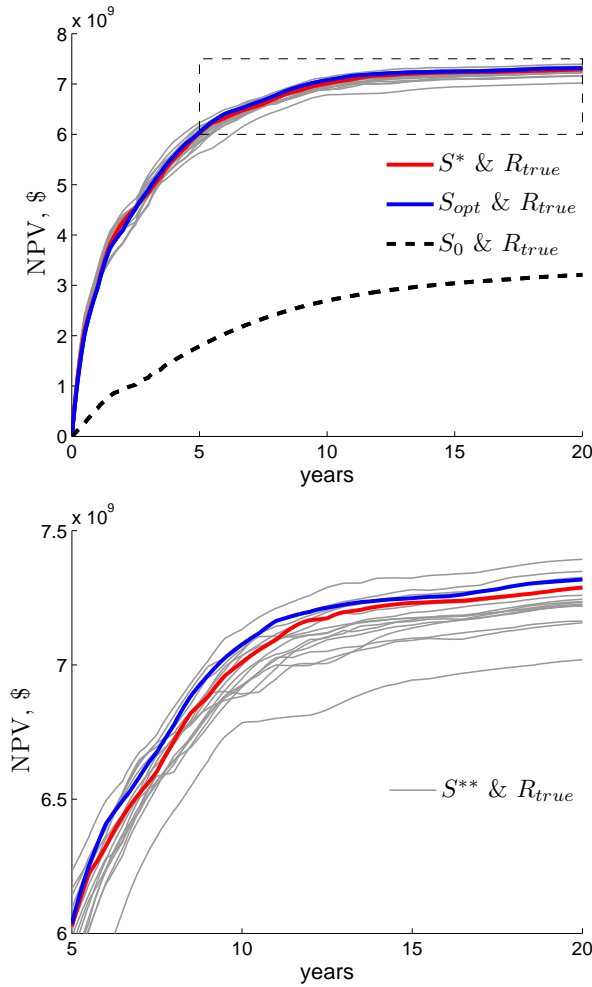


Fig. 7 (top) NPV as a function of time for the optimal solution \mathcal{S}_{opt} (blue line), reference solution \mathcal{S}^* (red line), initial guess solution \mathcal{S}_0 (dashed black line), and a set of reference solutions \mathcal{S}_i^{**} , $i = 1, \dots, 15$ (thin gray lines). All NPVs are computed with $\mathcal{R}_{\text{true}}$. (bottom) Magnification of the region indicated by the dashed rectangle showing the range of reference solutions \mathcal{S}_i^{**} for years 5 to 20.

The impact of production optimization can also be seen by considering the remaining oil at the end of the reservoir life. Figure 9 presents oil saturation distributions at 20 years for the initial controls \mathcal{S}_0 and optimal controls \mathcal{S}_{opt} and \mathcal{S}^* . It is apparent that \mathcal{S}_{opt} and \mathcal{S}^* lead to improved sweep and thus more oil production relative to \mathcal{S}_0 . This is particularly evident in layers 4 and 6.

4.6 Comparison of SQP and MMA

In Figure 6 we saw that the “convergence” of the model and reservoir NPVs occurred at smaller ℓ with MMA than SQP, and that MMA led to a slightly higher final NPV than SQP. These findings are consistent with the better overall performance observed for MMA. Additional comparisons for the first three closed-loop iterations are presented in Ta-

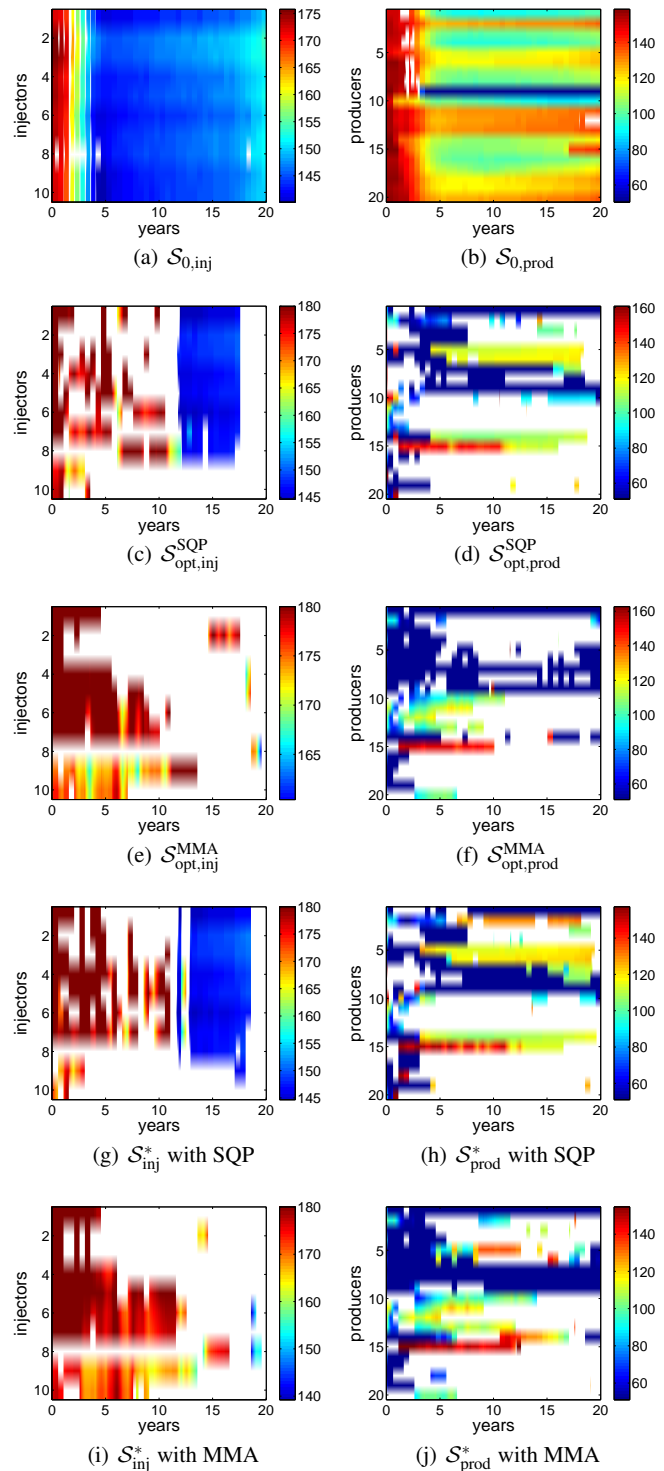


Fig. 8 BHP schedules for (a, c, e, g, i) 10 injection and (b, d, f, h, j) 20 production wells for five sets of controls: (a, b) \mathcal{S}_0 , (c, d) \mathcal{S}_{opt} with SQP, (e, f) \mathcal{S}_{opt} with MMA, (g, h) \mathcal{S}^* with SQP and (i, j) \mathcal{S}^* with MMA. White regions indicate well shut-in.

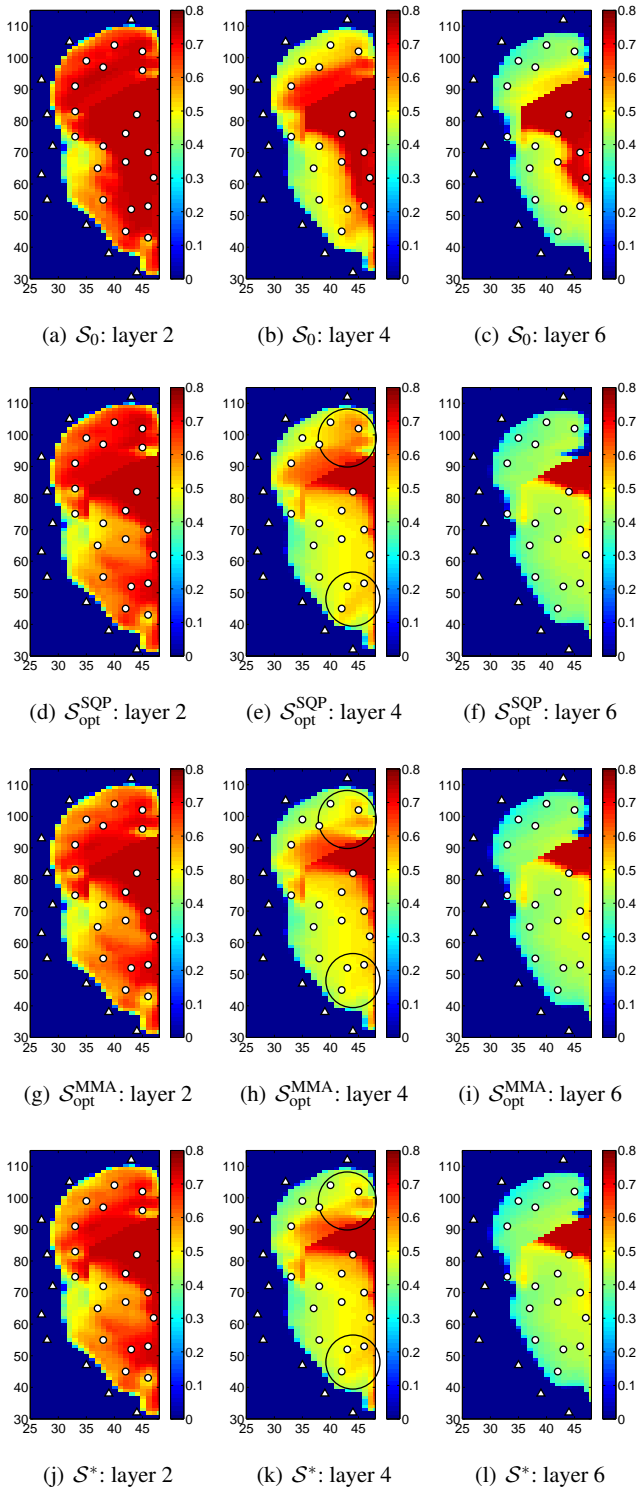


Fig. 9 Final oil saturation distribution in the vicinity of the wells (injectors indicated by white triangles, producers by white circles) in layers 2, 4 and 6 for four cases: (a, b, c) S_0 , (d, e, f) S_{opt} with SQP, (g, h, i) S_{opt} with MMA, and (j, k, l) S^* with MMA. \mathcal{R}_{true} is used in all cases. Black circles indicate regions of improved sweep efficiency relative to S_0 .

ble 2. Here, forward simulation and gradient evaluation refer to sequential steps within the “repeat” loop in Algorithm 1, where we solve (18) and (20)–(21). From the table we see that MMA requires fewer forward simulations and gradient evaluations (the number of gradient evaluations with MMA is equal to the number of forward simulations) for $\ell = 1, 2, 3$. This is likely due to the fact that, as discussed above, by adjusting the asymptotes, MMA is able to provide an accurate nonlinear approximation of the objective function even at the beginning of the optimization.

We note also that a slight improvement in sweep with MMA relative to SQP can be seen by comparing Figures 9(e) and (h) within the regions indicated by the black circles. The optimal BHP controls for the two cases differ, as is evident in Figure 8.

The relative performance of MMA and SQP cannot be judged based on a single test. Thus, we now perform multiple optimizations to assess the number of reservoir simulations required for convergence. MMA termination criteria are based on the relative change of the objective function and optimization variables, while the SQP criterion considers the reduction of the norm of the gradient. Table 3 compares the computational performance of SQP and MMA for 15 different optimization runs (in which S^{**} , described in Section 4.5, is determined). In this table, columns 3 and 6 provide the number of simulation runs needed for MMA and SQP to converge to the values shown in columns 2 and 5. The values in parentheses in column 4 specify the number of runs required for MMA to reach the same NPV value as SQP at the point of SQP termination. In 12 of the 15 cases, MMA achieves a higher NPV than SQP (the cases in which SQP reaches a higher NPV are indicated by asterisks in column 1). In addition, we see that MMA consistently requires fewer simulations.

The superior performance of MMA is further illustrated in Figure 10, where we compare the the norm $\frac{1}{N_{nb}} \|\nabla \mathcal{J}_{PO}\|_{1,nb}$ of the objective function gradient (20) obtained while computing the solutions S^{**} . This norm includes only the components corresponding to the N_{nb} control variables located away from the boundary of the feasibility region. It is evident that the MMA gradients decrease faster, and that there is less spread in the MMA results for the different runs. This suggests that MMA performance is less dependent than SQP on the initial guess.

Table 2 Comparison of the computational performance of SQP and MMA solvers for the production optimization problem over the first three CLRM iterations.

CLRM iteration, ℓ	SQP gradient evaluations	SQP forward simulations	MMA forward simulations
1	26	56	10
2	26	28	12
3	9	10	8

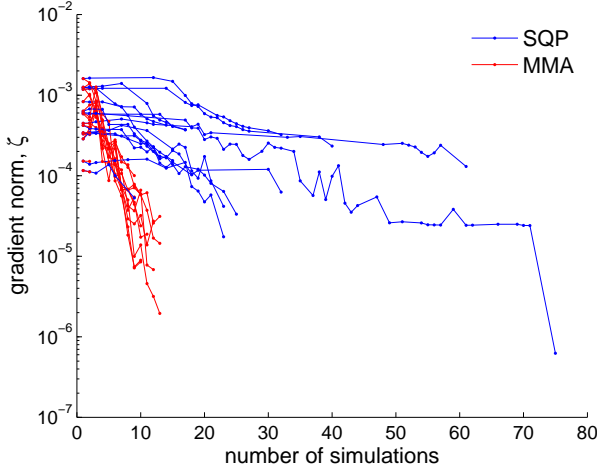


Fig. 10 Comparison of the reduction in the norm of the objective function gradient (20), $\zeta = \frac{1}{N_{nb}} \|\nabla \mathcal{J}_{PO}\|_{1,nb}$, using SQP (blue lines) and MMA (red lines) solvers for the computation of S_i^{**} , $i = 1, \dots, 15$. Points indicate the simulations used for gradient evaluations. $\|\cdot\|_{1,nb}$ includes only the components corresponding to the N_{nb} control variables located away from the boundary of the feasibility region.

Table 3 Comparison of the computational performance of MMA and SQP solvers for the production optimization problem. Performance is assessed for 15 different initial guesses, used to obtain solutions S_i^{**} , $i = 1, \dots, 15$. All NPVs are computed with \mathcal{R}_{true} .

i	MMA		SQP	
	NPV (10^9 \$)	simulations	NPV (10^9 \$)	simulations
1	7.41	13 (3)	7.15	15
2	7.49	12 (2)	7.25	17
3	7.43	10 (4)	7.39	32
4	7.35	9 (2)	7.22	17
5	7.35	9 (4)	7.32	61
6	7.38	13 (3)	7.16	32
7	7.42	10 (3)	7.21	40
8*	7.17	2	7.23	24
9	7.30	12 (4)	7.24	23
10*	7.17	2	7.23	13
11	7.16	10 (2)	7.01	14
12	7.23	11 (6)	7.22	25
13	7.37	10 (3)	7.28	18
14*	7.28	2	7.31	9
15	7.36	13 (7)	7.34	75
$\langle \text{NPV} \rangle$	7.33		7.24	

4.7 History matching performance

In order to assess the performance of the CLRM history matching, we compare the results obtained from the PO solution \mathcal{S}_{opt} for different reservoir properties \mathcal{R} . Figure 11 shows the comparison of NPV as a function of time for the initial guess \mathcal{R}_0 (thick dashed line), the optimal solution \mathcal{R}_{opt} (magenta line), and the true field \mathcal{R}_{true} (blue line). The NPV trajectories corresponding to the true and history

matched fields are nearly identical, which is indicative of the quality of \mathcal{R}_{opt} .

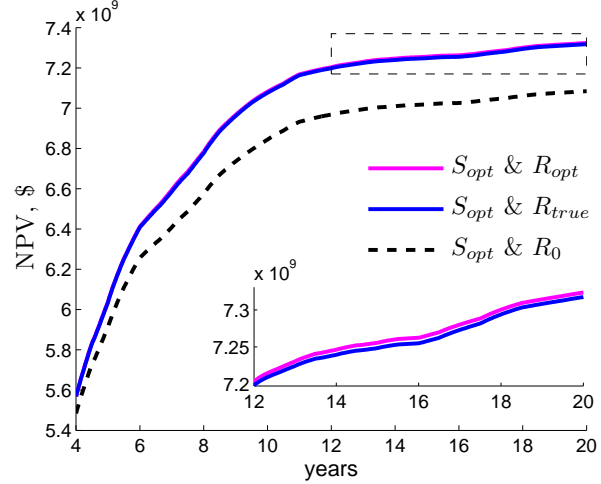


Fig. 11 Comparison of NPV as a function of time for the optimal solution \mathcal{R}_{opt} (magenta line), the true reservoir \mathcal{R}_{true} (blue line), and the initial guess \mathcal{R}_0 (dashed black line). All NPVs are computed using the same set of controls (\mathcal{S}_{opt}). Magnification of the region indicated by the dashed rectangle is shown in the subplot.

Figure 12 displays transmissibility fields in the x and z directions for the initial guess model (\mathcal{R}_0), the history match solution (\mathcal{R}_{opt}), and the true model (\mathcal{R}_{true}). Although the optimal and true solutions clearly differ, they do display general similarities for many of the large-scale features. It is important to note that \mathcal{R}_{opt} represents just one of many possible history matched models, so we cannot draw general conclusions from the fields in Figure 12. In addition, the history matching results are impacted by the weightings used for production and seismic data, and the specific effects of these weightings requires further study. In any event, based on these results and those in Figure 11, we can conclude that the CLRM history matching is performing as expected.

The results presented above were computed using production data that did not contain noise, along with proxy seismic data that were unfiltered and also without noise. We now introduce noise and filtering into the proxy seismic data. For these computations, we focus on the HM problem. Thus, all simulations involve the use of the fixed production schedule \mathcal{S}_{opt} , obtained from the full 20-year CLRM optimization, as discussed in Section 4.4. History matching uses only proxy seismic data generated yearly for the first 10 years, with different noise levels (η ranges from 0 to 50%) and application of the spatial filtering described in Section 3.3. Figure 13 shows the comparison of (signed) NPV error as a function of noise level η . The error ε is computed as the difference between the NPV for the particular case (at 20 years) and the NPV obtained with \mathcal{R}_{true} . As expected, for small noise values, ε is small, which suggests the

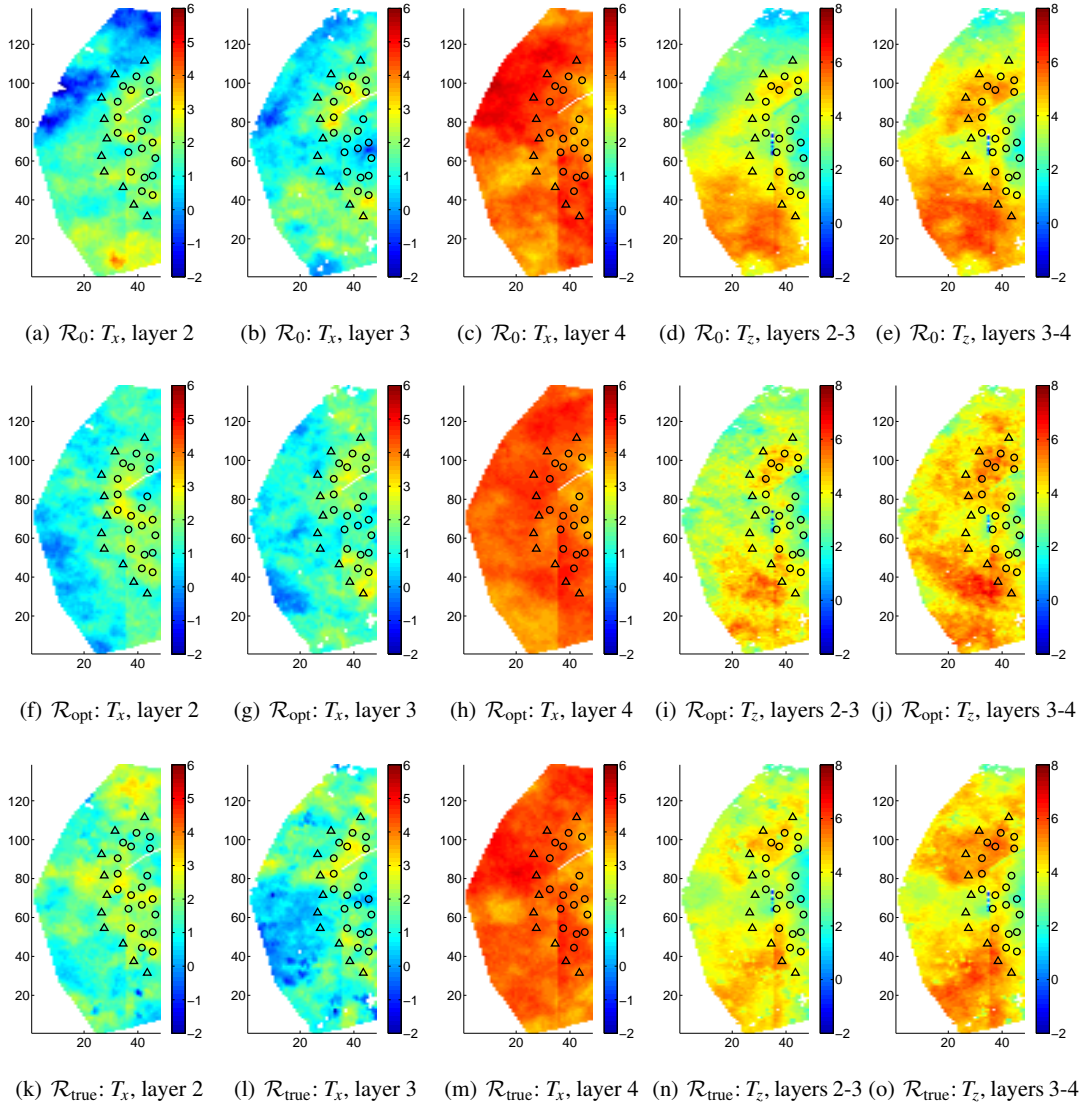


Fig. 12 (a-e) Initial \mathcal{R}_0 , (f-j) optimized (history matched) \mathcal{R}_{opt} , and (k-o) true $\mathcal{R}_{\text{true}}$ interblock x -transmissibility T_x and z -transmissibility T_z (log-scale) fields for layers 2, 3 and 4. Triangles and circles indicate injectors and producers, respectively.

history matched model is accurate (the 1% error interval is shown in gray in Figure 13). As η increases beyond about 25%, however, ε grows significantly, approaching 10% error in NPV for $\eta = 50\%$.

We now consider the same history matching problem (again using 10 years of data) for cases where only production data, only proxy seismic data, or both production and seismic data, are available. Production data contain noise, with reference to (44), defined by $\gamma_q = 0.05$ (5% of flow rates), $\sigma_{q,\min} = 2.0$, and $\sigma_{q,\max} = 25.0$. Noise in the proxy seismic data is as described above, with $\eta = 30\%$. Production and seismic data, when used, are available every month and every 5 years, respectively.

Figure 14 shows history matching results for nine different initial guesses for the three types of data. Error is

computed as described above. Interestingly, the use of only production data tends to underestimate NPV, while the use of only seismic data tends to overestimate NPV. The average error from using only seismic data is clearly less than that using only production data. This may be due in part to the relative noise levels in the two types of data, and/or to the fact that the seismic data is spatially dense. The error using seismic data here is less than that in Figure 13 with $\eta = 30\%$. This may be because a different noise realization was used in this case. Use of both types of data leads to lower accuracy (on average) than does the use of seismic data only, though errors are quite small in either case. The error is, however, less biased when both data types are used. Additional examples will need to be studied to identify the impact of various data types and noise levels on the HM so-

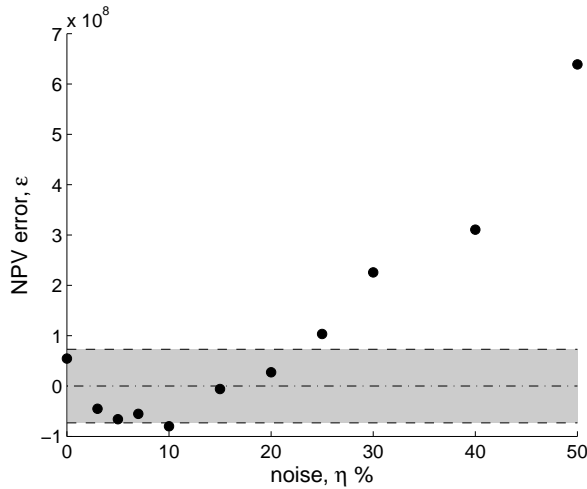


Fig. 13 Comparison of NPV error as a function of noise level η for different HM solutions \mathcal{R}_{opt} , obtained for cases when history matching uses only proxy seismic data (with noise and filtering) generated yearly for the first 10 years. Gray region indicates the 1% error interval.

lution. In future work, we plan to address this issue using a multiobjective optimization framework.

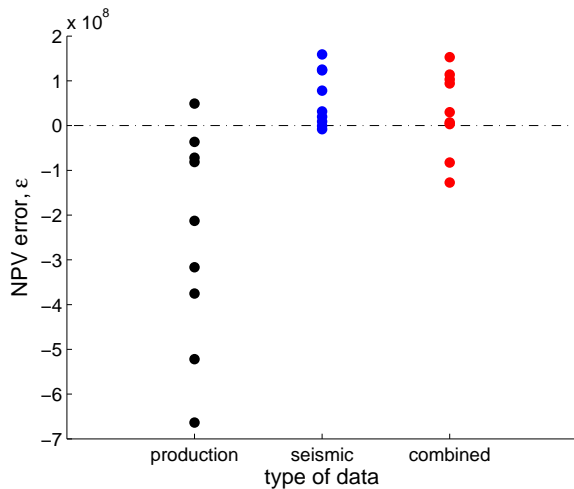


Fig. 14 Error in NPV computed for different HM solutions \mathcal{R}_{opt} obtained using only production data (black points), only proxy seismic data (blue points), and both production and seismic data (red points), for nine different initial guesses.

5 Concluding remarks

In this work, we presented an efficient adjoint-based computational approach for CLRM. The AD framework, built around the reservoir simulator AD-GPRS and combined with dedicated SCP solvers, enables a highly flexible implementation. In the history matching component of CLRM, the SQP method is applied to a problem parameterized using

PCA. Both production data and proxy seismic data are considered. In the production optimization component, a novel (in this context) SCP solver is employed. This solver is suitable for optimization problems where the objective function and PDE-constraints are quasi-linear and separable with respect to the optimization variables.

We investigated the performance of the closed-loop methodology for the well-documented Brugge model. For the production optimization problem, we showed that MMA outperformed SQP, particularly at early iterations of the closed-loop procedure. For history matching, by contrast, the commonly used SQP procedure is the method of choice. The use of synthetic 4D seismic data in the gradient-based history matching procedure was also demonstrated. This involved the use of interpreted data, which provided an estimated saturation field that was incorporated into the objective function. The effect on history matching of noise and spatial filtering in the proxy seismic data was assessed. The impact of the different data types on history matching results was also considered.

CLRM results for Brugge model NPV were shown to be in close agreement with those achieved for the “true” (synthetic) reservoir model. The detailed well settings differed between the true and optimized models, which reflects the nonuniqueness of the CLRM solution. The high level of agreement in NPV between the two solutions is likely due, in this case, to the characteristics of the 104 Brugge realizations and to our use of “true” seismic data. In any event, our overall results demonstrate that the components of the AD-based CLRM are functioning properly.

There are many ways in which the adjoint gradient-based CLRM can be tested and extended. The relative performance of SQP and MMA for production optimization should be assessed for other problems. Given that we are using data of different types, it will be of interest to apply multiobjective optimization techniques, in which the Pareto front is computed. We also plan to investigate the use of other control variable re-parameterization procedures. These could include the dynamical grouping of gradient components, according to theoretical or heuristic observations, for production optimization, and the use of other geological parameterizations, such as optimization-based PCA [42], for history matching. The application of CLRM for synthetic and real field models that involve more complicated physics, as well as the incorporation of robust optimization procedures for optimization under uncertainty, will also be addressed.

Acknowledgements We are grateful to the industrial affiliates of the Stanford University Reservoir Simulation Research (SUPRI-B) and Smart Fields Consortia for financial support. We thank Michael Saunders, Tapan Mukerji and Hai Vo (Stanford University) for useful discussions and suggestions, and David Echeverría Ciaurri (IBM Research) for assistance with the Brugge model.

References

1. Aziz, K., Settari, T.: *Petroleum Reservoir Simulation*. Applied Science Publishers (1979)
2. Brouwer, D., Jansen, J.D.: Dynamic optimization of waterflooding with smart wells using optimal control theory, SPE78278-PA. *SPE Journal* **9**(4), 391–402 (2004)
3. Brouwer, D., Nævdal, G., Jansen, J.D., Vefring, E., Van Kruijsdijk, C.: Improved reservoir management through optimal control and continuous model updating, SPE90149-MS (2004)
4. Bukshtynov, V., Protas, B.: Optimal reconstruction of material properties in complex multiphysics phenomena. *Journal of Computational Physics* **242**, 889–914 (2013)
5. Bukshtynov, V., Volkov, O., Protas, B.: On optimal reconstruction of constitutive relations. *Physica D: Nonlinear Phenomena* **240**(16), 1228–1244 (2011)
6. Cattell, R.B.: The scree test for the number of factors. *Multivariate Behavioral Research* **1**(2), 245–276 (1966)
7. Chen, Y., Oliver, D., Zhang, D.: Efficient ensemble-based closed-loop production optimization, SPE112873-PA. *SPE Journal* **14**(4), 634–645 (2009)
8. Chen, Y., Ollver, D.: Ensemble-based closed-loop optimization applied to Brugge field, SPE118926-PA. *SPE Reservoir Evaluation and Engineering* **13**(1), 56–71 (2010)
9. Dong, Y., Gu, Y., Oliver, D.: Sequential assimilation of 4D seismic data for reservoir description using the ensemble Kalman filter. *Journal of Petroleum Science and Engineering* **53**(1–2), 83–99 (2006)
10. Dong, Y., Oliver, D.: Quantitative use of 4D seismic data for reservoir description, SPE84571-PA. *SPE Journal* **10**(1), 91–99 (2005)
11. Engl, H., Hanke, M., Neubauer, A.: *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer (1996)
12. Feng, T., Mannseth, T.: Impact of time-lapse seismic data for permeability estimation. *Computational Geosciences* **14**(4), 705–719 (2010)
13. Fichtner, A.: *Full Seismic Waveform Modelling and Inversion*. Springer (2011)
14. Fleury, C.: First and second order convex approximation strategies in structural optimization. *Structural Optimization* **1**(1), 3–10 (1989)
15. Foss, B., Jensen, J.: Performance analysis for closed-loop reservoir management, SPE138891-PA. *SPE Journal* **16**(1), 183–190 (2011)
16. Gill, P., Murray, W., Saunders, M.: *User's Guide for SNOPT Version 7: Software for Large-Scale Nonlinear Programming*. Stanford University (2008)
17. Gonzalez, R., Woods, R.: *Digital Image Processing*, 3rd edn. Prentice Hall (2007)
18. Huang, X., Meister, L., Workman, R.: Reservoir characterization by integration of time-lapse seismic and production data, SPE38695-MS (1997)
19. Jansen, J., Douma, S., Brouwer, D., Van Den Hof, P., Bosgra, O., Heemink, A.: Closed-loop reservoir management, SPE119098-MS. pp. 856–873 (2009)
20. Jansen, J.D.: Adjoint-based optimization of multi-phase flow through porous media - a review. *Computers and Fluids* **46**(1), 40–51 (2011)
21. Jansen, J.D., Bosgra, O., Van den Hof, P.: Model-based control of multiphase flow in subsurface oil reservoirs. *Journal of Process Control* **18**(9), 846–855 (2008)
22. Johnson, S.G.: *The NLOpt nonlinear-optimization package* (2013). URL <http://ab-initio.mit.edu/nlopt>
23. Kaiser, H.F.: The application of electronic computers to factor analysis. *Educational and Psychological Measurement* **20**(1), 141–151 (1960)
24. Kourounis, D., Durlafsky, L., Jansen, J.D., Aziz, K.: Adjoint formulation and constraint handling for gradient-based optimization of compositional reservoir flow. *Computational Geosciences* **18**(2), 117–137 (2014)
25. Li, R., Reynolds, A., Oliver, D.: History matching of three-phase flow production data. *SPE Journal* **8**(4), 328–340 (2003)
26. Luenberger, D.: *Optimization by Vector Space Methods*. Series in Decision and Control. Wiley (1976)
27. Nævdal, G., Brouwer, D., Jansen, J.D.: Waterflooding using closed-loop control. *Computational Geosciences* **10**(1), 37–60 (2006)
28. Nocedal, J., Wright, S.J.: *Numerical Optimization*, 2nd edn. Springer, New York (2006)
29. Oliver, D., Chen, Y.: Recent progress on reservoir history matching: A review. *Computational Geosciences* **15**(1), 185–221 (2011)
30. Oliver, D.S., Reynolds, A.C., Liu, N.: *Inverse Theory for Petroleum Reservoir Characterization and History Matching*. Cambridge University Press (2008)
31. Peters, E., 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 optimization and history matching, SPE119094-PA. *SPE Reservoir Evaluation and Engineering* **13**(3), 391–405 (2010)
32. Peters, E., Chen, Y., Leeuwenburgh, O., Oliver, D.: Extended Brugge benchmark case for history matching and water flooding optimization. *Computers and Geosciences* **50**, 16–24 (2013)
33. Petrou, M., Petrou, C.: *Image Processing: The Fundamentals*, 2nd edn. Wiley (2010)
34. Sarma, P.: *Efficient Closed-Loop Optimal Control of Petroleum Reservoirs under Uncertainty*. PhD Thesis, Stanford University (2006)
35. Sarma, P., Chen, W., Durlafsky, L., Aziz, K.: Production optimization with adjoint models under nonlinear control-state path inequality constraints, SPE99959-PA. *SPE Reservoir Evaluation and Engineering* **11**(2), 326–339 (2008)
36. Sarma, P., Durlafsky, L., Aziz, K.: Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics. *Mathematical Geosciences* **40**(1), 3–32 (2008)
37. Sarma, P., Durlafsky, L., Aziz, K., Chen, W.: Efficient real-time reservoir management using adjoint-based optimal control and model updating. *Computational Geosciences* **10**(1), 3–36 (2006)
38. Skjervheim, J.A., Evensen, G., Aanonsen, S., Ruud, B., Johansen, T.: Incorporating 4D seismic data in reservoir simulation models using ensemble Kalman filter, SPE95789-PA. *SPE Journal* **12**(3), 282–292 (2007)
39. Svanberg, K.: The method of moving asymptotes - a new method for structural optimization. *International Journal for Numerical Methods in Engineering* **24**(2), 359–373 (1987)
40. Tarantola, A.: *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM (2005)
41. Vasco, D., Datta-Gupta, A., Behrens, R., Condon, P., Rickett, J.: Seismic imaging of reservoir flow properties: Time-lapse amplitude changes. *Geophysics* **69**(6), 1425–1442 (2004)
42. Vo, H., Durlafsky, L.: A new differentiable parameterization based on principal component analysis for the low-dimensional representation of complex geological models. *Mathematical Geosciences* **46**(7), 775–813 (2014)
43. Volkov, O., Voskov, D.: Advanced strategies of forward simulation for adjoint-based optimization, SPE163592-MS (2013)
44. Wang, C., Li, G., Reynolds, A.: Production optimization in closed-loop reservoir management, SPE109805-PA. *SPE Journal* **14**(3), 506–523 (2009)
45. Younis, R.: *Modern Advances in Software and Solution Algorithms for Reservoir Simulation*. PhD Thesis, Stanford University (2011)
46. Younis, R., Aziz, K.: Parallel automatically differentiable datatypes for next-generation simulator development, SPE106493-MS (2007)

-
47. Zhou, Y.: Parallel General-Purpose Reservoir Simulation with Coupled Reservoir Models and Multi-Segment Wells. PhD Thesis, Stanford University (2012)
 48. Zhou, Y., Tchelepi, H., Mallison, B.: Automatic differentiation framework for compositional simulation on unstructured grids with multi-point discretization schemes, SPE141592-MS (2011)
 49. Zilber, C.: A globally convergent version of the method of moving asymptotes. *Structural Optimization* **6**(3), 166–174 (1993)