Comput Geosci (2013) 17:117–138
DOI 10.1007/s10596-012-9320-4

ORIGINAL PAPER

Large-scale history matching with quadratic interpolation models

Hui Zhao · Gaoming Li · Albert C. Reynolds · Jun Yao

Received: 13 June 2011 / Accepted: 17 September 2012 / Published online: 19 October 2012

Abstract Due to the limited availability of adjoint code in commercial reservoir simulators for gradient calculations, there is a need to explore the applicability of derivative-free optimization algorithms for large-scale history matching. This paper tests the utility of three derivative-free optimization algorithms (stochastic Gaussian search direction (SGSD), new unconstrained optimization algorithm (NEWUOA), and quadratic interpolation model-based algorithm guided by approximate gradient (QIM-AG)) for history matching. The SGSD method uses a negative stochastic gradient which is obtained by simultaneously perturbing all the model parameters using a Gaussian random vector. For a continuous objective function and a sufficiently small perturbation size, the stochastic gradient is always uphill and the expectation of the stochastic gradient converges to the true gradient as the perturbation size goes to zero. NEWUOA is a quadratic interpolation model-based optimization algorithm. At each iteration, the objective function is first approximated by a quadratic interpolation model. The quadratic model is then minimized to obtain an updated reservoir description for the next iteration. The number of interpolation points (reservoir simulation runs) required by NEWUOA must be larger than the dimension of reservoir model parameter space in order to construct the initial quadratic model. QIM-AG reduces the required number of interpolation points by replacing the first-order coefficients that appear in the quadratic model by an approximate gradient. The approximate gradient used in this study is an average of several stochastic gradients from SGSD. To reduce the dimension of the optimization problem, a simple parameterization method based on the prior covariance matrix is applied. The prior covariance matrix is approximated using an ensemble of unconditional realizations. The parameterization avoids the calculation of the inverse of the prior covariance matrix during optimization and may further regularize the ill-posed inverse problem.

Keywords SGSD · NEWUOA · Parameterization · History matching · Uncertainty quantification

1 Introduction

In the Bayesian framework, the reservoir history-matching problem is equivalent to the minimization of an objective function which includes the mismatch between the predicted and observed data and the mismatch between the current reservoir model and the prior mean reservoir model. This objective function can be minimized with gradient-based or derivative-free optimization algorithms. Gradient-based optimization algorithms [2, 6, 7, 9, 10, 14, 23, 27, 30, 32, 42, 54, 63, 64, 68, 72, 74, 76] appear to be the most efficient

H. Zhao
Yangtze University, Jingzhou,
Hubei 434023, China

G. Li (✉) · A. C. Reynolds
University of Tulsa, 800 S. Tucker Dr. Tulsa,
OK 74104, USA
e-mail: gaoming-li@utulsa.edu

J. Yao
China University of Petroleum, Qingdao,
Shandong 266555, China

approach for reservoir history-matching problems as well as for the associated production optimization step of the closed-loop reservoir management [8, 66, 73, 79]. The required gradient may be calculated efficiently using the adjoint method based on the formulation developed by [43], [64], or [41]. However, the adjoint method requires specific knowledge of the mathematical details and numerics of reservoir simulators, and commercial reservoir simulators in most cases do not provide all gradients that are needed. A cogent discussion of gradient-based and other methods for history matching can be found in a recent review paper [55].

Derivative-free optimization algorithms [18, 40] require only the evaluation of the objective function, not the gradient, and hence the reservoir simulator may be treated as a black box. These algorithms may be divided into several categories. One category includes methods where the search direction is calculated based on an approximate gradient such as simultaneous perturbation stochastic approximation (SPSA) [4, 29, 75] or ensemble-based optimization [12, 13, 75]. In another category, the objective function is approximated with some polynomial function, which is then minimized with a gradient-based method [16, 18, 60]. Heuristic random search algorithms such as simulated annealing [5, 20], genetic algorithm [22, 65], particle swarm optimization (PSO) [37, 47, 56], ant colony optimization [31], and neighborhood algorithm [15] comprise a third category. This class of derivative-free algorithms uses heuristic random searches and may find global optima but typically have very slow convergence rates and require such a large number of function evaluations that they are feasible only when the number of variables is small. The first two categories of the derivative-free methods mimic the gradient-based optimization with local searches hence have better convergence rate compared to heuristic random search algorithms. Another class of derivative-free optimization algorithms is directional-direct search methods [18, 19, 39] which are popular in the optimization community as it can be proved that they converge to a local optimum under reasonable assumptions. Our experience with these methods is minimal, and to the best of our knowledge, these methods have not been tested for computational efficiency in history-matching problems. A survey of different derivative-free optimization methods for general optimization problems can be found in Conn et al. [17]. Zhao et al. [79] compared the performance of different derivative-free methods for reservoir production optimization. This paper will investigate three derivative-free algorithms: stochastic Gaussian search direction (SGSD) [45], new unconstrained optimization algorithm (NEWUOA) [60, 61], and quadratic interpolation model-based algorithm guided by an approximate gradient (QIM-AG) [79].

The SGSD algorithm has its root in the SPSA algorithm [69–71]. SPSA as a derivative-free optimization algorithm was first applied to history matching by Gao et al. [29]. In the first-order SPSA algorithm, the search direction is the negative SPSA gradient, which is obtained by simultaneously perturbing all the reservoir model parameters using a random vector. The elements of the random vector are usually chosen as samples from the symmetric ± 1 Bernoulli distribution. Although the SPSA gradient is stochastic, its expectation is approximately the true gradient and it is always an uphill direction in the vicinity of the current iterate [29]. To improve the performance of the first-order SPSA, Gao et al. [29] tested two second-order SPSA algorithms. The first one applies a stochastic Hessian inverse matrix, which is obtained using simultaneous perturbation, and the second one uses the prior covariance matrix as the approximate inverse Hessian matrix. The latter method is referred to as the simple second-order SPSA. Examples presented in Gao et al. [29] showed that the simple second-order SPSA performs better than all other SPSA algorithms. Li and Reynolds [45] proposed a new SGSD algorithm, which uses a negative stochastic gradient as its search direction. The stochastic gradient in the SGSD method is obtained by simultaneously perturbing the model parameters using unconditional realizations sampled from the prior distribution. The expectation of this stochastic gradient is the prior covariance matrix times the true gradient, so the search direction in the SGSD method is similar to the simple second-order SPSA algorithm. However, the examples show that SGSD algorithm is more reliable and efficient than the simple second-order SPSA algorithm in obtaining a set of conditional reservoir models using the randomized maximum likelihood (RML) method [38, 52, 54].

NEWUOA is an efficient and robust quadratic model-based derivative-free algorithm developed by Powell [60] for unconstrained optimization problems. Its companion algorithm for bound constrained problem is bound-constrained optimization by quadratic approximation (BOBYQA) [62]. As a good quadratic model can capture the curvature of the objective function, reasonably fast convergence may be achieved [46, 50, 58] once the initial quadratic model is built. The coefficients of this initial quadratic model are determined using the interpolation condition, i.e., the quadratic model is made equal to the objective function at some interpolation points. When the number of variables n is large, it is not feasible to generate enough interpolation points ($[(n+1)(n+2)]/2$) to fully deter-

mine a unique quadratic model. Thus, the number of interpolation points used to build this initial quadratic model is usually far smaller than the total number of coefficients in the quadratic model, in which case the coefficients in the constant and gradient part of the quadratic model are fully determined and only some of the coefficients in the Hessian matrix are calculated while others are equal to zero. During iteration, NEWUOA updates the quadratic model by determining the coefficients in the change of the quadratic model. These coefficients are calculated by minimizing the Frobenius norm of the change in the Hessian matrix subject to the constraints that the quadratic model satisfies the interpolation conditions [59]. The construction and update of the quadratic model in NEWUOA is similar to that of the Broyden–Fletcher–Goldfarb–Shanno quasi-Newton method [49], which obtains the updated Hessian matrix by minimizing the change in the Hessian matrix between iterations subject to the secant equation. In NEWUOA, the quadratic model is minimized using a trust-region method to obtain an updated reservoir model at each iteration. Another comparable derivative-free optimization (DFO) package developed by Conn et al. [16] also uses the quadratic interpolation model. In DFO, the basis functions for the quadratic model are Newton polynomials while NEWUOA uses Lagrangian polynomials. The DFO package chooses a "well-poised" set of interpolation points close to the current optimal estimate in building the quadratic interpolation model to avoid numerical difficulties in optimization, while NEWUOA does not explicitly check the poisedness of the interpolation points used to build the quadratic model. Although NEWUOA does not directly check the interpolation set to ensure it is poised, NEWUOA does contain a procedure for changing the interpolation set which aims to improve the position of interpolation set and maintain the quality of the quadratic model. Fasano et al. results suggest it may be possible to skip checks on the quality of the interpolation model. The Fasano et al. [24] results, however, are based on computational experiments and are not supported by rigorous theory. Conn et al. [18] suggest that DFO should be used for small-scale optimization problems with less than 50 parameters.

The major challenge in applying NEWUOA and BOBYQA to a large-scale optimization problem is that the methods require a large number of interpolation points (this number must be larger than the dimension of the optimization problem) to build the initial quadratic model. Note that the computational time required may be significantly reduced with parallel computing in a cluster environment. For computing environments without computer-cluster access, Zhao et al. [79] proposed a QIM-AG to reduce the number of interpolation points required for building initial quadratic model. In the QIM-AG algorithm, the objective function is approximated by a quadratic model as in NEWUOA. The gradient part of the quadratic model at the current iterate is set equal to an approximate gradient, which can be either a stochastic gradient as in SGSD or a simplex gradient. The Hessian matrix of the quadratic model is determined by minimizing the Frobenius norm of the Hessian matrix subject to the interpolation condition at some points as in NEWUOA. The bound constraints in the reservoir model parameters are dealt with by simple truncation as discussed in Zhao et al. [79], i.e., when a parameter is outside of the bounds, it is simply set equal to the value of its nearest bound.

In a history-matching problem that adjusts the reservoir gridblock properties, the number of parameters can be quite large. To reduce the computational cost and ill-posedness in this inversion problem, one often reduces the number of parameters through some parameterization method. A natural way to parameterize a rock property field is to divide the reservoir into zones and assume uniform properties in each zone which may contain several reservoir gridblocks. Zonation was introduced to the history-matching literature by [33]. The gradzone procedure of [6] is similar to zonation. Although a property field does not have to be constant within a gradzone, the property at all gridblocks within a specific gradzone is multiplied by the same constant during history matching. Although the zonation and gradzone methods are simple, they can introduce nonphysical discontinuities in a rock property field across the boundary of a zone or gradzone. Parameterization methods using basis vectors avoid these discontinuities. When using basis vectors, the original vector of reservoir parameters is represented as a linear combination of basis vectors and the new parameters are the coefficients in this linear combination. The subspace method [2, 51, 63] uses gradients of subgroups of data mismatch term and gradients of the model mismatch term in the objective function as the basis vectors. Tavakoli and Reynolds [72] argued that the principal right singular vectors of the dimensionless sensitivity matrix form the optimal basis for parameter reduction [72]. This parameter reduction method was also used in Rodrigues [64]. As the basis vectors in these methods use gradient information, they cannot be used in derivative-free methods. Assuming that a good prior geostatistical knowledge of the reservoir is available and reservoir models are Gaussian, one may use the eigenvectors corresponding to the largest eigenvalues

of covariance matrix as the basis vectors. The method is known as linear principal component analysis (PCA) or Karhunen–Loeve decomposition, which has been applied with gradient-based history matching [30, 53, 63], with the ensemble Kalman filter (EnKF) [78] and with PSO [47]. For non-Gaussian models (such as facies reservoir models or reservoirs with fluvial channels), one may use nonlinear kernel PCA (KPCA), which preserves the high-order non-Gaussian moments [34, 67]. As in the linear PCA, KPCA relies on good prior knowledge of the reservoir. When this prior geological information is not available, one may apply the discrete cosine transform [34–36], which constructs basis vectors from cosine functions and does not explicitly require prior geological knowledge, although the prior geological knowledge may lead to better reservoir parameter estimation.

In this paper, we always assume a Gaussian prior model and apply a simple parameterization based on the prior covariance matrix, which is approximated using a set of unconditional reservoir models. This parameterization is effectively the same as the one used in the ensemble Kalman filter where any updated vector of model parameters is a linear combination of the initial ensemble of vectors of model parameters [1, 44]. The optimization is done in the parameterized space, which has a much smaller dimension. The three optimization algorithms—SGSD, QIM-AG, and NEWUOA—are compared through three synthetic reservoir cases including the Brugge test reservoir [57].

2 Model parameterization

Based on the Bayesian statistical framework, the posterior probability density function (pdf) conditioned to the observation data (d_{obs}) is given by

$$f(m|d_{\text{obs}}) = a\exp(-O(m)), \tag{1}$$

where $O(m)$ is the objective function,

$$O(m) = \frac{1}{2}(m - m_{\text{prior}})^T C_M^{-1}(m - m_{\text{prior}}) + \frac{1}{2}(g(m) - d_{\text{obs}})^T C_D^{-1}(g(m) - d_{\text{obs}}). \tag{2}$$

In Eq. 2, m is an N_m-dimensional column vector, which includes all the reservoir model parameters. The prior model for m is Gaussian with mean m_{prior} and covariance matrix C_M; d_{obs} is an N_d-dimensional observation data vector and the measurement errors for d_{obs} are assumed to be Gaussian with zero mean and covariance matrix C_D; $g(m)$ is the corresponding predicted data vector obtained by running the reservoir simulator. The maximum a posteriori (MAP) estimate is obtained by minimizing the objective function of Eq. 2.

For any realistic history-matching problem where we wish to estimate gridblock rock properties, the dimension N_m of the model parameter vector m is large, on the order of 10^6 or more, which makes the history-matching problem very ill-conditioned for typical covariance matrices. Because of this ill conditioning, one is often forced to damp the changes in the objective function at early iterations of any gradient-based optimization algorithm by artificially increasing the data measurement error variance (see, for example, Gao et al. [27] and Tavakoli and Reynolds [72]. Here, we apply a simple model parameterization based on the prior model to improve the conditioning of the inverse problem.

We first generate a set of unconditional reservoir models sampled from the prior Gaussian distribution, i.e., $m_j, \quad j = 1, \cdots, N_e$. Note that $N_e \ll N_m$ for the large-scale history-matching problems. The mean reservoir model is

$$\bar{m} = \frac{1}{N_e}\sum_{j=1}^{N_e} m_j. \tag{3}$$

The prior covariance matrix can be approximated by a set of realizations, i.e.,

$$C_M \approx \frac{1}{N_e - 1}\sum_{j=1}^{N_e}(m_j - \bar{m})(m_j - \bar{m})^T = \frac{1}{N_e - 1}\Delta M \Delta M^T, \tag{4}$$

where ΔM is an $N_m \times N_e$ matrix with its jth column equal to $(m_j - \bar{m})$. Applying the singular value decomposition to ΔM yields

$$\Delta M = U \Lambda V^T, \tag{5}$$

where U and V consist of the left and right singular vectors and Λ contains the singular values on its diagonal. Note that there are at most $N_p \le N_e - 1$ nonzero singular values because the matrix on the right-hand side of Eq. 4 has a rank less than or equal to $N_e - 1$. Substituting Eq. 5 into Eq. 4 yields

$$C_M \approx \frac{1}{N_e - 1} U \Lambda V^T V \Lambda^T U^T = \frac{1}{N_e - 1} U_p \Lambda_p^2 U_p^T, \tag{6}$$

where the diagonal matrix Λ_p contains the N_p nonzero singular values and the columns of U_p are the corresponding left singular vectors. The pseudo-inverse of C_M is

$$C_M^{\dagger} = (N_e - 1)U_p \Lambda_p^{-2} U_p^T. \tag{7}$$

We now parameterize the model vector in terms of the vector p, which is defined by

$$m = \bar{m} + \frac{U_p \Lambda_p}{\sqrt{N_e - 1}} p. \tag{8}$$

Note that Eq. 8 simply indicates that we restrict the model space so that $m - \bar{m}$ is a linear combination of the columns of U_p. If p is a random vector with mean zero and $N_p \times N_p$ identity covariance matrix I_{N_p}, m defined by Eq. 8 represents an approximate sample from $N(\bar{m}, C_M)$. Replacing C_M^{-1} with the pseudo-inverse $C_M^{\dagger}$ and m_{prior} with $\bar{m}$ in Eq. 2 yields

$$O(p) = \frac{1}{2} p^T p + \frac{1}{2}\big(g(m(p)) - d_{\text{obs}}\big)^T C_D^{-1}\big(g(m(p)) - d_{\text{obs}}\big). \tag{9}$$

Note that $\bar{m} = m_{\text{prior}}$ and $C_M^{\dagger} = C_M^{-1}$ are only true when $N_e \to \infty$.

During optimization, we adjust the parameter vector p to minimize the objective function $O(p)$ for the MAP estimate. Normally, one wishes to quantify the uncertainty in the reservoir model and in the reservoir performance predictions. One useful method for uncertainty quantification is the RML method [52]. With the parameterization method, a set of conditional realizations, which provide an approximate sampling of the pdf in Eq. 1, can be obtained using the RML method by minimizing the following objective function,

$$O_r(p) = \frac{1}{2}(p - p_{\text{uc}})^T (p - p_{\text{uc}}) + \frac{1}{2}\big(g(m(p)) - d_{\text{uc}}\big)^T C_D^{-1}\big(g(m(p)) - d_{\text{uc}}\big), \tag{10}$$

where p_{uc} is an unconditional sample from $N(0, I_{N_p})$ and d_{uc} is a perturbed observation vector obtained by sampling the Gaussian distribution, $N(d_{\text{obs}}, C_D)$.

In the examples considered here, we use $p = 0$, which corresponds to the prior mean model as the initial guess for the MAP estimate. For uncertainty quantification using the RML method, we use $p = p_{\text{uc}}$ as the initial guess.

3 SGSD algorithm

The SGSD algorithm iteratively minimizes the objective function of Eq. 9 to generate a MAP estimate and Eq. 10 to generate a conditional realization. The reservoir model update equation for iteration $\ell + 1$ is

$$p^{\ell+1} = p^{\ell} + a^{\ell} \frac{d^{\ell}}{\| d^{\ell} \|_{\infty}}, \tag{11}$$

where d^{ℓ} is the stochastic search direction and $d^{\ell} = -\hat{g}^{\ell}$. Here, $\hat{g}^{\ell}(p^{\ell})$ is a stochastic gradient obtained with simultaneous Gaussian perturbation as

$$\hat{g}^{\ell}(p^{\ell}) = \frac{O(p^{\ell} + \epsilon^{\ell} z^{\ell}) - O(p^{\ell})}{\epsilon^{\ell}} z^{\ell}, \tag{12}$$

where the elements of the perturbation vector z^{ℓ} are Gaussian random standard normal deviates and ϵ^{ℓ} is the perturbation size. It is shown in Li and Reynolds [45] that this stochastic gradient is an uphill direction in the vicinity of p^{ℓ} and the expectation of $\hat{g}^{\ell}$ is the true gradient of the objective function as the perturbation size ϵ^{ℓ} goes to zero. Due to the latter property of this stochastic gradient, the search direction is defined as the average of several stochastic gradients, i.e., $d^{\ell} = -\overline{\hat{g}^{\ell}}$ where

$$\overline{\hat{g}^{\ell}} = \frac{1}{N_g} \sum_{i=1}^{N_g} \hat{g}_i^{\ell}\left(p^{\ell}\right). \tag{13}$$

We would expect that averaging a larger number of SPSA gradients will yield an improved approximation of the true gradient. However, experiments in [21] indicate that using the average of five to ten SPSA gradients to compute the search direction is optimal in terms of overall computational efficiency. As shown in [21] for production optimization problems, this result appears to hold regardless of the number of parameters. In the examples of this paper, we use $N_g = 5$ as an approximate to the true gradient. However, in the first example, we explore the effect of N_g on the results.

The stochastic gradient in Eq. 12 is sensitive to the choice of the perturbation size ϵ^{ℓ}. After some experimentation, we set ϵ^{ℓ} to 0.01 for all examples presented in the paper. The stepsize a^{ℓ} is determined by a simple line search with the initial stepsize $a^0 = 0.25$. If the objective function does not decrease, a^{ℓ} is cut by half until we got a lower objective function or the maximum number of stepsize cuts is reached, which is 4 in the examples presented here. If the maximum number of cuts is reached without finding a p with $O(p) < O(p^{\ell})$, then we do not update ℓ or p. Instead, we simply generate a new search direction by averaging five new stochastic gradients and do the iteration again. If at this repeat

of the iteration, we cannot obtain a decrease in the objective function, the algorithm is terminated. Whenever we do obtain a decrease in the objective function, we evaluate for convergence based on the convergence criterion

$$\frac{\mid O(p^{\ell+1}) - O(p^{\ell}) \mid}{O(p^{\ell+1})} \leq 10^{-3}, \tag{14}$$

with the maximum number of simulation runs allowed equal to 600 for the first example and 400 for other examples.

4 NEWUOA

NEWUOA first constructs the initial quadratic model using N_i specially designed interpolation points. This quadratic model is then updated at each iteration as a new interpolation point is included in the interpolation set. In NEWUOA, the number of interpolation points, N_i, is fixed; thus, one interpolation point has to be removed when one new interpolation point is included [60]. The quadratic model at the ℓth iteration is

$$Q^{\ell}(p) = c^{\ell} + (p - p_0)^T g^{\ell} + \frac{1}{2}(p - p_0)^T G^{\ell}(p - p_0), \tag{15}$$

where p_0 is a fixed point. The coefficients c^{ℓ}, g^{ℓ}, and G^{ℓ} are, respectively, a scalar, an N_p-dimensional vector, and an $N_p \times N_p$-dimensional symmetric matrix. With a new interpolation point, NEWUOA updates the change of the quadratic model,

$$\begin{aligned}\delta Q^{\ell}(p) &= Q^{\ell}(p) - Q^{\ell-1}(p) \\ &= \delta c^{\ell} + (p-p_0)^T \delta g^{\ell} + \frac{1}{2}(p-p_0)^T \delta G^{\ell}(p - p_0),\end{aligned} \tag{16}$$

subject to the interpolation conditions at the interpolation set $\hat{p}_k^{\ell}$, i.e.,

$$\delta Q^{\ell}(\hat{p}_k^{\ell}) = O(\hat{p}_k^{\ell}) - Q^{\ell-1}(\hat{p}_k^{\ell}) \quad k = 1, 2, ..., N_i. \tag{17}$$

The total number of coefficients for the quadratic model in Eq. 16 is

$$N_i^* = \frac{(N_p + 1)(N_p + 2)}{2}, \tag{18}$$

so N_i^* interpolation conditions (Eq. 17) are required to fully determine all the coefficients, which in turn requires N_i^* objective function evaluations at the interpolation points. In our problems, each objective function evaluation requires a reservoir simulation run. Thus, when the number of model parameters (N_p) is large and the reservoir simulation model is large, it is worthwhile to try to minimize the number of function evaluations required. In general, NEWUOA only requires the number of interpolation points N_i be within the interval $[N_p + 2, N_i^*]$ and the originally recommended value for N_i was $2N_p + 1$ [60]. However, for many test problems with a large number of variables, experiments in [61] show that using $N_i = N_p + 6$ results in faster convergence. The extra degrees of freedom are used to minimize the Frobenius norm of δG, which is defined as

$$\|\delta G^{\ell}\|_F^2 = \sum_{i=1}^{N_p}\sum_{j=1}^{N_p}\left(\delta G_{ij}^{\ell}\right)^2. \tag{19}$$

This constrained minimization problem can be solved by minimizing the Lagrange function

$$\begin{aligned}L(\delta c^{\ell}, \delta g^{\ell}, \delta G^{\ell}, \lambda^{\ell}) = &\frac{1}{4}\sum_{i=1}^{N_p}\sum_{j=1}^{N_p}\left(\delta G_{ij}^{\ell}\right)^2 - \sum_{k=1}^{N_i}\lambda_k^{\ell} \\ &\times\left\{\delta Q^{\ell}(\hat{p}_k^{\ell}) - \left(O(\hat{p}_k^{\ell}) - Q^{\ell-1}(\hat{p}_k^{\ell})\right)\right\}.\end{aligned} \tag{20}$$

The minimum is achieved when the first-order derivatives $\nabla_{\delta c^{\ell}} L$, $\nabla_{\delta g^{\ell}} L$, $\nabla_{\delta G^{\ell}} L$, and $\nabla_{\lambda^{\ell}} L$ are all zero which yields the following $(N_i + N_p + 1)$ system of equations,

$$\begin{pmatrix} A & P^T \\ P & O \end{pmatrix}^{\ell} \begin{pmatrix} \lambda^{\ell} \\ \delta c^{\ell} \\ \delta g^{\ell} \end{pmatrix} = \begin{pmatrix} R^{\ell} \\ 0 \\ 0 \end{pmatrix}, \tag{21}$$

where the matrix A^{ℓ} is $N_i \times N_i$ and the entry in the ith row and the kth column is given by

$$A_{ik}^{\ell} = \frac{1}{2}\left\{\left(\hat{p}_i^{\ell} - p_0\right)^T \left(\hat{p}_k^{\ell} - p_0\right)\right\}^2. \tag{22}$$

R^{ℓ} is an N_i-dimensional vector with the components $O(\hat{p}_k^{\ell}) - Q^{\ell-1}(\hat{p}_k^{\ell})$, $k = 1, 2, ..., N_i$ and P^{ℓ} is $(N_p + 1) \times N_i$ matrix, defined by

$$P^{\ell} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \hat{p}_1^{\ell} - p_0 & \hat{p}_2^{\ell} - p_0 & \cdots & \hat{p}_{N_i}^{\ell} - p_0 \end{pmatrix}. \tag{23}$$

Note that setting $\nabla_{\delta G^{\ell}} L = 0$ yields the relationship between G^{ℓ} and λ_k^{ℓ}'s,

$$\delta G^{\ell} = \sum_{k=1}^{N_i} \lambda_k^{\ell} \left(\hat{p}_k^{\ell} - p_0\right)\left(\hat{p}_k^{\ell} - p_0\right)^T. \tag{24}$$

The parameters δc^{ℓ}, δg^{ℓ}, and λ^{ℓ} can then be obtained by solving Eq. 21. Substituting these parameters into Eq. 24 and then 16 yields the updated quadratic model $Q^{\ell}(p)$.

The updated quadratic model is minimized using a trust-region conjugate gradient method [60], i.e.,

$$\min Q^{\ell}\left(p_{\text{opt}}^{\ell-1}+\delta p^{\ell}\right), \text{ subject to } \parallel \delta p^{\ell} \parallel \leq \Delta^{\ell}, \quad (25)$$

where $p_{\text{opt}}^{\ell-1}$ is the optimal point found so far and Δ^{ℓ} is the trust-region radius.

5 QIM-AG algorithm

The major challenge in applying NEWUOA to large-scale optimization problems is that the number of interpolation points has to be larger than the dimension of the problem for initial quadratic model construction as it requires the estimation of the gradient of the quadratic model. To alleviate this problem, Zhao et al. [79] proposed QIM-AG. In the proposed QIM-AG algorithm, instead of determining c^{ℓ} and g^{ℓ} in the quadratic model (Eq. 15) using interpolation points, c^{ℓ} is fixed at the best function evaluation, i.e., $c^{\ell} = O(p_{\text{opt}}^{\ell-1})$, and g^{ℓ} is replaced with $\overline{\hat{g}^{\ell}}$, the average of several stochastic gradients, i.e., Eq. 13 at each iteration. The quadratic function now becomes

$$Q^{\ell}(p) = O\left(p_{\text{opt}}^{\ell-1}\right) + \left(p - p_{\text{opt}}^{\ell-1}\right)^{T} \overline{\hat{g}^{\ell}} + \frac{1}{2}\left(p - p_{\text{opt}}^{\ell-1}\right)^{T} G^{\ell}\left(p - p_{\text{opt}}^{\ell-1}\right). \quad (26)$$

To determine G^{ℓ}, the QIM-AG algorithm minimizes the Frobenius norm of G^{ℓ} under the constraints that the quadratic model of Eq. 26 is equal to the objective function at all interpolation points $\hat{p}_k^{\ell}$,

$$Q^{\ell}\left(\hat{p}_k^{\ell}\right) = O\left(\hat{p}_k^{\ell}\right), \ k = 1, \cdots, N_i. \quad (27)$$

QIM-AG requires a minimum of one interpolation point, i.e., $N_i \geq 1$ in addition to $p_{\text{opt}}^{\ell-1}$ while NEWUOA requires $N_i \geq N_p + 2$.

Following the derivation in NEWUOA and Zhao et al. [79], the updated quadratic model at the ℓth iteration is given by

$$Q^{\ell}(p) = O\left(p_{\text{opt}}^{\ell-1}\right) + \left(p - p_{\text{opt}}^{\ell-1}\right)^{T} \overline{\hat{g}^{\ell}} + \frac{1}{2}\sum_{k=1}^{N_i} \lambda_k^{\ell}\left\{\left(p - p_{\text{opt}}^{\ell-1}\right)^{T}\left(\hat{p}_k^{\ell} - p_{\text{opt}}^{\ell-1}\right)\right\}^2. \quad (28)$$

Using the interpolation condition of Eq. 27 in Eq. 28 yields N_i linear equations. These equations can be written in a matrix form as

$$A^{\ell}\lambda^{\ell} = R^{\ell}, \quad (29)$$

where $\lambda^{\ell} = [\lambda_1^{\ell}, \lambda_2^{\ell}, \cdots, \lambda_{N_i}^{\ell}]^T$; the matrix A^{ℓ} has the elements

$$A_{ik}^{\ell} = \frac{1}{2}\left\{\left(\hat{p}_i^{\ell} - u_{\text{opt}}^{\ell-1}\right)^{T}\left(\hat{p}_k^{\ell} - p_{\text{opt}}^{\ell-1}\right)\right\}^2, \quad (30)$$

and vector R^{ℓ} has the elements

$$R_k^{\ell} = O\left(\hat{p}_k^{\ell}\right) - O\left(p_{\text{opt}}^{\ell-1}\right) - \left(\hat{p}_k^{\ell} - p_{\text{opt}}^{\ell-1}\right)^{T} \overline{\hat{g}^{\ell}}, \quad (31)$$

for $k = 1, \cdots, N_i$. Solving Eq. 29 for λ^{ℓ} and substituting it into Eq. 28 yields the interpolation quadratic model for the QIM-AG algorithm. Similar to Eq. 25, this quadratic model is then minimized with a trust-region method (DGQT routine from the MINIPACK2 [48]) to obtain an updated model, $\tilde{p}_{\text{opt}}^{\ell} = p_{\text{opt}}^{\ell-1} + \delta p_{\text{opt}}^{\ell}$. If $\tilde{p}_{\text{opt}}^{\ell}$ yields a strict decrease in the objective function, i.e., $O(\tilde{p}_{\text{opt}}^{\ell}) < O(p_{\text{opt}}^{\ell-1})$, it is accepted as the optimal point for the ℓth iteration, i.e., $p_{\text{opt}}^{\ell} = \tilde{p}_{\text{opt}}^{\ell}$. Then, the quadratic model for the next iteration, $Q^{\ell+1}$, is updated using $O(p_{\text{opt}}^{\ell})$ as the constant part, a new average stochastic gradient from Eq. 13 as the gradient part, and p_{opt}^{ℓ} as an additional interpolation point for the Hessian matrix. If the norm of the change in model $\|\delta p^{\ell}\|$ is equal to the trust-region radius, i.e., $\|\delta p^{\ell}\| = \Delta^{\ell}$, the trust-region radius is doubled for the next iteration as long as the updated trust-region radius is less than the maximum trust-region radius specified, $\Delta_{\max}$, for the algorithm, i.e., $\Delta^{\ell+1} = \min(\Delta_{\max}, 2\Delta^{\ell})$. However, if $\|\delta p^{\ell}\| < \Delta^{\ell}$, the trust-region radius does not change, i.e., $\Delta^{\ell+1} = \Delta^{\ell}$. If the current iteration does not yield a strict decrease in the objective function, i.e., if $O(\tilde{p}_{\text{opt}}^{\ell}) \geq O(p_{\text{opt}}^{\ell-1})$, p_{opt}^{ℓ} is set to $p_{\text{opt}}^{\ell-1}$ and the trust-region radius is cut by half, i.e., $\Delta^{\ell+1} = 0.5\Delta^{\ell}$. However, the quadratic model is updated using $\tilde{p}_{\text{opt}}^{\ell}$ as an additional interpolation point, which is the updating strategy used in "wedge method" [46]. The algorithm converges when the maximum number of iteration is reached or $\Delta^{\ell} \leq \Delta_{\min}$, where $\Delta_{\min}$ is the minimum trust-region radius specified.

In the examples, the initial trust-region radius Δ^0 for QIM-AG is always set equal to 2; this value was determined by experimentation (see the first example) and has worked well for all examples we have tried. $\Delta_{\min} = 0.01$ and $\Delta_{\max} = 5.0$ in all examples. In the QIM-AG algorithm, the number of interpolation points, N_i, increases as the iterations proceed as long as $N_i \leq 5N_p$. When the number of interpolation points is greater than $5N_p$, the farthest points from the optimal point are removed from the interpolation set.

6 Examples

In this section, we apply the three derivative-free optimization algorithms (SGSD, NEWUOA, and QIM-AG) to three synthetic reservoir cases: a 2D synthetic reservoir model, the well-known PUNQ reservoir [25], and the Brugge field [57].

6.1 Example 1: 2D synthetic model

This horizontal 2D synthetic model is the same as the one used in Zafari and Reynolds [77]. The reservoir has 20×30 gridblocks. For parameterization, we generate 100 initial unconditional realizations of the porosity and log-permeability fields (i.e., $N_e = 100$). Figure 1 shows three initial realizations and the corresponding mean of log-permeability. The true horizontal log-permeability map with the location of six wells is shown in Fig. 2a. Note that permeability is isotropic. Wells INJ1, Pro2, and Pro5 lie along a high-permeability channel and the other wells are drilled in relatively low-permeability regions. The true reservoir porosity and log-permeability fields were generated using sequential Gaussian co-simulation. The well injection and production schedule is the same as that in Zafari and Reynolds [77]. The first well was put on production at time 0 with an oil rate of 100 STB/day, and every 90 days, one more well was put on production at the same rate of 100 STB/day. After the sixth well had produced for 90 days, the first well was converted to an injector with an injection rate of 1,000 STB/day and the oil production rate of all other wells was increased to 200 STB/day. At day 5,850, the injection rate of well INJ1 was increased to 2,000 STB/day, Pro5 was shut in due to excessive water production, and the oil production rate of all other production wells was increased to 300 STB/day. The dynamic well data to be history-matched are bottom hole pressure (BHP) of the injector and producers, gas production rate (GPR), and water production rate (WPR) of the production wells. The observed data (d_{obs}) are generated by adding Gaussian noise to the predicted data generated with the true porosity and permeability fields. Random normally distributed noise with zero mean and standard deviation equal to 5 % of the true data was added to the true data to define the observed data.

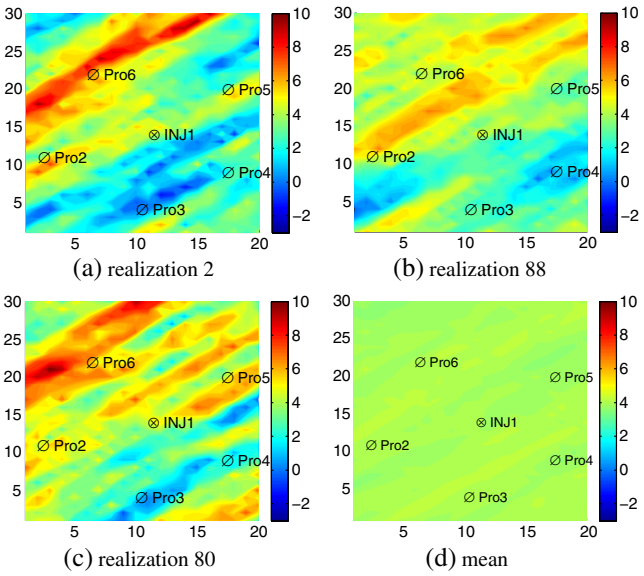


Fig. 1 Some initial realizations of log-permeability field and the mean of the ensemble of realizations, example 1

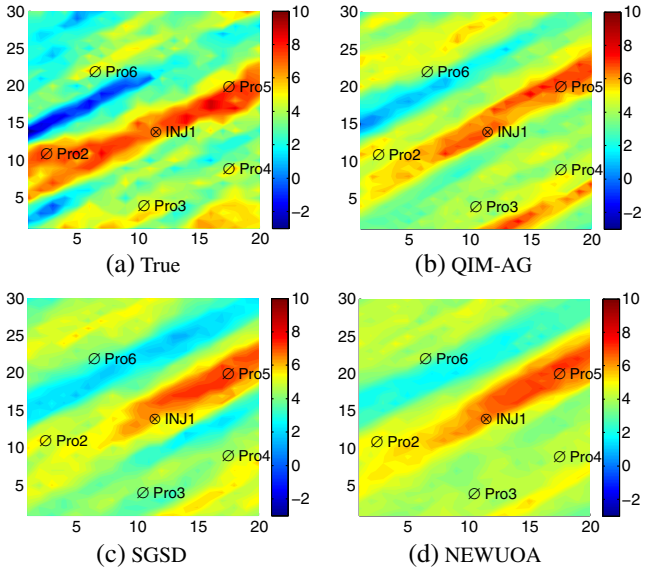


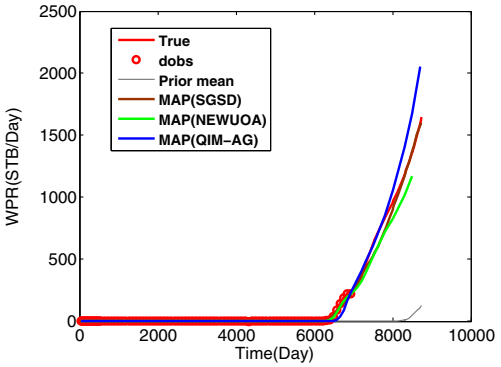
Fig. 2 The MAP estimate of log-permeability field compared with the truth, example 1

In NEWUOA, the initial trust-region radius Δ^0 and the initial value of its lower bound, ρ, are both equal to 0.4. For convergence, the minimum value for ρ is $\rho_{\text{end}} = 0.1$. The total number of interpolation points is $N_i = N_p + 6$ where $N_p = N_e - 1$ after parameterization. For the QIM-AG algorithm, the initial trust-region radius is set to $\Delta^0 = 2.0$ with an upper and lower bound of $\Delta_{\max} = 5.0$ and $\Delta_{\min} = 0.01$, respectively. An average of five stochastic gradients is used to define the search direction, i.e., $N_g = 5$ in Eq. 13. Figure 2b–d shows the MAP estimates of the log-permeability fields obtained with algorithms QIM-AG, SGSD, and NEWUOA. As shown in the figure, all of the above methods give results that capture the main geological features of the true permeability field, especially for

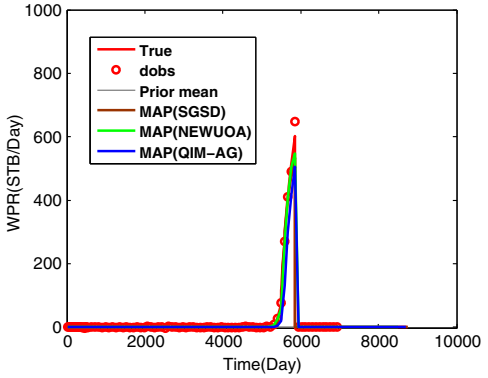
the high-permeability region connecting Pro5 and Pro2. However, the MAP estimates of the log-permeability field is smoother than the truth model as expected. The porosity field has a correlation coefficient of 0.8 with the permeability field, and similar MAP estimates of porosity are obtained from the three optimization algorithms.

Figure 3 illustrates the BHP data match up to day 7,290 and prediction of BHP until day 9,000 of two wells (INJ1 and Pro4). In the figure, the red curve is the true BHP data obtained by running the true geological model and the red circles are the observed BHP data. The initial guess (prior mean) gives a prediction far from that of the true model. After history matching, the predictions from the MAP estimates generated with the three optimization algorithms match the observations reasonably well. The WPR data match for Pro2 and Pro5 is shown in Fig. 4. Note that Pro2 and Pro5 are the only two wells that have water breakthrough before day 9,000 and Pro5 was shut in around day 6,000 due to excessive water production. The prior mean model predicts no water breakthrough for both wells. Figures 5 and 6, respectively, show the matches and future predictions of the gas production rate at Pro4 and the field cumulative water production. As shown in Figs. 3, 4, 5, and 6, good data matches and future predictions are obtained for all the algorithms.

Figure 7 shows the convergence rates of different algorithms starting with the prior mean as the initial guess. With the maximum allowable simulation runs equal to 600, the three optimization algorithms yield almost the same objective function value of about 3,400,

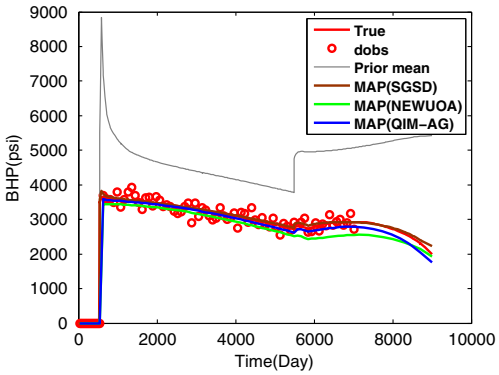


(a) Pro2

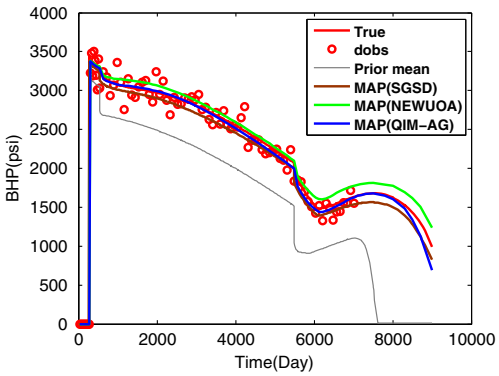


(b) Pro5

Fig. 4 Water production rate data match, example 1



(a) INJ1



(b) Pro4

Fig. 3 BHP data match, example 1

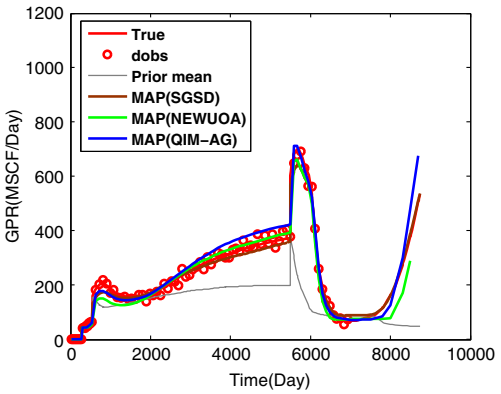


Fig. 5 Gas production rate data match of Pro4, example 1

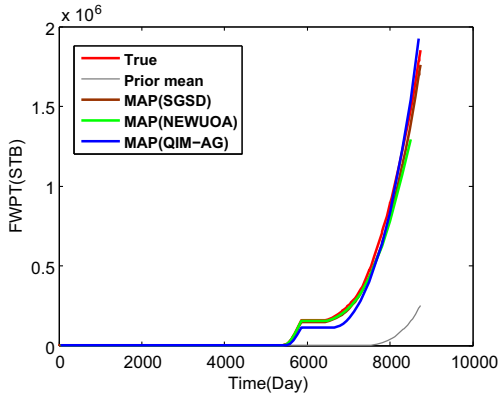


Fig. 6 Field cumulative water production, example 1

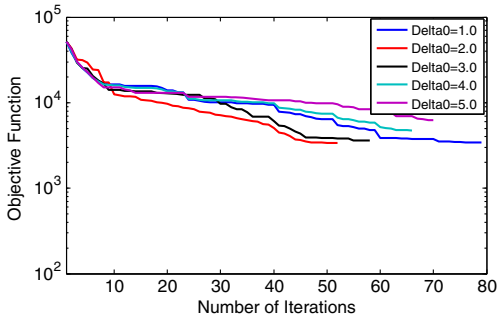


Fig. 8 Objective function versus iterations with different Δ^0 for QIM-AG algorithm, example 1

while the QIM-AG method (red curve) requires far fewer simulation runs than SGSD and NEWUOA. Note that optimization in NEWUOA starts after the construction of the initial quadratic model, which uses about 100 simulation runs shown as the flat section in the objective function of Fig. 7. At the early iterations, the QIM-AG algorithm behaves similar to that of the SGSD method as there are very few interpolation points available and the minimization of the quadratic model is basically done along the hyperplane determined by the previous optimal point and the SGSD search direction. At late iterations where more interpolation points are added, the rate of decrease of the objective function in the QIM-AG algorithm is similar to that of NEWUOA.

For QIM-AG and NEWUOA algorithms, we performed some tests for parameter selection of the initial trust-region radius Δ^0 in this example. The other parameters are all the same as defined above. The convergence performance for different values of Δ^0 for QIM-AG algorithm is shown in Fig. 8 and Table 1. For the three cases, when $\Delta^0 \leq 3.0$, the final objective function value at convergence is about 3,400, which is obviously smaller than the final objective function

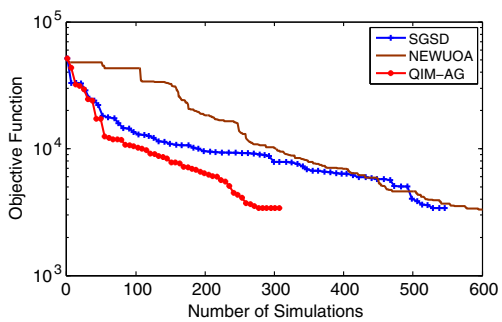


Fig. 7 Objective function Eq. 9 versus simulation runs, example 1

obtained when $\Delta^0 \geq 4.0$. When $\Delta^0 = 1.0$, we obtain a slightly slower convergence rate, i.e., more simulation runs are required to obtain convergence, than for $\Delta^0 = 2.0$. In the following two examples, we set the parameter for Δ^0 to 2.0 in QIM-AG algorithm. The performance of convergence rates for NEWUOA can be seen in Fig. 9 and Table 2. As shown in Fig. 9, with a larger Δ^0, NEWUOA may result in a faster decrease in the objective function at early iterations but results in a final value of the objective function that is larger, especially when $\Delta^0 \geq 1.0$. A good choice for Δ^0 seems to be between 0.1 and 1.0, and we simply set Δ^0 to 0.4 in the next two examples.

For the QIM-AG algorithm, another key parameter that may affect the computational behavior is the number of stochastic gradients N_g averaged to obtain the search direction at each iteration (see Eq. 13). Because the expectation of the SPSA gradient is equal to the true gradient, one might expect that averaging more SPSA gradients in Eq. 13 would yield superior results. While it is true that one might expect to obtain a better approximation to the true gradient by using a larger value of N_g, increasing N_g also increases the number of reservoir simulation runs required to reach convergence. For production optimization problems, [21] found that using an average of five to ten SPSA gradients was optimal for virtually all of the example considered, specifically using $N_g > 10$ did not generally yield a significant improvement in the optimum but

Table 1 The convergence rates with different Δ^0 for QIM-AG algorithm, example 1

Δ^0	Total simulations	Final objective function
1.0	469	3,403.5
2.0	313	3,390.2
3.0	349	3,427.4
4.0	403	4,722.6
5.0	421	6,261.1

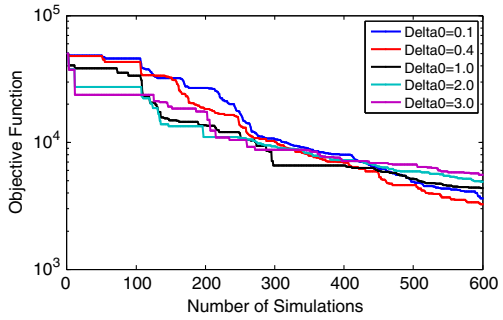


Fig. 9 Objective function versus simulations with different Δ^0 for NEWUOA algorithm, Example 1

Table 2 The convergence rates with different Δ^0 for NEWUOA algorithm, example 1

Δ^0	Total simulations	Final objective function
0.1	600	3,534.9
0.4	600	3,440.7
1.0	600	4,369.8
2.0	600	4,904.1
3.0	600	5,541.6

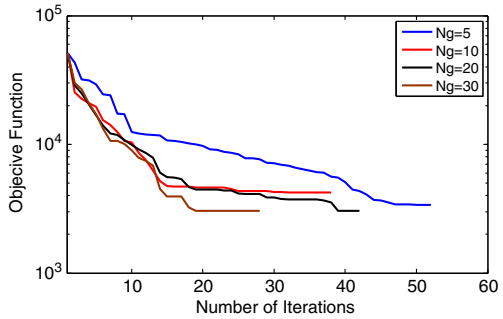


Fig. 10 Objective function versus iterations with different N_g for QIM-AG algorithm, example 1

Table 3 The convergence rates with different N_g for QIM-AG algorithm, example 1

N_g	Total simulations	Final objective function
5	313	3,390.2
10	419	4,225.0
20	883	3,023.1
30	834	2,938.6

invariably required far more simulation runs to reach convergence. For the history-matching problem of example 1, we present in Fig. 10 and Table 3 the convergence rates and the minimum values of the objective function obtained for different values of N_g. In this experiment, we use $\Delta^0 = 2$. The lowest value of the objective function (best history match) was obtained by averaging 30 SPSA gradients but the result is not substantially better than the result obtained with $N_g = 5$, and when averaging 30 gradients, 834 simulation runs were required to reach convergence whereas only 313 simulation runs were required for the termination of the algorithm when only five SPSA gradients were averaged. As it would be costly to do experiments to determine the optimal value of N_g, we have opted to simply use $N_g = 5$ in all cases.

6.2 Example 2: PUNQ reservoir

PUNQ-S3 is a three-phase, three-dimensional reservoir model [3, 25, 28, 29]. Figure 11 shows the true horizontal log-permeability distribution for all five layers. The blue region at the bottom and left corner is a strong aquifer. In this study, we simulate the aquifer numerically with high porosity (0.95) and 100 % water saturation for the aquifer gridblocks as in Gao [26]. The reservoir is bounded with two faults to the right and top. A small gas cap is located at the center of the dome-shaped structure. There are six production wells drilled around the gas cap. No producers are completed in the top two layers. The production data used for

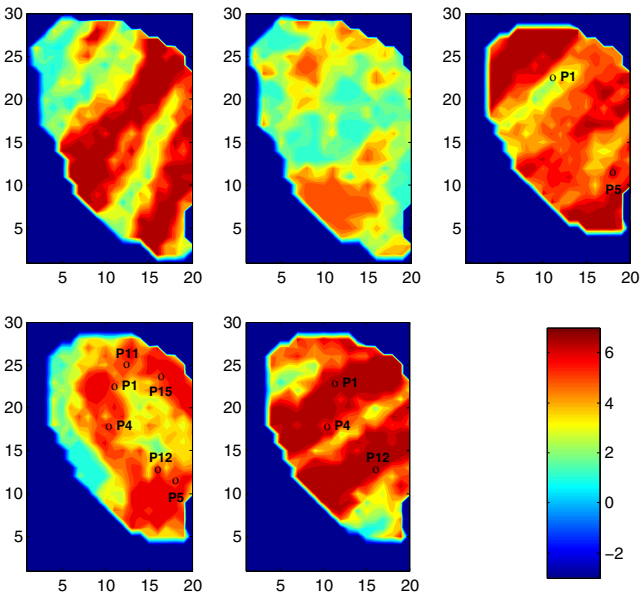


Fig. 11 The true log-permeability field, example 2

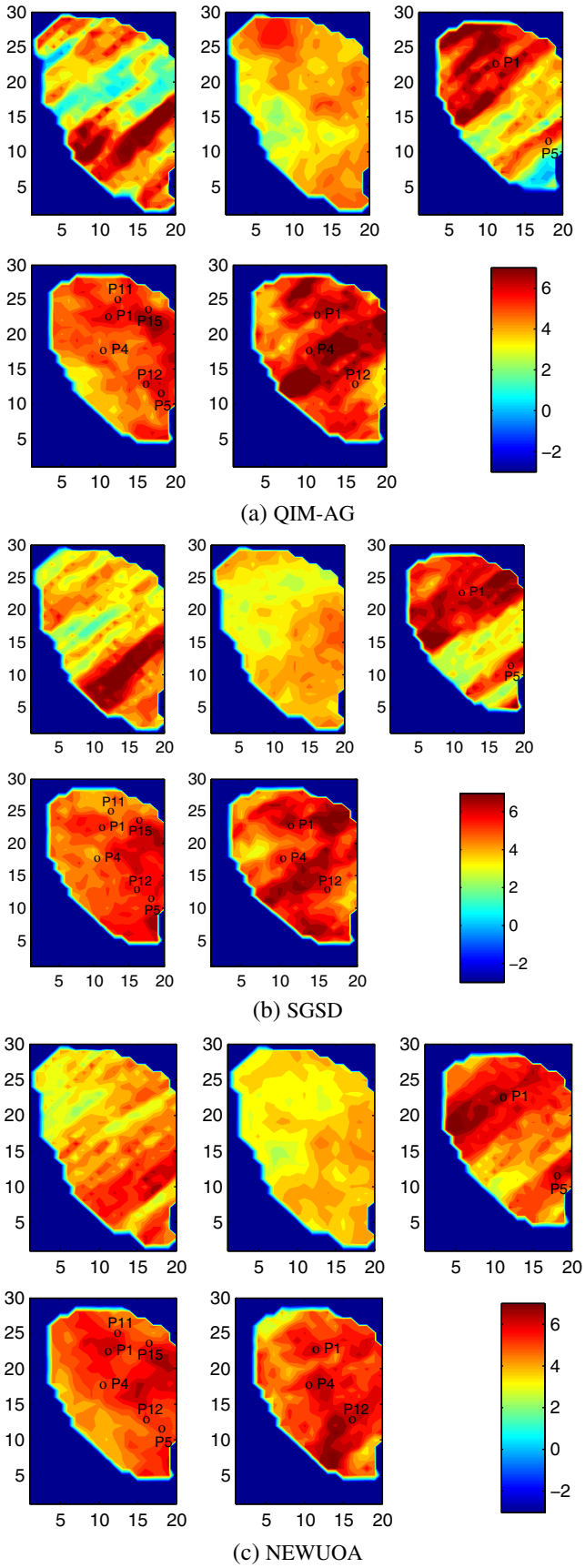


Fig. 12 The MAP estimate of log-permeability field, example 2

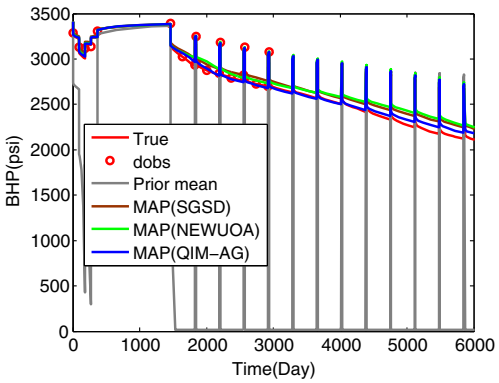


Fig. 13 BHP match data for PRO1, example 2

history matching are BHP, GOR, and WCT, which are the same as in Gao et al. [29] with the same production schedule. The wells were produced for 1 year and then shut in for a 3-year extended well test. After they were reopen, the wells were producing at a rate of 943.5 STB/day with a minimum bottomhole pressure constraint of 14.7 psi. The observations between time 0 and day 2,936 are used for history matching, and then the estimated reservoir models are run for prediction up to day 6,000. The reservoir model parameters are horizontal and vertical gridblock log-permeability and porosity. The number of realizations used for parameterization is 100 and $N_p = 99$.

The optimization parameters for the QIM-AG algorithm are set to $\Delta^0 = 2.0$ with an upper and lower bound of $\Delta_{\max} = 5.0$ and $\Delta_{\min} = 0.01$. For NEWUOA, the initial trust-region radius is 0.4 and its minimum value for convergence is 0.1. The total number of interpolation points is $N_i = N_p + 6$. The final MAP estimates of the horizontal log-permeability for each layer obtained by the optimization algorithms are presented in Fig. 12. As shown in the figure, all algorithms yield log-permeability fields that give a vague hint of the

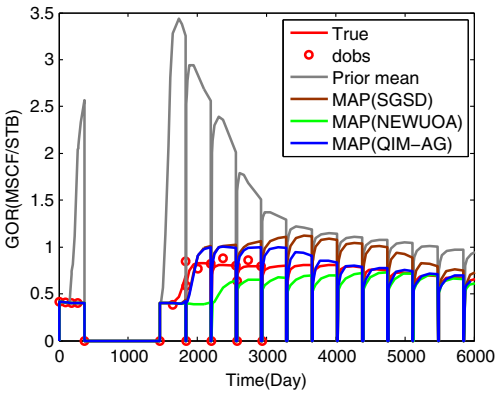


Fig. 14 GOR match data for PRO1, example 2

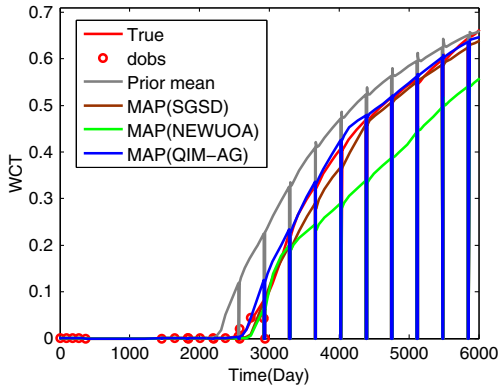


Fig. 15 WCT match data for PRO11, example 2

channelized geological features of the true permeability field (Fig. 11), but the estimated log-permeability fields are roughly right considering that the initial guess for history matching is the prior mean, which has a constant log-permeability value for each layer. Also by nature, the MAP estimate is smoother than individual realizations (see Gao et al. [28]).

Figure 13 shows the BHP data match of well PRO1. The spikes are due to the 15-day shut-in each year. The initial model cannot maintain the oil production rate specified, so the well produces at a minimum bottomhole pressure of 14.7 psi. After history matching, the data predictions from the MAP estimates of all the three methods are quite close to the true (red curve), even in the forecasting period from day 2,936 to day 6,000. Figures 14 and 15 illustrate the GOR data match of well PRO1 and the WCT data match of well PRO11, respectively. All methods lead to an improved data match over that obtained with the prior model. However, the NEWUOA algorithm underestimates the WCT of PRO11 compared to the true in the prediction period.

The convergence performance of the different algorithms is shown in Fig. 16. The QIM-AG and SGSD algorithms, respectively, converged after 193 and 313 simulation runs, and NEWUOA was terminated as it reached the maximum number of simulation runs, which is 400. All algorithms reached a minimum objective function value close to 100. Similar to example 1, if we plotted the NEWUOA results in Fig. 13 with 106 subtracted from the number of simulation runs, then the NEWUOA and QIM-AG results would roughly overlay.

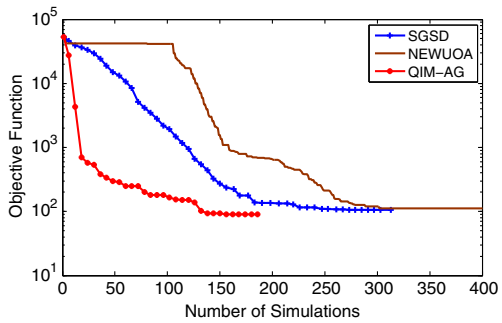
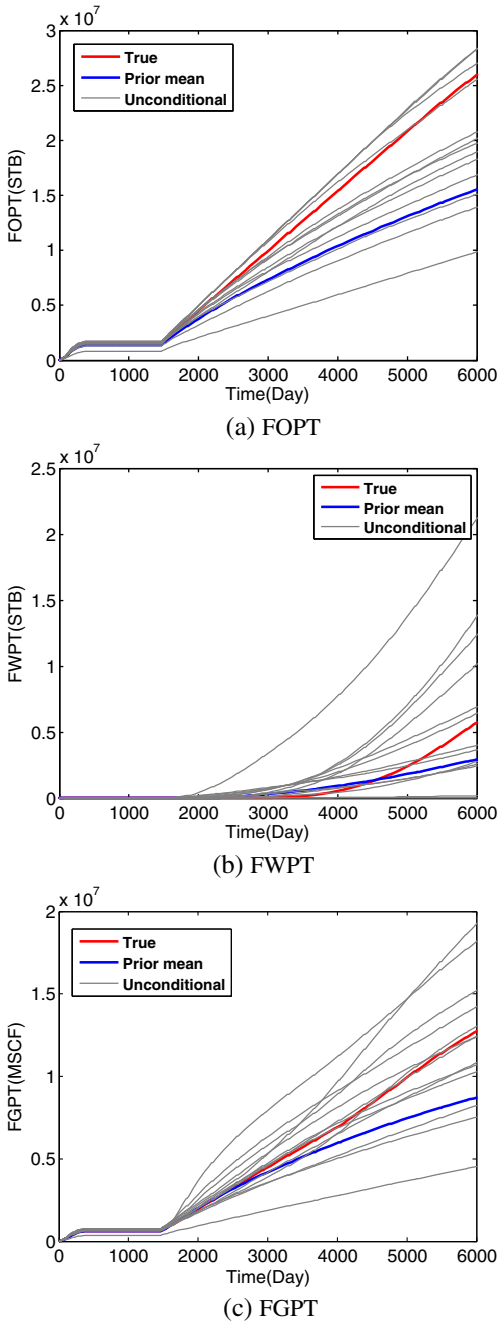


Fig. 16 Objective function versus simulation runs, example 2



(a) FOPT

(b) FWPT

(c) FGPT

Fig. 17 Data match from unconditional realization, example 2

To quantify uncertainty in the future reservoir performance predictions, ten conditional realizations were generated using the RML method by minimizing Eq. 10 with different (p_{uc}, d_{uc}) pairs using the QIM-AG algorithm. The realization p_{uc} is sampled from the normal distribution $N(0, I_{N_p})$. The predictions of cumulative oil, water, and gas production for the unconditional and conditional realizations are shown in Figs. 17 and 18. In all the figures, the red curve represents the true prediction, the blue curve represents either the prediction based on the prior mean model or the MAP estimate, and the gray curves are the predictions from the unconditional or conditional realizations depending on the figure considered. Figure 17 shows that the unconditional and prior mean models do not provide good predictions. After history matching with the QIM-AG algorithm, the conditional realizations and MAP estimate reduce the uncertainty in predictions and give predictions that are close to those of the truth (Fig. 18).

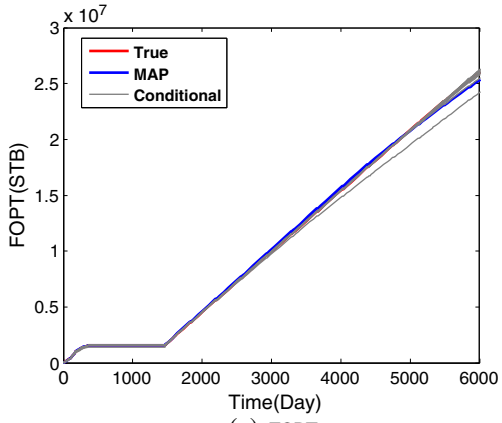


Fig. 18 Data match from conditional realization, example 2

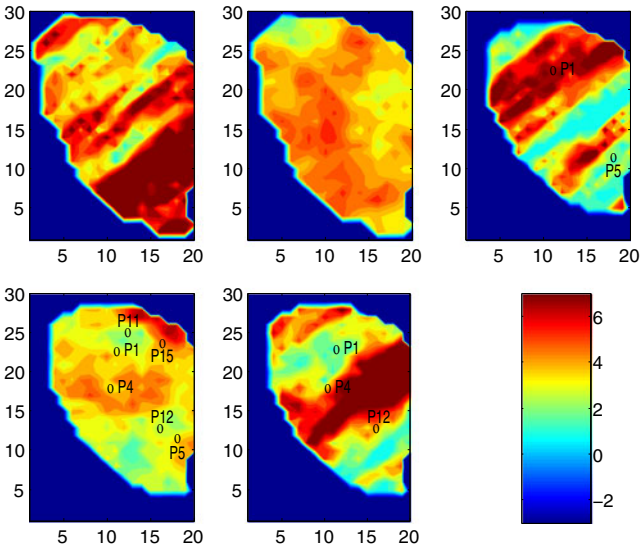


Fig. 19 The log-permeability field of one unconditional realization, example 2

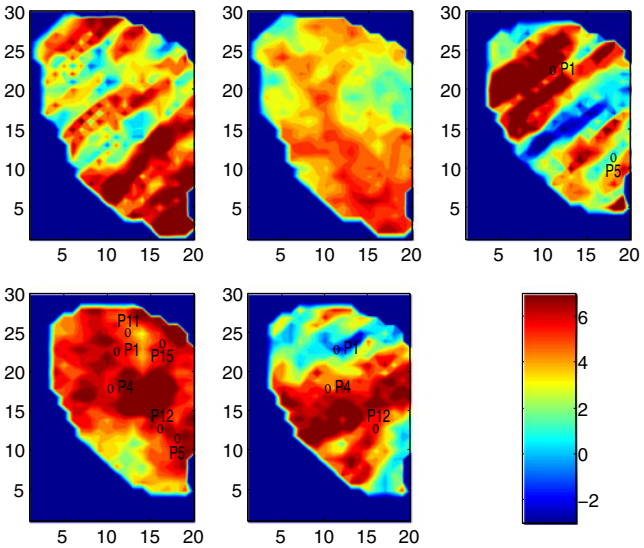


Fig. 20 The log-permeability field of one conditional realization, example 2

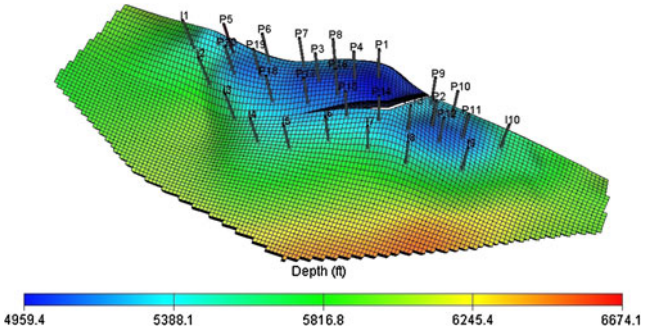


Fig. 21 The top structure of Brugge field, example 3

Figures 19 and 20, respectively, present an unconditional realization of the log-permeability field and its corresponding conditional realization after matching the data. The figures show that limited model changes are made to match the data. The most noticeable change is in layer 4, where the permeability in most gridblocks was increased. The conditional realization is much rougher than the MAP estimate in Fig. 12.

6.3 Example 3: Brugge reservoir

The Brugge field is a synthetic reservoir developed by TNO [57] as a benchmark study to test different algorithms for closed-loop reservoir management. The simulation model has a grid system of $139 \times 48 \times 9$. The total number of active gridblocks is 44,550. The top structure map and well locations are shown in Fig. 21.

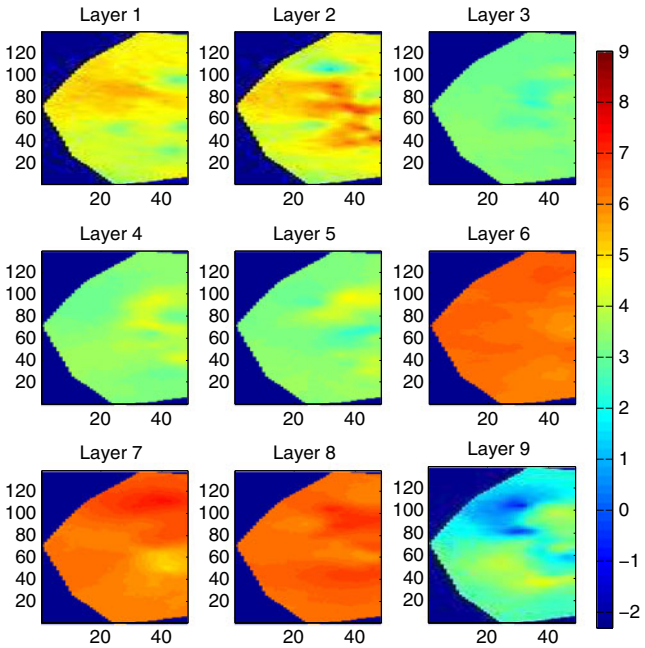


Fig. 22 Prior mean log-permeability model, example 3

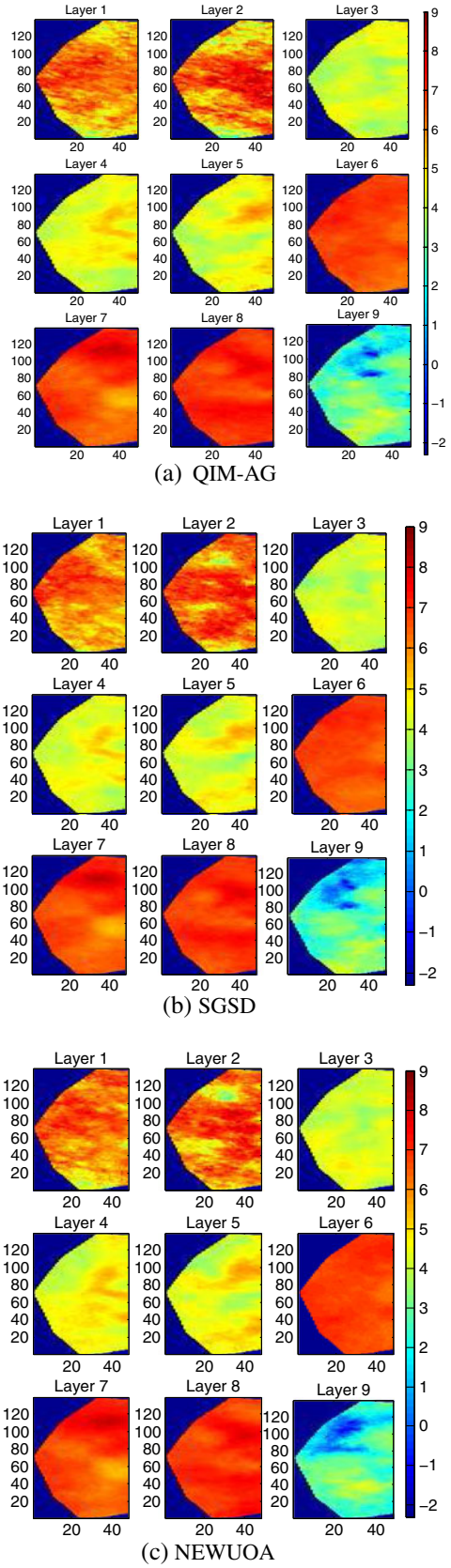
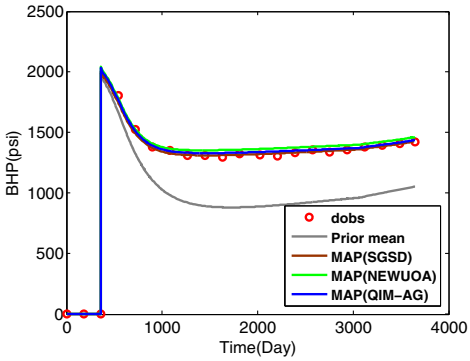
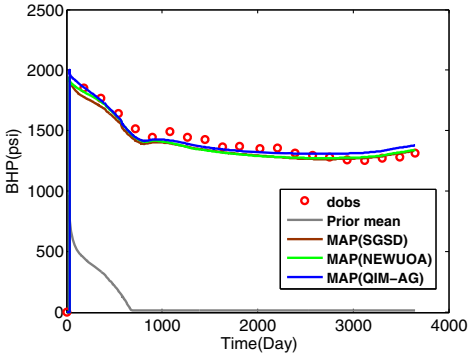


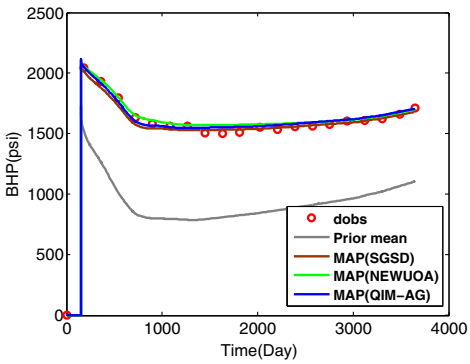
Fig. 23 MAP estimate of log-permeability fields, example 3



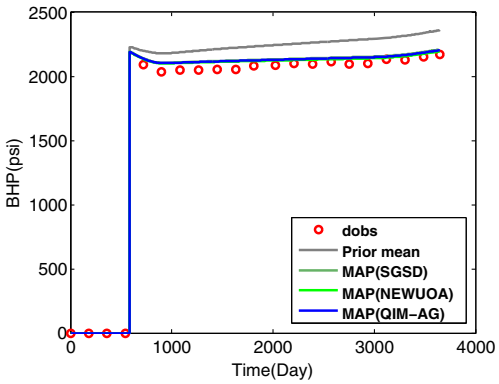
(a) P1



(b) P10

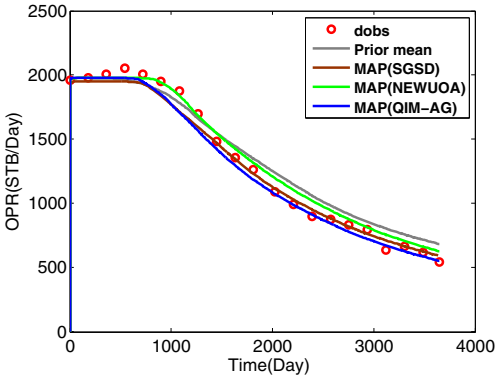


(c) P14

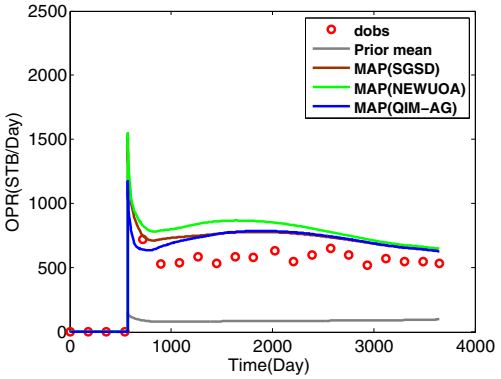


(d) I1

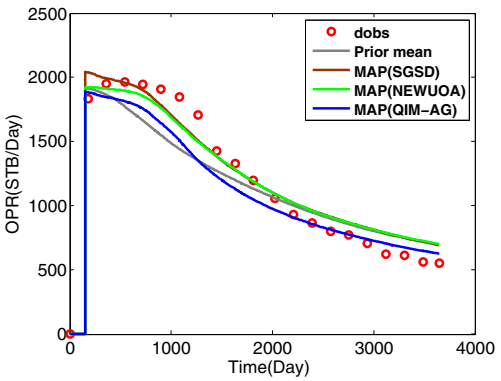
Fig. 24 BHP data match, example 3



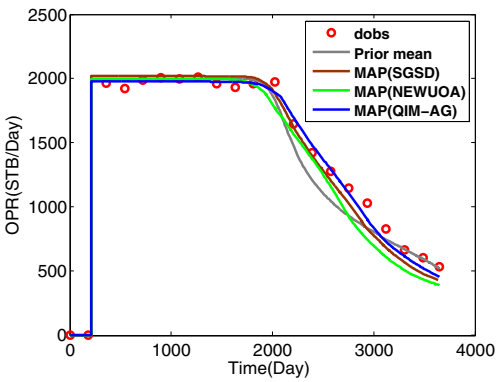
(a) P5 OPR



(b) P9 OPR



(c) P14 OPR



(d) P16 OPR

Fig. 25 Oil production rate (OPR) data match, example 3

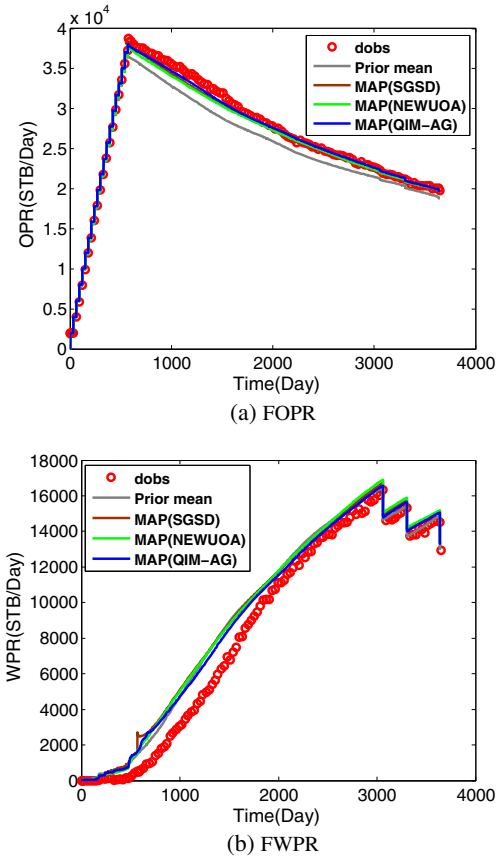


Fig. 26 Field production rate, example 1

There are 30 vertical wells in the reservoir including 20 producing wells and ten water injection wells. The target liquid rate for each producer is 2,000 STB/day with a minimum bottomhole pressure constraint of 725 psi. The target rate for each water injector is 4,000 STB/day and the maximum allowable pressure is 2,611 psi. Observation data including bottomhole pressure or water injection rates at injection wells and bottomhole pressure or oil and water production rates at producers were provided at 1-month intervals for the first 10 years. For history matching, we use the same noise level as in Chen et al. [11]. The parameters to be estimated are the following properties of each gridblock: net-to-gross ratio, porosity, initial water saturation, and absolute permeability in the x-, y-, and z-directions. Figure 22 shows the prior mean reservoir model of horizontal log permeability obtained by averaging the ensemble of 104 initial realizations. For parameterization, we use all the 104 realizations provided by TNO and the number of model parameters.

The optimization parameters for the QIM-AG algorithm are set to $\Delta^0 = 2.0$ with an upper and lower bound of $\Delta_{\max} = 5.0$ and $\Delta_{\min} = 0.01$. For NEWUOA, the initial trust-region radius is 0.4 and its minimum value for convergence is 0.1. The total number of interpolation points is $N_i = N_p + 2$.

The MAP estimates of log-permeability field obtained by the three algorithms are shown in Fig. 23. By visual comparison, QIM-AG, NEWUOA, and SGSD give similar results in the MAP estimates of horizontal

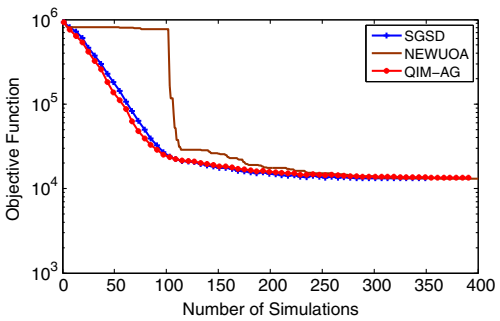


Fig. 27 Objective function versus the simulation numbers, example 3

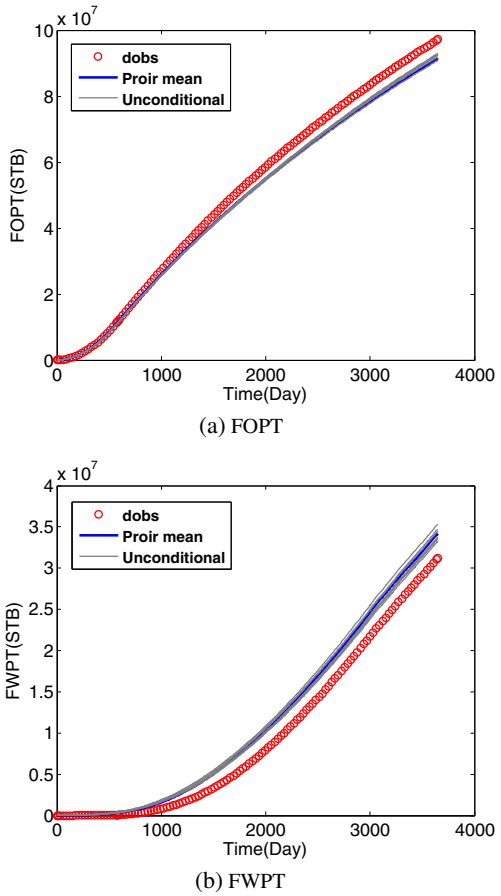
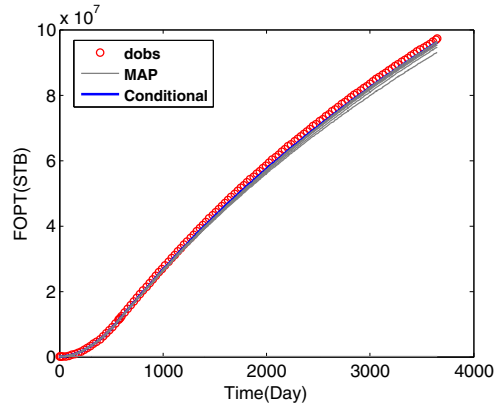
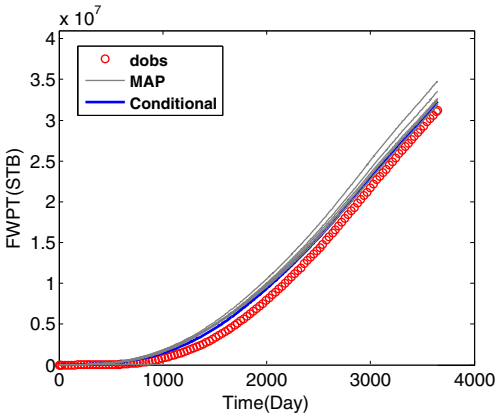


Fig. 28 Data match from unconditional realization, example 3



(a) FOPT



(b) FWPT

Fig. 29 Data match from conditional realization, example 3

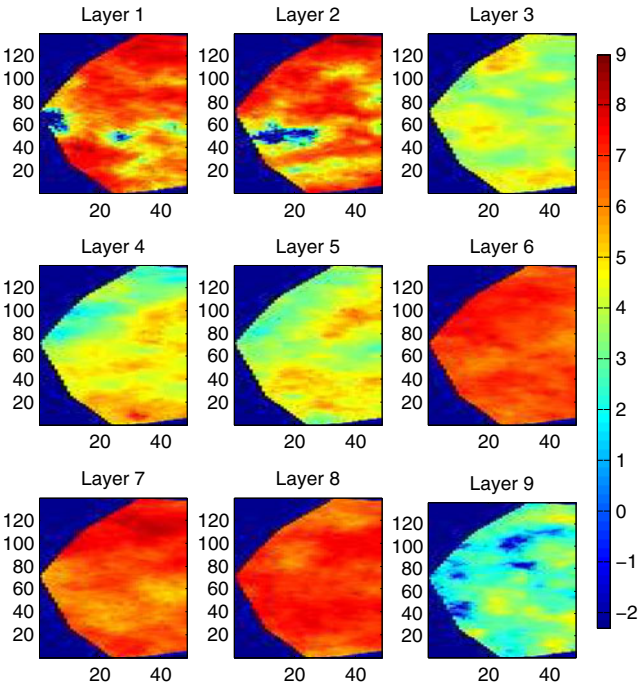


Fig. 30 The log-permeability field of one unconditional realization, example 3

log permeability and they all look smoother than the mean model obtained from EnKF [11, 13, 57].

Figure 24 illustrates the BHP data match of four wells for the first 10 years. In the figure, the red circles are the observed BHP data. The predicted data obtained from the initial guess (prior mean) are far from the observed data. After history matching, the predictions from the MAP estimates generated from the three algorithms match the data very well. The oil production rate data match for four producers and the field production rate data match are shown in Figs. 25 and 26, respectively. The oil production rate data match is greatly improved compared to the initial guess (prior mean) for producer P9. The improvement on the oil production rate data match for other wells is not as obvious as for well P9. The BHP and oil production rate data match obtained from all algorithms is comparable to or better than the results shown in Peters et al. [57].

The convergence performance of different algorithms with simulation runs is shown in Fig. 27. In this example, all the optimization algorithms yield a final objective function value of about 12,300 within 400 simulation runs. Compared to the previous two examples, the decreasing rate in the objective function at early iterations is much slower in QIM-AG than that of NEWUOA. This may indicate that the quality of the

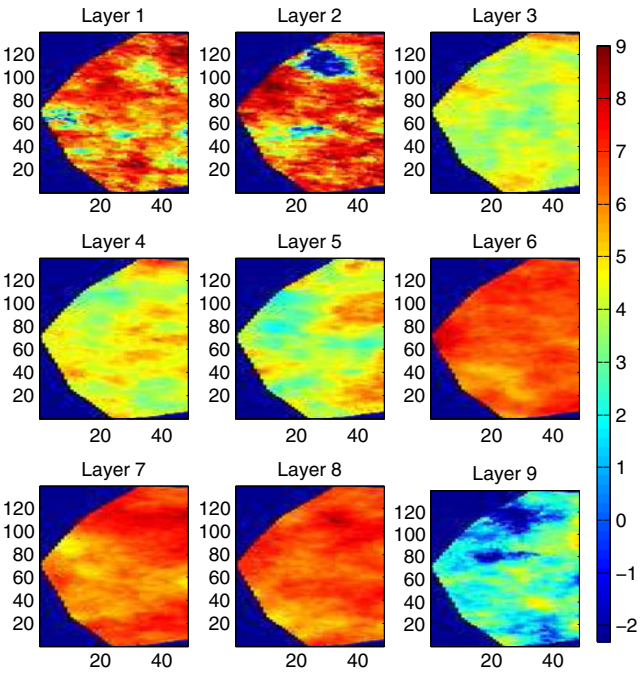


Fig. 31 The log-permeability field of one conditional realization, example 3

quadratic approximation model in QIM-AG is not as good as the one in NEWUOA.

We also applied the RML method using QIM-AG algorithm for uncertainty quantification. As shown in Figs. 28 and 29, the predictions in FOPT obtained from the set of conditional models significantly improved after history matching and are quite close to the true observation data compared with that of the unconditional realizations. However, the data match for FWPT only improved slightly. The results are comparable to those obtained with other algorithms shown in Peters et al. [57]. Figures 30 and 31 show the log-permeability field of one unconditional realization and its corresponding conditional realization.

7 Note on the performance of SGSD and QIM-AG

Based on nonexhaustive experiments, we choose to use an initial trust region radius of $\Delta^0 = 2.0$ for the QIM-AG algorithm and $\Delta^0 = 0.4$ in NEWUOA. Similarly, based on computational experiments, we choose to use $N_g = 5$, i.e., to use the average of five SPSA gradients to compute a search direction in SGSD and QIM-AG. Although there is no guarantee that these choices are optimal, or even appropriate for all problems, it would not be feasible to do a large number of experiments to determine optimal values for each problem considered. Moreover, these values have worked well for other problems similar to those considered here.

There currently exists no formal convergence proof for the QIM-AG and SGSD algorithms, and as we have experimented with these methods only for a handful of problems, there is no guarantee that the methods will exhibit a similar rate of convergence for all problems. However, the experiments we have done suggest that the behavior of the SGSD and QIM-AG algorithms is fairly illustrative of what we can expect for other history-matching problems. Moreover, we observed a similar rate of convergence when SGSD and QIM-AG were applied to the optimal well control problem when the only constraints were simple bounds on the well controls [79]. For this optimal well control problem, the number of reservoir simulation runs required to obtain numerical convergence ranged from about 400 to 1,000 but [79] did not use reparameterization and the number of variables (well controls) estimated was equal to 3,360 for the largest problem considered. Although the use of any method which does not have a formal convergence proof should be used with some caution, the results presented here and in [79] suggest that SGSD and/or QIM-AG are useful algorithms for both the history matching and production optimization step of closed-loop reservoir. If this is indeed true, and we believe it is, then these algorithms would not be the first ones of value that are not supported by rigorous theory; consider, for example, the now highly popular particle swarm algorithm. Moreover, it is important to note that the SGSD shares the most important features of the classical SPSA algorithm of [70], which has a convergence proof.

8 Conclusions

In this paper, we compared three derivative-free methods for large-scale history matching: the SGSD method, NEWUOA and a newly developed QIM-AG. The SGSD methods mimics the steepest descent method using a stochastic gradient, while NEWUOA approximates the quasi-Newton method using a quadratic interpolation model. QIM-AG is a hybrid of the above two methods, which uses the SGSD stochastic gradient as the first-order coefficients in the quadratic interpolation model, hence reduces the required number of the interpolation points compared to NEWUOA for initial quadratic model construction.

To further simplify the history-matching problem, we applied a simple parameterization method. The reservoir property fields are parameterized as a linear combination of the principal eigenvectors of the prior covariance matrix, which is approximated with a set of unconditional realizations. This not only reduced the parameter space for optimization but also eliminated the need to calculate the inverse of the prior covariance matrix in evaluating the objective function.

All the algorithms are applied to three synthetic reservoir examples: a simple 2D reservoir, the 3D PUNQ reservoir, and the Brugge test reservoir for history matching as well as uncertainty quantification using the RML method.

In obtaining the MAP estimate during history matching, QIM-AG results in a much more rapid decrease in the objective function at early iterations and converges in fewer iterations than NEWUOA. Both QIM-AG and NEWUOA exhibit better convergence than the SGSD algorithm for the first two reservoir test cases, but they all perform similarly in the more realistic Brugge reservoir test case. All algorithms greatly improved the production data match in most wells.

The QIM-AG is applied to sample the posterior pdf using the RML method for the PUNQ and Brugge reservoir examples. The algorithm could effectively reduce the uncertainty of the cumulative production forecast by assimilating the production data.

Acknowledgements The support of the member companies of TUPREP is very gratefully acknowledged. Hui Zhao would like to express his gratitude to the China Scholarship Council and China Important National Science & Technology Specific Projects (2011ZX05014) for the generous financial support of the research and to Yang Li (China University of Petroleum) for his help in this study.

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. Abacioglu, Y., Oliver, D.S., Reynolds, A.C.: Efficient reservoir history matching using subspace vectors. Comput. Geosci. **5**(2), 151–172 (2001)
3. Barker, J.W., Cuypers, M., Holden, L.: Quantifying uncertainty in production forecasts: another look at the PUNQ-S3 problem. SPE J. **6**(4), 433–441 (2001)
4. Bangerth, W., Klie, H., Wheeler, M., Stoa, P., Sen, M.: On optimization algorithm for the reservoir oil well placement problem. Comput. Geosci. **10**, 303–319 (2006)
5. Beckner, B., Song, X.: Field development planning using simulated annealing—optimal economic well scheduling and placement. SPE Annual Technical Conference and Exhibition, Dallas, TX (1995)
6. Bissell, R.: Calculating optimal parameters for history matching. In: 4th European Conference on the Mathematics of Oil Recovery, Roros, Norway (1994)
7. Bissell, R., Dubrule, O., Lamy, P., Swaby, P., Lepine, O.: Combining goestatistical modelling with gradient information for history matching: the pilot point method, SPE 38730. In: Proc. of the SPE Annual Technical Conference and Exhibition, San Antonio, TX (1997)
8. Brouwer, D., Jansen, J.: Dynamic optimization of water flooding with smart wells using optimial control theory. SPE J. **9**(4), 391–402 (2004)
9. Chavent, G.M., Dupuy, M., Lemonnier, P.: History matching by use of optimal control theory. SPE J. **15**(1), 74–86 (1975)
10. Chen, W.H., Gavalas, G.R., Seinfeld, J.H., Wasserman, M.L.: A new algorithm for automatic history matching. SPE J. **14**(6), 593–608 (1974)
11. Chen, C., Wang, Y., Li, G., Reynolds, A.C.: Closed-loop reservoir management on the Brugge test case. Comput. Geosci. **14**(4), 691–703 (2010)
12. Chen, Y., Oliver, D.S.: Ensemble-based closed-loop reservoir management (SPE-118926). In: Proceedings of SPE Reservoir Simulation Symposium (2009)
13. Chen, Y., Oliver, D.S., Zhang, D.: Efficient ensemble-based closed-loop production optimization. SPE J. **14**(4), 634–645 (2009)
14. Cheng, H., Datta-Gupta, A., He, Z.: Acomparison of traveltime and amplitude matching for field-scale production-data integration: sensitivity, nonlinearity, and practical implications. SPE J. **10**(1), 75–90 (2005)
15. Christie, M., MacBeth, C., Subbey, S.: Multiple history-matched models for Teal South. The Leading Edge **21**(3), 286–289 (2002)
16. Conn, A.R., Toint, Ph.L.: An algorithm using quadratic interpolation for unconstrained derivative free optimization. In: Di Pillo, G., Gianessi, F. (eds.) Nonlinear Optimization and Applications, pp. 27–47. Plenum, New York (1996)
17. Conn, A.R., Schienberg, K., Toint, Ph.L.: Recent progress in unconstrained nonlinear optimization without derivatives. Math. Program. **79**, 397–414 (1997)
18. Conn, A.R., Scheinberg, K., Vicente, L.N.: Introduction to Derivative-Free Optimization. MPS-SIAM Book Series on Optimization, SIAM, Philadelphia (2009)
19. Custsodio, A.L., Vicente, L.N.: Using sampling and simplex derivatives in pattern search methods. SIAM J. Optim. **18**, 537–555 (2007)
20. Deutsch, C.V., Journel, A.G.: The application of simulated annealing to stochastic reservoir modeling. SPE Advanced Technology Series **2**(2), 222–227 (1994)
21. Do, S.: Application SPSA-type algorithms to production optimization. Ph.D. thesis, The University of Tulsa, Tulsa, Oklahoma (2012)
22. Emerick, A., Silva, E., Messer, B., Almeida, L., Szwarcman, D., Pacheco, M., Vellasco, M.: Well placement optimization using a genetic algorithm with nonlinear constraints. SPE-99690. In: Proceedings of SPE Reservoir Simulation Symposium, The Woodlands, TX, USA (2009)
23. Eydinov, D., Aanonsen, S.I., Haukas, J., Aavatsmark, I.: A method for automatic history matching of a compositional reservoir simulator with multipoint flux approximation. Comput. Geosci. **12**(2), 209–225 (2008)
24. Fasano, G., Morales, J.L., Nocedal, J.: On the geometry phase in model-based algorithms for derivative-free optimization. Optim. Methods Softw. **24**(1), 145–154 (2009)
25. Floris, F.J.T., Bush, M.D., Cuypers, M., Roggero, F., Syversveen, A.-R.: Methods for quantifying the uncertainty of production forecasts: a comparative study. Petroleum Geoscience **7**, S87–S96 (2001)
26. Gao, G.: Data integration and uncertainty evaluation for large scale automatic history matching problems. Ph.D. thesis, University of Tulsa (2005)
27. Gao, G., Reynolds, A.C.: An improved implementation of the LBFGS algorithm for automatic history matching. SPE J. **11**(1), 5–17 (2006)
28. Gao, G., Zafari, M., Reynolds, A.C.: Quantifying uncertainty for the PUNQ-S3 problem in a Bayesian setting with RML and EnKF. SPE J. **11**(4), 506–515 (2006)
29. Gao, G., Li, G., Reynolds, A.C.: A stochastic optimization algorithm for automatic history matching. SPE J. **12**(2), 196–208 (2007)
30. Gavalas, G.R., Shah, P.C., Seinfeld, J.H.: Reservoir history matching by Bayesian estimation. SPE J. **16**(6), 337–350 (1976)
31. Hajizadeh, Y., Christie, M., Demyanov, V.: Ant colony optimization for history matching. SPE 121193. In: Proceedings of the EUROPEC/EAGE conference and Exhitbition, Amsterdam, The Netherlands (2009)
32. He, Z., Yoon, S., Datta-Gupta, A.: Streamline-based production data integration with gravity and changing field conditions. SPE J. **7**(4), 423–436 (2002)
33. Jacquard, P., Jain, C.: Permeability distribution from field pressure data. SPE J. **5**(4), 281–294 (1965)
34. Jafarpour, B., McLaughlin, D.B.: History matching with an ensemble Kalman filter and discrete cosine parameterization. Comput. Geosci. **12**(2), 227–244 (2008)
35. Jafarpour, B., Goyal, V.K., McLaughlin, D.B., Freeman, W.T.: Exploiting transform-domain sparsity in reservoir history matching. SPE 117819. In: Proceedings of the 2009 SPE Simulation Symposium, The Woodlands, TX (2009)
36. Jafarpour, B., Goyal, V.K., McLaughlin, D.B., Freeman, W.T.: Compressed history matching: exploiting transformdo-

main sparsity for regularization of nonlinear dynamic data integration problems. Math. Geosc. **42**(1), 1–27 (2010)
37. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings IEEE International Conference on Neural Networks, IV. pp. 1942–1948. IEEE Service Ceneter, Piscataway, NJ (1995)
38. Kitanidis, P.K.: Parameter uncertainty in estimation of spatial functions: Bayesian estimation. Water Resour. Res. **22**(4), 499–507 (1986)
39. Kolda, T.K., O'Leary, D.P., Nazareth, L.: BFGS with update skipping and varying memory. SIAM J. Optim. **8**(4), 1060–1083 (1998)
40. Kolda, T.G., Lewis, R.M., Torczon, V.: Optimization by direct search: new perspectives on some classical and modern methods. SIAM REV. **45**(3), 385–482 (2003)
41. Kraaijevanger, J.F.B.M., Egberts, P.J.P., Valstar, J.R., Buurman, H.W.: Optimal waterflood design using the adjoint method. SPE 105764. In: Proceedings of the SPE Reservoir Simulation Symposium (2007)
42. Li, R., Reynolds, A.C., Oliver, D.S.: Sensitivity coefficients for three-phase flow history matching. J. Canadian Pet. Tech. **42**(4), 70–77 (2003)
43. Li, R., Reynolds, A.C., Oliver, D.S.: History matching of three-phase flow production data. SPE J. **8**(4), 328–340 (2003)
44. Li, G., Reynolds, A.C.: Iterative ensemble Kalman filters for data assimilation. SPE J. **14**(3), 496–505 (2009)
45. Li, G., Reynolds, A.C.: Uncertainty quantification of reservoir performance predictions using a stochastic optimization algorithm. Comput. Geosci. **15**(2), 421–429 (2011)
46. Marazzi, M., Nocedal, J.: Wedge trust region methods for derivative free optimization. In: Report on Optimization Technology Center, Northwestern University, Evanston (2000)
47. Mohamed, L., Christie, M., Demyanov, V.: Application of Particle Swarms for history matching in the Brugge reservoir. SPE 135264, SPE. In: Proceedings of Annual Technical Conference and Exhibition, Florence, Italy (2010)
48. Moré, J.J., Sorensen, D.: Computing a trust region step. SIAM J. Sci. **4**, 553–572 (1983)
49. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (2006)
50. Oeuvray, R.: Trust-region methods based on radial basis functions with application to biomedical imaging. Ph.D. thesis, EPF, Lausanne, Switzerland (2005)
51. Oldenberg, D.W., McGillivray, P.R., Ellis, R.G.: Generalizaed subspace methods for large-scale inverse problems. Geophysical Journal International **114**(1), 12–20 (1993)
52. Oliver, D.S., He, N., Reynolds, A.C.: Conditioning Permeability Fields to Pressure Data. European Conference for the Mathematics of Oil Recovery V (1996)
53. Oliver, D.S.: Multiple realizations of the permeability field from well-test data. SPE J. **1**(2) 145–154 (1996)
54. Oliver, D.S., Reynolds, A.C., Liu, N.: Inverse Theory for Petroleum Reservoir Characterization and History Matching. Cambridge University Press, Cambridge (2008)
55. Oliver, D.S., Chen, Y.: Recent progress on reservoir history matching: a review. Comput. Geosci. **15**, 185–221 (2011)
56. Onwunalu, J.E., Durlofsky, L.J.: A new well-pattern-optimization procedure for large-scale field development. SPE J. **16**(3), 594–607 (2011)
57. Peters, L., Arts, R.J., Brouwer, G.K., Geel, C.R., Cullick, S., Lorentzen, R.J., Chen, Y., Dunlop, K.N.B., Vossepoel, F.C., Xu, R., Sarma, P., Alhuthali, A.H., Reynolds, A.C.: Results of the Brugge benchmark study for flooding optimization and history matching. SPE Reserv. Eval. Eng. **13**(3), 391–405 (2010)
58. Powell, M.J.: UOBYQA: unconstrained optimization by quadratic approximation. Math. Program. **92**, 555–582 (2002)
59. Powell, M.J.: Least Frobenius norm updating of quadratic models that satisfy interpolation conditions. Math. Program. **100**, 183–215 (2004)
60. Powell, M.J.: The NEWUOA software for unconstrained optimization without derivatives in large-scale nonlinear optimization. In: Di Pillo, G., Roma, M. (eds.) Large-Scale Nonlinear Optimization. Springer, Netherlands (2006)
61. Powell, M.J.: Developments of NEWUOA for unconstrained minimization without derivatives. Technical Report DAMTP NA2004/08, Department of Applied Mathematics and Theoretical Physics, Cambridge University, Cambridge, UK (2007)
62. Powell, M.J.: The BOBYQA algorithm for bound constrained optimization without derivatives. http://www6.cityu.edu.hk/rcms/publications/preprint26.pdf. Optimization, NA2009/06, (2009)
63. Reynolds, A.C., He, N., Chu, L., Oliver, D.S.: Reparameterization techniques for generating reservoir descriptions conditioned to variograms and well-test pressure data. SPE J. **1**(4), 413–426 (1996)
64. Rodrigues, J.R.P.: Calculating derivatives for automatic history matching. Comput. Geosci. **10**, 119–136 (2006)
65. Romero, C.E., Carter, J.N., Zimmerman, R.W., Gringarten, A.C.: Modified genetic algorithm for reservoir characterization, SPE-64765. In: Proceedings of SPE Annual Technical Conference and Exhibition, Dallas, TX (2000)
66. Sarma, P., Durlofsky, L., Aziz, K., Chen, W.: Efficient real-time reservoir management using adjoint-based optimal control and model updating. Comput. Geosci. **10**, 3–36 (2006)
67. Sarma, P., Durlofsky, L., Aziz, K.: A new approach to automatic history matching using kernel PCA. SPE 106176. In: Proceedings of the 2007 SPE Simulation Symposium, Houston TX (2007)
68. Shah, P.C., Gavalas, G.R., Seinfeld, J.H.: Error analysis in history matching: the optimum level of parameterization. SPE J. **18**(6), 219–228 (1978)
69. Spall, J.C.: A stochastic approximation technique for generating maximum likelihood parameter estimates. In: Proceedings of the American Control Conference, pp. 1161–1167. Minneapolis, MN (1987)
70. Spall, J.C.: Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE Trans. Automat. Control **37**, 332–341 (1992)
71. Spall, J.C.: Implementation of the simultaneous perturbation algorithm for stochastic optimization. IEEE T. Aero. Eclec. Sys. **34** 817–823 (1998)
72. Tavakoli, R., Reynolds, A.C.: History matching with parameterization based on the SVD of a dimensionless sensitivity matrix. SPE J. **15**(2), 495–508 (2010)
73. van Essen, G., Zandvliet, M., den Hof, P.V., Bosgra, O., Jansen, J.: Robust water flooding optimization of multiple geological scenarios. SPE J. **14**(1), 202–210 (2009)
74. Vasco, D.W., Yoon, S., Datta-Gupta, A.: Integrating dynamic data into high-resolution reservoir models using streamline-based analytic sensitivity coefficients. SPE J. **4**(4), 389–399 (1999)
75. Wang, C., Li, G., Reynolds, A.C.: Production optimization in closed-loop reservoir managment. SPE J. **14**(3), 506–523 (2009)

76. Wu, Z., Reynolds, A.C., Oliver, D.S.: Conditioning geostatistical models to two-phase production data. SPE J. **4**(2), 142–155 (1999)
77. Zafari, M., Reynolds, A.C.: Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter. SPE J. **12**(3), 382–391 (2007)
78. Zhang, D., Lu, Z., Chen, Y.: Dynamic reservoir data assimilation with an efficient, dimension-reducd Kalman filter. SPE J. **12**(8), 108–117 (2007)
79. Zhao, H., Chen, C., Do, S., Li, G., Reynolds, A.C.: Maximization of a dynamic quadratic model for production optimization. SPE-141317. In: Proceedings of the 2011 SPE Simulation Symposium, Houston TX (2011)