V European Conference on Computational Fluid Dynamics
ECCOMAS CFD 2010
J. C. F. Pereira and A. Sequeira (Eds)
Lisbon, Portugal,14-17 June 2010

MODEL-REDUCED GRADIENT-BASED HISTORY MATCHING

M.P. Kaleta*, R.G. Hanea†, A.W. Heemink††, J.D. Jansen †††

* Delft University of Technology
Mekelweg 4, 2628 CD, Delft, The Netherlands
e-mail: m.p.kaleta@tudelft.nl
† TNO Built Environment and Geosciences, Business Unit Geo Energy and Geo Information
Princetonlaan 6, 3584 CB, Utrecht, The Netherlands
†† Delft University of Technology
Mekelweg 4, 2628 CD, Delft, The Netherlands
††† Delft University of Technology and Shell Int. E&P
Stevinweg 1, 2628 CN, Delft, The Netherlands

Key words: history matching, proper orthogonal decomposition, balanced proper orthogonal decomposition, reservoir model

Abstract. *In typical simulation models for subsurface flow in oil and gas reservoirs many geological parameters are uncertain. Their estimated values can sometimes be improved by using surface and downhole production data in a process called history matching. History matching can be performed very efficiently by gradient-based algorithms. Those methods require, however, the implementation of an adjoint method. In reservoir models, the Jacobian matrices of the model with respect to state variables are usually available. Nevertheless the implementation of the adjoint method is an immense programming effort because of the need to derive the Jacobians of the model with respect to the parameters, which are required for the gradient computation. We propose a gradient-based history matching algorithm which is based on model reduction, where the original (nonlinear and high-order) forward model is replaced by a linear reduced-order forward model. Consequently, the adjoint of the tangent linear model is replaced by the adjoint of a linear reduced-order forward model. Due to the linear character of the reduced-order model, the corresponding adjoint model is easily obtained. The gradient of the objective function is approximated and the minimization problem is solved in the reduced space; the procedure is iterated with the updated estimate of the parameters if necessary. The reduced-order model is constructed with the aid of the conventional proper orthogonal decomposition (POD) method and the balanced POD method. The conventional POD-based approach is adjoint-free and can be used with any reservoir simulator. The balanced POD-based method requires the adjoint states but it is free from the Jacobians with respect to parameters. The methods were evaluated for a waterflooded reservoir with uncertain permeability field. A comparison with an adjoint-based approach shows that the model-reduced approaches give comparable quality of history matches and predictions. Their computational efficiencies are lower than of an adjoint-based approach but higher than of an approach where the gradients are obtained with simple finite differences.*

1 INTRODUCTION

To design optimal recovery strategies for subsurface oil and gas reservoirs frequent use is made of large-scale numerical simulation methods for flow through heterogeneous porous media. Such *reservoir models* consist of time-and space-discretized systems of nonlinear partial differential equations to describe the time evolution of the state variables, i.e. the spatially varying pressures and water and hydrocarbon accumulations, over the producing life of the reservoir. Typical reservoir models contain up to millions of grid blocks with strongly varying parameter values representing porosity and permeability (i.e. inverse flow resistance) of the subsurface. Simulation is normally performed with fully implicit time stepping schemes in combination with Newton-Raphson iteration to solve the nonlinear equations at every time step. The values of the gridblock parameters, and of other model parameters representing geological features, are usually only known in the neighbourhood of the wells but very uncertain otherwise. However, during the producing life of a reservoir it is sometimes possible to improve the estimate of the parameter values by comparison of measured production data (pressures and oil, water and gas flow rates in the wells) with their model predictions, a process known as *history matching*. History matching identifies the parameter values that minimize an *objective function* that represents the mismatch between modeled and observed production data.[1] Usually the objective function is defined as a sum of weighted squared differences between observed and modeled outputs of the system, and gradient-based routines are used for iterative minimization of its value. A numerically efficient way to calculate the gradients of the objective function with respect to the parameters is the adjoint method. Full implementation of the adjoint method requires the availability of Jacobians matrices of the model with respect to the state variables and the parameters. Usually, in reservoir models, the Jacobians of the system with respect to the states are readily available because they are used in the Newton-Raphson iterations during the forward simulation. Even so, the implementation of the adjoint method to compute the gradients with respect to the parameters is an immense programming effort. This is caused by the need for Jacobians of the system with respect to parameters, which are difficult to derive because of the nonlinear nature of the equations and which are therefore often not available. Moreover, in some simulators there is an adjoint model available for history matching and the gradients for a selected group of parameters, but not necessarily for all. In an earlier publication we proposed a model-reduced gradient-based history matching method where a reduced-order model is constructed with the aid of the proper orthogonal decomposition (POD) method,[2].[3] Here we extend this approach to the use of a balanced POD method. In general, model reduction techniques aim to obtain a low-order approximation to a high-order dynamical system. Depending on the application, the reduced-order model should preserve relevant features of the original model regarding the application. Both proposed reduction methods in our application are data-driven projection-based methods. The POD method is also known as the Karhunen-Loève method, principal component analysis or the method

of empirical orthogonal functions. It was introduced independently by Karhunen[4] and Loève[5] as a statistical tool to analyze random process data. The method was called for the first time the POD by Lumley,[6] when it was used for the study of turbulent flow. In the POD method a low-order projection subspace is determined by processing data obtained from numerical simulations of the high-order model, which are expected to provide the essential information about the dynamic behavior of the system. The high-order equations are projected on the low-order subspace resulting in a low-order model. The balanced POD approach is a combination of the POD method and another well-known reduced order technique called balanced truncation (BT). The concept of balancing was first time presented in the work of Mullis and Roberts[7] on the design of digital filters. Subsequently, balanced truncation was introduced by Moore[8] in the control theory community for stable, linear, input-output systems. It has been extended to nonlinear systems by Lall et al.[9] and Scherpen.[10] The POD and balanced truncation have been combined by Moore,[8] Lall et al[9] and Rowley.[11] Here we use the method proposed by Rowley[11] which is an approximation to balanced truncation with computational costs similar to double the costs of the POD method. In this method we aim to build the reduced-order system with the input-output behavior similar to the original system. We preserve in the reduced-order model the part of the state that is relevant for the input-output behavior of the system. This relevant part of the state can be interpreted in terms of the system-theoretical concepts observability and controllability. Since in the history matching problems the observed production data are very sparse, we expect that the observable part of the state is very limited and therefore a significant reduction can be achieved; see.[12] It is suspected that the detailed description of the pressure and saturation in the reduced-order model may not be important to reconstruct the measurements. In this work we evaluate if an improvement in the computational efficiency can be achieved by neglecting those parts of the states.

Typically, balanced truncation is used to reduce the order of the model describing a controllable system, that is, a system in which input variables are changed in order to reach desirable outputs. The reduced-order model must be capable of reproducing those input-output dependencies.

Here, we adopt the terminology of the optimal control community, and we apply the idea of balancing to the parameter estimation problem. The control variables (input variables) in our problem are the unknown parameters that we want to estimate. Note that this is unlike in the production optimization problem, where the controls are production settings, e.g. controllable well pressures or flow rates. In our case, we know the past production settings with which the reservoir was operated, and we want to find the model that best represents the measured production data with those settings. The outputs in our problem are the simulated production data.

The proposed methods require different data and therefore have different limitations. The balanced POD method requires the adjoint states to be available, but it is free from the Jacobian of the system with respect to model parameters, and therefore can be used to

estimate the parameters for which the gradient is not available yet. The POD method does not require the adjoint model at all and can therefore be used with any simulator. None of the methods does require access to the simulation code, because we only need to collect the states of the model at different time instances for different sets of parameters, and the states of the adjoint model (in case of balanced POD).

2 HISTORY MATCHING AS INVERSE MODELING

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

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

where $\mathbf{x}(t_i) \in X \subset \mathbb{R}^n$ denotes the state vector of pressures and saturations at time t_i, $\boldsymbol{\theta}$ denotes the vector of uncertain parameters and N denotes the total number of simulation time steps. The dynamic operator $\mathbf{f}_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$ represents a nonlinear and deterministic reservoir simulator. Another operator, $\mathbf{h}_i : \mathbb{R}^n \rightarrow \mathbb{R}^m$, describes the relationship between measured production data $\mathbf{y}(t_i)$ and state variables $\mathbf{x}(t_i)$ (see e.g.[13] for further details). If we assume that observations are imperfect, then the simulated measurements are described by

$$\mathbf{y}(t_i) = \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}] + \boldsymbol{\eta}_i, i = 1, ..., N_o, \tag{2}$$

where $\boldsymbol{\eta}_i$ is the observation error and N_o is the number of time steps where the observations are taken. The uncertain parameters in the model can be estimated by minimizing an objective function that measures the difference between simulated data $\mathbf{y}$ and observed data $\mathbf{d}$. In the case of assimilation of production data, the data are sparse and therefore this information is not sufficient to correctly estimate all parameters. The parameter-estimation problem is ill-posed. One way to make the history matching problem well-posed is to rely on some additional information in the form of a prior estimate of the model parameters $\boldsymbol{\theta}^b_{init}$ and their uncertainty $\mathbf{R}_b^{-1}$, see.[14] The objective function then consists of a prior (background) term and an observation term:

$$\begin{aligned} J(\boldsymbol{\theta}) \;=\; & \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^b_{init})^T \mathbf{R}_b^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^b_{init}) + \\ & + \frac{1}{2}\sum_{i=1}^{N_o} \left[\mathbf{d}(t_i) - \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]\right]^T \mathbf{R}_i^{-1} \left[\mathbf{d}(t_i) - \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]\right], \end{aligned} \tag{3}$$

where $\mathbf{R}_i$ is the covariance matrix of the observation errors at time t_i. By minimizing objective function (3) we find a model that is close to the prior model, while it simultaneously minimizes the misfit between the modeled and observed data.

3 METHODOLOGY

Here we use the methodology proposed for the first time by Vermeulen and Heemink[15] and then applied to the history matching problem in,[2,3] It is a gradient-based procedure

which is based on model reduction, where the original (nonlinear and high-order) forward model is replaced by a linear reduced-order forward model. Consequently, the adjoint of the tangent linear approximation of the original forward model is replaced by the adjoint of a linear reduced-order forward model. Due to the linear character of the reduced-order model, the corresponding adjoint model is easily obtained. The gradient of the objective function is approximated and the minimization problem is solved in the reduced space; the procedure is iterated with the updated estimate of the parameters if necessary.
The history matching procedure turns into a scheme that consists of two loops: an inner loop, which finds the minimum of the approximate objective function by using the low-order tangent linear model and an outer loop, where the original model is used to calculate the original objective function and to build the reduced-order model. It iterates until some predefined convergence criteria are met, and can be classified as an incremental approach.[16] The advantage of the proposed method is that we do not need to acquire a tangent linear model or a simplified model, because we construct the reduced-order tangent linear model by using information from perturbations of the original model. To construct the reduced-order model we need to collect data to build the projection matrices and approximate the Jacobians of the reduced-order model along the projection vectors. When the reduced-order model is available, we derive the reduced-order adjoint model and we improve the parameters in reduced space. The minimization is performed using a quasi-Newton optimization method where the Hessian of the objective function is updated using the Broyden Fletcher Goldfarb Shanno (BFGS) method. The solution of the inner loop is an optimum for the reduced-order linearized system but not necessarily for the original one. Therefore, we check the value of the original objective function for the new parameters. The described methodology is depicted by the following chart in Figure 1. For further details see,[2],[3] where this approach, using the POD method, has been applied to the history matching problem.

4 REDUCED ORDER METHODS

The balanced POD method and the POD method fall in the category of projection methods. They involve the projection of the dynamic equations onto a subspace of the original model space. If we define the projection as $\mathbf{\Pi} = \mathbf{\Phi}\mathbf{\Psi}^T$, where $\mathbf{\Phi}, \mathbf{\Psi} \in \mathbb{R}^{n\times k}$, $\mathbf{\Psi}^T\mathbf{\Phi} = \mathbf{I}_k$, then the system described by Equations (1) and (2) is approximated by the reduced-order system in the following form

$$\mathbf{z}(t_i) = \mathbf{\Psi}^T\mathbf{f}_i(\mathbf{\Phi}\mathbf{z}(t_{i-1}), \boldsymbol{\theta}) \tag{4}$$

$$\mathbf{y}(t_i) = \mathbf{h}_i(\mathbf{\Phi}\mathbf{z}(t_i), \boldsymbol{\theta}). \tag{5}$$

In the used methodology we reduce the order of the tangent linear approximation of the original model derived around the background trajectory $\mathbf{x}^b(t_i)$ corresponding to the best

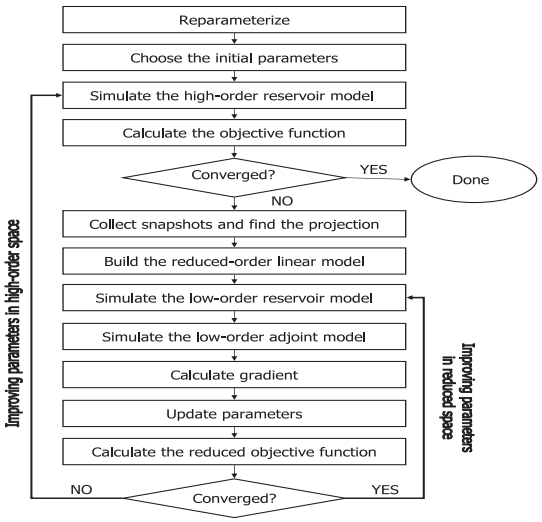


Figure 1: Model-reduced gradient-based history matching algorithm.

current estimate $\boldsymbol{\theta}^b$. In this case the reduced-order system is

$$\bar{\mathbf{z}}(t_i) = \boldsymbol{\Psi}^T \frac{\partial \mathbf{f}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)]}{\partial \mathbf{x}(t_{i-1})} \boldsymbol{\Phi} \bar{\mathbf{z}}(t_{i-1}) + \boldsymbol{\Psi}^T \frac{\partial \mathbf{f}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)]}{\partial \theta} \Delta\theta \tag{6}$$

$$\mathbf{y}(t_i) = \mathbf{h}_i[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b] + \frac{\partial \mathbf{h}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)]}{\partial \mathbf{x}(t_i)} \boldsymbol{\Phi} \bar{\mathbf{z}}(t_i) + \frac{\partial \mathbf{h}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)]}{\partial \boldsymbol{\theta}} \boldsymbol{\theta}. \tag{7}$$

The difference between POD and balanced POD is in the way how those projection matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are constructed and the kind of data that is used to find them.

4.1 PROPER ORTHOGONAL DECOMPOSITION

The data required for POD is built from the selected state vectors of a forward simulation of the large-scale numerical model, called *snapshots*. Since the reduced-order model is used for parameter estimation, the snapshots should be able to represent the behavior of the system for modified parameter values. Therefore snapshots are created in the following way:

$$\mathbf{x}_j(t_i) = \mathbf{f}_i[\mathbf{x}^b(t_{i-1}), \boldsymbol{\theta}^b + \mathbf{i}_j \Delta\theta_j] - \mathbf{f}_i[\mathbf{x}^b(t_{i-1}), \boldsymbol{\theta}^b], \tag{8}$$

where $\mathbf{i}_j$ is a zero vector with 1 on the jth coordinate. Then snapshots are put as columns in matrix

$$\mathbf{X} = \{\mathbf{x}_1(t_1), ..., \mathbf{x}_1(t_N), ..., \mathbf{x}_p(t_1), ..., \mathbf{x}_p(t_N)\} \tag{9}$$

of size $n \times s$, where s is the total number of snapshots and since typically $s \ll n$ the reduced eigenvalue problem

$$(\mathbf{X}^T \mathbf{X})\mathbf{v}_i = \gamma_i \mathbf{v}_i; \quad i \in \{1, ..., s\} \tag{10}$$

is solved. This results in s non-negative eigenvalues γ_i and s corresponding eigenvectors $\mathbf{v}_i$ which are transformed into eigenvectors $\boldsymbol{\phi}_i$ of $\mathbf{X}\mathbf{X}^T$ by $\mathbf{X}\mathbf{V}\boldsymbol{\Lambda}^{-\frac{1}{2}}$, where $\boldsymbol{\Lambda}$ is a matrix with eigenvalues γ_i on the main diagonal and $\mathbf{V}$ is the matrix consisting of all eigenvectors $\mathbf{v}_i$. The projection $\boldsymbol{\Pi} = \boldsymbol{\eta}\boldsymbol{\Phi}^T$is defined as Galerkin projection that is $\boldsymbol{\Phi} = \boldsymbol{\eta}$, where $\boldsymbol{\Phi}$ consist of $\boldsymbol{\phi}_i$ corresponding to the leading eigenvalues. The reduced-order system is then described by

$$\bar{\mathbf{z}}(t_i) = \boldsymbol{\Phi}^T \frac{\partial \mathbf{f}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)]}{\partial \mathbf{x}(t_{i-1})} \boldsymbol{\Phi}\bar{\mathbf{z}}(t_{i-1}) + \boldsymbol{\Phi}^T \frac{\partial \mathbf{f}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)]}{\partial \theta} \Delta\theta \tag{11}$$

$$\mathbf{y}(t_i) = \mathbf{h}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)] + \frac{\partial \mathbf{h}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)]}{\partial \mathbf{x}(t_i)} \boldsymbol{\Phi}\bar{\mathbf{z}}(t_i) + \frac{\partial \mathbf{h}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)]}{\partial \boldsymbol{\theta}} \boldsymbol{\theta}. \tag{12}$$

This reduced-order system preserves the states which are evident in the set of snapshots.

4.2 BALANCED PROPER ORTHOGONAL DECOMPOSITION

The controllability and observability of a dynamic system reflect to what extent the input influences the states, and to what extent the states influence the output. We can express controllability and observability of a linear dynamic system with the aid of Gramians, which are square and symmetric matrices defined in terms of the systme properties only see e.g.[17] Balancing is a technique to combine the controllability and observability properties of a system by finding a linear combination of the states that produces an equivalent model with equal and diagonal controllability and observability Gramians. The equality of the Gramians implies that combinations of states that are difficult to reach are simultaneously difficult to observe. The equality of the Gramians implies that combinations of states that are difficult to reach are simultaneously difficult to observe. The diagonal entries of the balanced Gramians, known as Hankel singular values, reflect the combined controllability and observability of individual linear combinations of states in the system. Accordingly, they give a measure of the energy of each individual combination, or its contribution to the input-output behavior. Therefore, Hankel singular values can be used to identify those combinations of the states that represent the most important input-output characteristics of the system. A reduced-order system representation, known as balanced truncation, can therefore be obtained by expressing the system dynamics in terms of a limited number of linear combinations of the states (Hankel singular vectors) corresponding to the leading Hankel singular values. Unfortunately, computation of the Gramians for large systems is computationally very demanding, if possible at all. Balanced POD aims to obtain an approximation to balanced truncation which is computationally feasible for large systems. In linear systems the Gramians are computed by solving Lyapunov equations or they are computed from the data from numerical simulations. When the balanced truncation is generalized to nonlinear systems empirical Gramians may be used. Empirical Gramians were proposed by Moore[8] and were used by Lall et al.[18] Here, however, we follow the approach of Rowley,[11] and we compute the bal-

ancing transformation directly from snapshots of empirical Gramians, without computing the Gramians themselves. Similar approaches were used by[19] and[17] (Section 9.1).

- *Controllability Gramian using snapshots*
  To compute the empirical controllability Gramian for a system with p parameters, the state responses to the perturbation of the parameters are collected according to

$$\mathbf{x}_j(t_i) = \mathbf{f}_i[\mathbf{x}^b(t_{i-1}), \boldsymbol{\theta}^b + \mathbf{i}_j \Delta\theta_j] - \mathbf{f}_i[\mathbf{x}^b(t_{i-1}), \boldsymbol{\theta}^b], \tag{13}$$

  where $\mathbf{i}_j$ is a zero vector with 1 on the jth coordinate, and are put as columns in matrix

$$\mathbf{X} = \{\mathbf{x}_1(t_1), ..., \mathbf{x}_1(t_N), ..., \mathbf{x}_p(t_1), ..., \mathbf{x}_p(t_N)\} \tag{14}$$

  of size $n \times s$, where s is the total number of snapshots. We will show later that the controllability Gramian can be approximated by

$$\mathbf{W}_c = \mathbf{X}\mathbf{X}^T. \tag{15}$$

  This implies that the POD patterns for the dataset of perturbation responses are the eigenvectors of $\mathbf{W}_c$ corresponding to the largest eigenvalues. I.e., the POD patterns are the most controllable patterns for the given set of parameter perturbations.

- *Observability Gramian using snapshots*
  To compute the empirical observability Gramian we compute the impulse responses of the adjoint model

$$\begin{aligned} \boldsymbol{\lambda}(t_i) \quad = \quad & \left(\frac{\partial \mathbf{f}_{i+1}[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_i)}\right)^T \boldsymbol{\lambda}(t_{i+1}) + \\ & \left(\frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_i)}\right)^T \mathbf{R}_i^{-1} \left[\mathbf{d}_j(t_i) - \mathbf{h}_i^j[\mathbf{x}(t_i), \boldsymbol{\theta}]\right] \end{aligned} \tag{16}$$

  for $i = N, ..., 1$ with an end condition $\boldsymbol{\lambda}(t_{N+1}) = \mathbf{0}$ and we put as columns in matrix

$$\mathbf{Y} = \{\boldsymbol{\lambda}(t_1), ..., \boldsymbol{\lambda}_1(t_N), ..., \boldsymbol{\lambda}_p(t_1), ..., \boldsymbol{\lambda}_p(t_N)\}. \tag{17}$$

  Then the empirical observability Gramian is approximated by

$$\mathbf{W}_c = \mathbf{Y}\mathbf{Y}^T. \tag{18}$$

  The adjoint variable defined by (16) can be interpreted as a multiplier which informs us about the relevance of each state in the calculation of the objective function. The objective function is simply a weighted square of the outputs of the system. I.e. the

adjoint variables also inform us which states are relevant in the calculation of the outputs. If we call $\mathbf{v}_i = \mathbf{R}_i^{-1}\left[\mathbf{d}_j(t_i) - \mathbf{h}_i^j[\mathbf{x}(t_i), \boldsymbol{\theta}]\right]$ then

$$\boldsymbol{\lambda}(t_i) \;=\; \left(\frac{\partial \mathbf{f}_{i+1}[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_i)}\right)^T \boldsymbol{\lambda}(t_{i+1}) + \left(\frac{\partial \mathbf{h}_i[\mathbf{x}(t_i), \boldsymbol{\theta}]}{\partial \mathbf{x}(t_i)}\right)^T \mathbf{v}_i$$

is the adjoint of the tangent linear model. In theory, we should excite the system and the adjoint system, with impulses to obtain a best opproximation to the true Gramians. However, because the measurements have different orders of magnitude, an impulse response of the adjoint model is not recommended. Moreover, since we analyze the past data, the outputs of the 'true' system are known. Therefore, we excite the adjoint model with the observed data weighted with the measurement covariance matrix. The reduced-order model is supposed to approximate the behavior of the system in the past, where the output data is known. Hence, we do not need to preserve the part of the state corresponding to the data that were not observed in that period of time (for example if there was no water production we do not take it as excitation of the adjoint model). If N_o is the number of time instances where the outputs are observed and m is the number of measurement at each of those time instances then this approach requires $N_o \times m$ integrations of the adjoint system. Since observations are sparse we take all measurements available at a certain point in time together and we integrate the adjoint model once. In our case the adjoint state takes non-zero values from the time instance at which the observation is included, and it is integrated backward in time to quantify which part of the state is relevant in the gradient calculation. We include all time instances at once as well. Because the system is nonlinear with respect to parameters we collect snapshots using adjoint models corresponding to the forward models that were used to collect snapshots for controllability Gramian.

- *Balanced POD using the method of snapshots*
  The controllability and observability Gramians describe the variation in the states due to changes in the parameter values (inputs) and changes in the observed data (outputs), respectively. By performing a singular value decomposition (SVD) on those matrices we obtain the dominant patterns of the controllable and observable parts of the state. If we write that $\mathbf{W}_c = \mathbf{X}\mathbf{X}^T$ and $\mathbf{W}_o = \mathbf{Y}\mathbf{Y}^T$, then we have square roots of the Gramians. In the method of snapshots used here, the balancing patterns are computed by solving the SVD of the matrix $\mathbf{Y}^T\mathbf{X}$, which gives

$$\mathbf{Y}^T\mathbf{X} = \mathbf{W}\boldsymbol{\Sigma}\mathbf{V}^T = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_2 \end{bmatrix} \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix}. \tag{19}$$

  where $\mathbf{Y}^T\mathbf{X}$ has dimensions of columns of $\mathbf{X}$ times columns of $\mathbf{Y}$, that is, the number of snapshots collected for controllability Gramian times the number of snapshots

collected for the observability Gramian. The Gramians themselves never need to be computed and only a single SVD is needed. Since usually the number of snapshots is much smaller than the number of states, this is computationally efficient. The computation time required to build the projection subspace is about twice more expensive than in the case of the POD method. In this approach we need to generate snapshots of the adjoint model which is computationally as expensive as generating snapshots of the forward model.
After the SVD of matrix $\mathbf{Y}^T\mathbf{X}$ is computed, the balancing transformation is given by the matrix:

$$\mathbf{T} = \mathbf{\Sigma}^{-1/2}\mathbf{W}_1^T\mathbf{Y}^T \tag{20}$$

and its inverse is given by:

$$\mathbf{T}^{-1} = \mathbf{X}\mathbf{V}_1\mathbf{\Sigma}^{-1/2}. \tag{21}$$

It can be seen that $\mathbf{T}^{-1}$ is an inverse of $\mathbf{T}$, namely

$$\mathbf{T}^{-1}\mathbf{T} = \mathbf{\Sigma}^{-1/2}\mathbf{W}_1^T\mathbf{Y}^T \cdot \mathbf{X}\mathbf{V}_1\mathbf{\Sigma}^{-1/2} = \mathbf{\Sigma}^{-1/2}\mathbf{V}_1^T\mathbf{W}_1\mathbf{\Sigma}\mathbf{V}_1^T\mathbf{V}_1\mathbf{\Sigma}^{-1/2} = \mathbf{I}, \tag{22}$$

because $\mathbf{W}_1^T\mathbf{W}_1 = \mathbf{I}$ and $\mathbf{V}_1^T\mathbf{V}_1 = \mathbf{I}$. Next, we have to show now that the derived transformation results in the balanced representation of the controllability and observability Gramians, that is:

$$\mathbf{T}\mathbf{W_c}\mathbf{T}^T = \mathbf{T}^{-T}\mathbf{W_o}\mathbf{T}^{-1} = \mathbf{\Sigma}. \tag{23}$$

For the controllability Gramian we have

$$\begin{aligned}\left(\mathbf{\Sigma}^{-1/2}\mathbf{W}_1^T\mathbf{Y}^T\right)\mathbf{X}\mathbf{X}^T\left(\mathbf{\Sigma}^{-1/2}\mathbf{W}_1^T\mathbf{Y}^T\right)^T = \mathbf{\Sigma}^{-1/2}\mathbf{W}_1^T(\mathbf{Y}^T\mathbf{X})(\mathbf{Y}^T\mathbf{X})^T\mathbf{\Phi}_1\mathbf{\Sigma}^{-1/2} = \\ \mathbf{\Sigma}^{-1/2}\mathbf{W}_1^T(\mathbf{W}\mathbf{\Sigma}\mathbf{V}^T)(\mathbf{W}\mathbf{\Sigma}\mathbf{V}^T)^T\mathbf{W}_1\mathbf{\Sigma}^{-1/2} = \mathbf{I}.\end{aligned} \tag{24}$$

In similar way we can show that

$$\left(\mathbf{X}\mathbf{V}_1\mathbf{\Sigma}^{-1/2}\right)^T\mathbf{Y}\mathbf{Y}^T\left(\mathbf{X}\mathbf{V}_1\mathbf{\Sigma}^{-1/2}\right) = \mathbf{I}. \tag{25}$$

The projection $\mathbf{\Pi} = \mathbf{\Psi}\mathbf{\Phi}^T$is defined as Petrov-Galerkin projection, where $\mathbf{\Psi}$ consists of the columns of matrix $\mathbf{T}$ corresponding to the largest Hankel singular values that we want to retain, and $\mathbf{\Phi}$ consists of the first n_r rows of $\mathbf{T}^{-1}$. The reduced-order system is then described by

$$\bar{\mathbf{z}}(t_i) = \mathbf{\Psi}^T\frac{\partial\mathbf{f}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)]}{\partial\mathbf{x}(t_{i-1})}\mathbf{\Phi}\bar{\mathbf{z}}(t_{i-1}) + \mathbf{\Psi}^T\frac{\partial\mathbf{f}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)]}{\partial\theta}\Delta\theta \tag{26}$$

$$\mathbf{y}(t_i) = \mathbf{h}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)] + \frac{\partial\mathbf{h}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)]}{\partial\mathbf{x}(t_i)}\mathbf{\Phi}\bar{\mathbf{z}}(t_i) + \frac{\partial\mathbf{h}_i[\boldsymbol{\theta}^b, \mathbf{x}^b(t_i)]}{\partial\boldsymbol{\theta}}\boldsymbol{\theta}. \tag{27}$$

This reduced-order system preserves the states which are easier to control and at the same time easier to observe.

- *Discussion*
  If we consider the tangent linear approximation of the original system

$$\Delta\bar{\mathbf{x}}(t_i) = \frac{\partial \mathbf{f}_i[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b]}{\partial \mathbf{x}(t_{i-1})}\Delta\bar{\mathbf{x}}(t_{i-1}) + \frac{\partial \mathbf{f}_i[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b]}{\partial \theta}\Delta\theta, \tag{28}$$

  then its adjoint model is

$$\boldsymbol{\lambda}(t_i) = \frac{\partial \mathbf{f}_{i+1}[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b]}{\partial \mathbf{x}(t_i)}\boldsymbol{\lambda}(t_{i+1}) + \frac{\partial \mathbf{h}(t_i)[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b]}{\partial \mathbf{x}(t_i)}\mathbf{v}(t_i). \tag{29}$$

  Projecting those equations by using $\boldsymbol{\Pi}$ we obtain a reduced-order linear model

$$\mathbf{z}(t_i) = \boldsymbol{\Psi}^T \frac{\partial \mathbf{f}_i[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b]}{\partial \mathbf{x}(t_{i-1})}\boldsymbol{\Phi}\mathbf{z}(t_{i-1}) + \boldsymbol{\Psi}^T \frac{\partial \mathbf{f}_i[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b]}{\partial \theta}\Delta\theta. \tag{30}$$

  Then, the adjoint model of the reduced-order linearized model has the following form

$$\boldsymbol{\gamma}(t_i) = \left(\boldsymbol{\Psi}^T \frac{\partial \mathbf{f}_{i+1}[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b]}{\partial \mathbf{x}(t_i)}\boldsymbol{\Phi}\right)^T \boldsymbol{\gamma}(t_{i+1}) + \left(\frac{\partial \mathbf{h}(t_i)[\mathbf{x}^b(t_i), \boldsymbol{\theta}^b]}{\partial \mathbf{x}(t_i)}\boldsymbol{\Phi}^T\right)^T \boldsymbol{\Psi}(t_i). \tag{31}$$

  One can notice that if we assume that $\boldsymbol{\lambda} \approx \boldsymbol{\Psi}\boldsymbol{\gamma}$, then the projection of the original adjoint model corresponds to the adjoint of the reduced-order model. This does not have to be true if the external projection $\boldsymbol{\Psi}$ is chosen only based on the reconstruction of the forward model. Patterns that accurately represent the original state do not necessary accurately represent the adjoint state. Therefore, we should aim to find a projection that represents both the original forward model and the adjoint model with acceptable accuracy. This is exactly what is done in balanced POD where we reduce the model by constructing the projection matrix $\boldsymbol{\Pi}$ using both the forward and the adjoint states.

5 NUMERICAL EXPERIMENTS

In this experiment the two methods introduced above are used to estimate the parameters of a simple 2D reservoir model, which describes iso-thermal slightly compressible two-phase (oil-water) flow in a five-spot configuration.

- *Model assumptions:*
  We assume that reservoir operates under water flooding condition. During that process the dispersive effect of geological heterogeneities and the diffusive effect of capillary pressure take place. Additionally to diffusion and dispersion caused by physical phenomena, diffusion caused by the numerical errors of the discretization occurs. Since in many cases this numerical diffusion is of the same order of

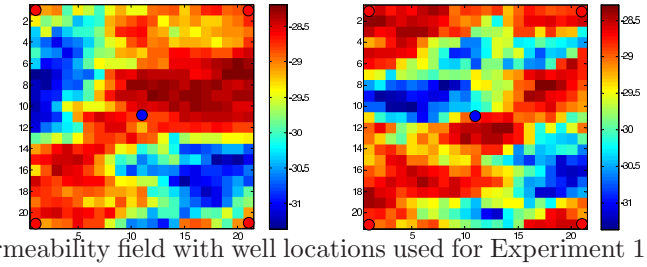


Figure 2: The reservoir permeability field with well locations used for Experiment 1 (left) and Experiment 2 (right). The red dots correspond to the production wells and the blue dot to the injection well.

magnitude as the physical diffusion and dispersion, we neglect the capillary forces and dispersion. We assume that the effects of gravity can be neglected, that the permeability is isotropic, and that porosity, density, viscosity and permeability are pressure-independent.

- *Fluids properties:*
  The relative permeabilities which represent the additional resistance to flow of a phase caused by the presence of the other phase is described with a Corey model. Corey exponents for oil and water are equal to 2, the end point relative permeabilities for oil and water equal 0.9 and 0.6, respectively, and the residual oil saturation and connate water saturation are both equal to 0.2. The oil and water viscosities equal $0.5e-3$ $[Pas]$ and $1.0e^{-3}$ $[Pas]$, respectively. The water, oil and rock compressibilities equal $1.0e^{-9}$ $[1/Pa]$.

- *Reservoir geometry:*
  The total size of the reservoir is $700m \times 700m \times 2m$ and it is divided into $21 \times 21 \times 1$ uniform Cartesian grid blocks.

- *Reservoir properties:*
  The porosity is assumed to be uniform and equals 0.3. The 'true' permeability field on a natural log scale in $[m^2]$ is shown in Figure 2.

- *Initial conditions:*
  The initial reservoir pressures is $30 \cdot 10^6 Pa$ and the initial water saturation is taken as connate water saturation.

- *Well locations and constraints:*
  Water is injected at a rate of 0.003 m^3/s in a vertical injector placed in the middle of the reservoir, while four vertical producers at the corners are operated at constant pressures of $25 \cdot 10^6 Pa$. The well configuration consists of four producers in the corners and one central water injector, and is depicted in Figure 2.

5.1 Data assimilation settings

All assimilation processes are performed with the same settings. The assimilation period is 250 days and observations are taken from the four production wells and the injector well after 60, 120, 180 and 240 days of reservoir life, resulting in 20 observations. During this period no water is observed in the production wells. The observations are generated from the model with the 'true' permeability field. In the injection well the bottom hole pressure is measured and is assumed to have 10% error. In the production wells the total production rates are measured and they are assumed to have 5% error. The errors are generated from normal distributions with zero mean and a standard deviation equal to 10% and 5% of actual data in the injection and production wells, respectively. We assume that distinct observations are affected by physically independent noise, resulting in a diagonal covariance matrix for the observation errors. The 'true' permeability field is chosen from an ensemble of 1000 permeability fields and the remaining ensemble members are used to create the prior permeability $\boldsymbol{\theta}^b$ and to estimate the prior covariance matrix.

5.2 Model-reduced settings

Additional settings are required for the model-reduced approach related to the number of snapshots and patterns. We need to be aware that, to cover the dynamic behavior of the system, snapshots should be taken during all the assimilation times at each model step. Therefore we collect snapshots at each step of the model. The number of eigenvectors chosen for the projection matrix corresponds to 99% of the total eigenvalues. We estimate 20 parameters (after parameter reduction, see,[23] for further details).

5.3 Experiment 1

In this work our aim is to show that the model-reduced approaches provide results comparable to those of the adjoint-based approach. We performed parameter estimation using the adjoint-based approach, the POD-based approach and the balanced POD-based approach. In all cases we have iterated until the objective function reached comparably low values.

The results of the first data assimilation experiment for the adjoint-based approach, the POD-based approach and the balanced POD-based approach are summarized in Table 1, Table 2 and Table 3, respectively . The adjoint-based approach is the most efficient one, as was to be expected. The balanced POD-based approach is only twice less efficient than the adjoint-based approach. If we assume that the finite difference approach converges as the presented adjoint-based approach then it would require approximately ten times as many simulations as the adjoint approach. This means that for this example the computational effort of the POD-based approach is almost twice as efficient as the finite difference approach. The balanced POD-based approach is about four times as efficient as the finite difference approach.

A qualitative measure to compare the results can be obtained by visual inspection of

Outer loop	$J(\Delta\boldsymbol{\theta})$	RMSE	Nr of parameters	Nr of sim	Total nr of sim
0	784	1.15	20	1	1
1	521	1.12	20	1×2	3
2	42	1.04	20	1×2	5
10	15	0.80	20	1×2	21
12	12	0.77	20	1×2	25
15	11.2	0.76	20	1×2	31
29	9.65	0.65	20	1×2	60

Table 1: Experiment 1. Adjoint-based history matching

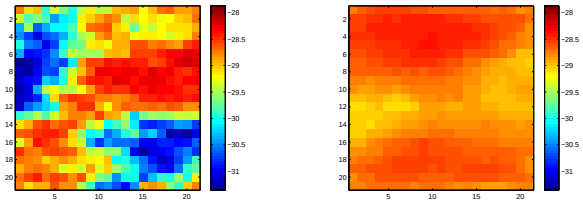
Outer loop	$J(\Delta\boldsymbol{\theta})$	RMSE	Nr of parameters	Nr of sim	Total nr of sim
0	784	1.15	20	1	1
1	15.88	0.75	20	$70 + 20 + 20 + 1$	112
2	10.92	0.73	20	$70 + 20 + 20 + 1 + 1$	224
3	9.68	0.57	20	$59 + 20 + 20 + 1 + 1$	325

Table 2: Experiment 1. POD-based history matching

Outer loop	$J(\Delta\boldsymbol{\theta})$	RMSE	Nr of parameters	Nr of sim	Total nr of sim
0	784	1.15	20	1	1
1	70.31	0.86	20	$20 + 20 + 20 + 16 + 1$	78
1	15.89	0.57	20	5	83
2	9.14	0.54	20	$20 + 20 + 20 + 18 + 1$	162

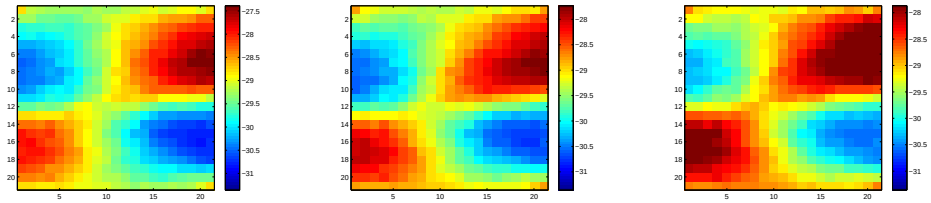
Table 3: Experiment 1. Balanced-POD-based history matching

the estimated permeability fields, see Figure 3 and Figure 4. Those figures depict the



(a) True permeability field. (b) Prior permeability field.

Figure 3: True and prior permeability fields.



(a) The POD-based estimate of the permeability field. (b) The balanced POD-based estimate of the permeability field. (c) The adjoint-based estimate of the permeability field.

Figure 4: Estimated permeability fields.

'true' permeability field, the prior permeability field, the estimates obtained using the POD-based approach, the balanced POD-based approach and the adjoint approach, respectively. In all the final estimates the high and low permeability channels were detected. A first quantitative measure to compare the results in the twin experiment is the *Root Mean Squared Error* ($RMSE$) defined as the square root of the mean squared difference between the 'true' parameter vector and its estimate. RMSEs are presented in Table 1, Table 2 and Table 3 and they show that all three estimates are very close to the true permeability field. Since the objective function derived for this problem contains a prior term, the true permeability field is not an argument that minimizes the given objective function, and therefore, RMSE can be only an indication of the quality of the obtained estimate. As a second quantitative measure to compare the results we used the *water breakthrough time prediction capability*. The water breakthrough time is evaluated over a period of 3 years and is depicted in Figure 5. The curves represent produced water rates for the 'true' permeability field (red), the prior permeability field (blue), the permeability estimate obtained from the adjoint-based method (pink), the estimate obtained from the balanced POD-based approach (black) and the the estimate obtained from the POD-based approach (green). Figure 5 shows that the prediction of water breakthrough

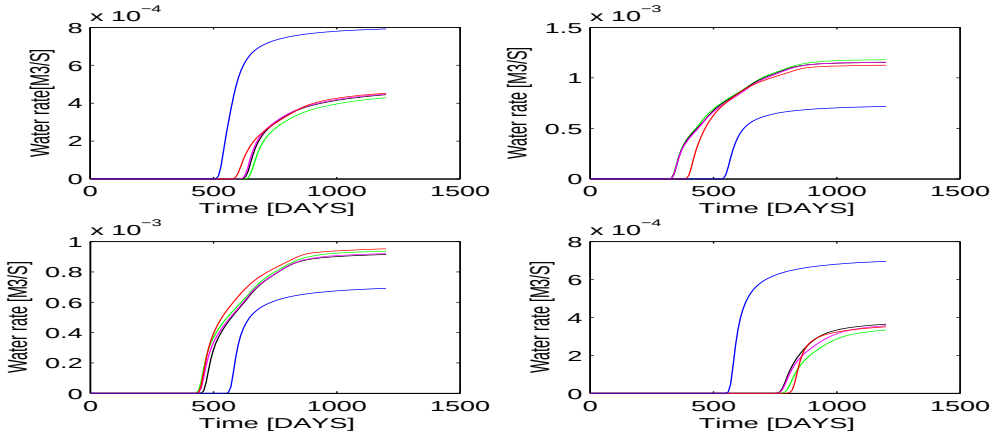


Figure 5: The prediction of water breakthrough time and produced water flow rates using the 'true' permeability field (red), the prior estimate (blue) and estimates obtained from three data assimilation approaches: the adjoint approach (pink), the POD-based approach (green) and the balanced POD-based approach (black).

time is very good for all the estimates.
Summarizing, we obtained very good results for all three methods. From a computational point of view the most efficient one is the adjoint-based method. However, the balanced POD method is still about half as efficiency as the adjoint-based method.

5.4 Experiment 2

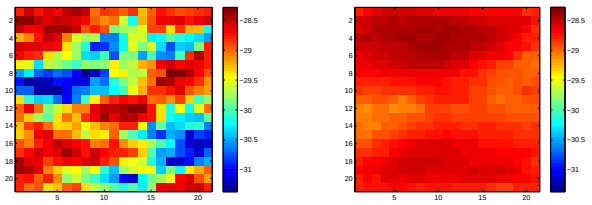
In the first experiment we concluded that the balanced POD-based approach and adjoint-based approach are the most efficient. Here, we performed an experiment for a more complicated 'true' permeability field which has two low-permeability channels on the diagonal, see Figure 2. In order to estimate this field from the data a longer assimilation period is required, namely 500 days. The rest of the setting stayed the same as in the first experiment.
The results of the second data assimilation experiment for the adjoint-based approach and the balanced POD-based approach are summarized in Table 4 and 5, respectively.
The obtained parameters are as follows: Similar to the first experiment we compared two quantitative measures, that is, the RMSEs and water breakthrough time prediction capability. RMSEs are presented in Table 4 and Table 5 and they show that both estimates are close to the true permeability field. The low permeable channels are clearly identified. The water breakthrough time is evaluated over a period of 1200 days and is depicted in Figure 8. The curves represent produced water rates for the 'true' permeability field (red), the prior permeability field (blue), the permeability estimate obtained from the adjoint-based method (pink), the estimate obtained from the balanced POD-based approach (black). The prediction of water breakthrough time is much better for both the estimates than for the prior permeability field.

Outer loop	$J(\Delta\boldsymbol{\theta})$	RMSE	Nr of parameters	Nr of sim	Total nr of sim
0	154	1.12	20	1	1
1	95.90	1.05	20	1×2	3
2	54.01	1.09	20	1×2	5
5	50.26	1.06	20	1×2	11
14	45.45	0.84	20	1×2	29
24	40.34	0.65	20	1×2	49
25	39.89	0.65	20	1×2	51
26	39.56	0.645	20	1×2	53
35	38.45	0.65	20	1×2	71

Table 4: Experiment 2. Adjoint-based history matching

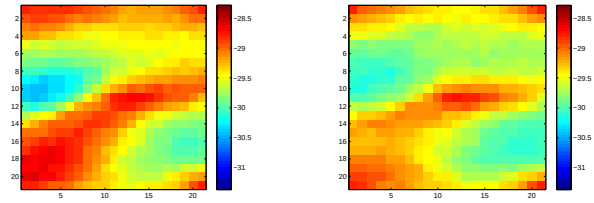
Outer loop	$J(\Delta\boldsymbol{\theta})$	RMSE	Nr of parameters	Nr of sim	Total nr of sim
0	154	1.12	20	1	1
1	44.35	0.64	20	$19 + 20 + 20 + 20 + 1$	81
1	31.57	0.58	20	1	82

Table 5: Experiment 2. Balanced-POD-based history matching



(a) True permeability field. (b) Prior permeability field.

Figure 6: True and prior permeability fields.



(a) The balanced POD-based estimate of the permeability field. (b) The adjoint-based estimate of the permeability field.

Figure 7: Estimated permeability fields.

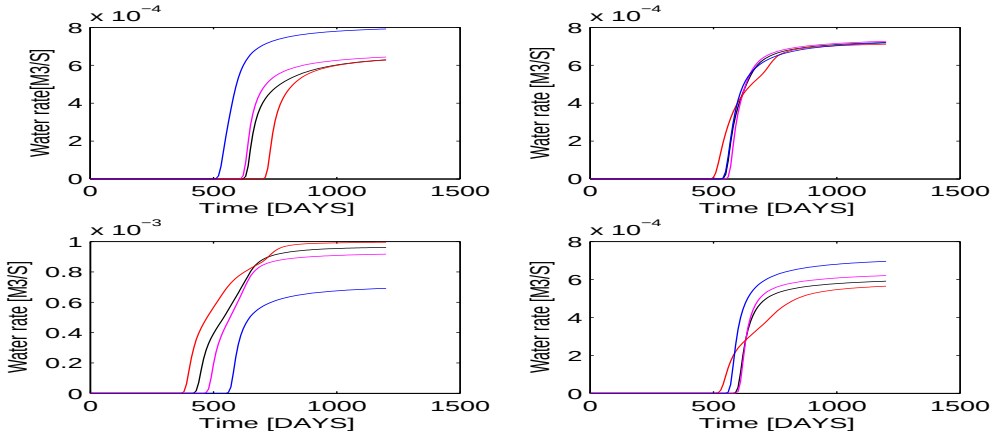


Figure 8: The prediction of water breakthrough time and produced water flow rates using the 'true' permeability field (red), the prior estimate (blue) and estimates obtained from two data assimilation approaches: adjoint approach (pink) and balanced POD-based approach (black).

We did not reach with the adjoint-based method an objective function value as low as with the balanced POD-based approach. We converged to different local optima. Contrary to the first experiment, the computational efficiency of the balanced POD-based approach is almost identical to the adjoint-based approach.

6 CONCLUSIONS

This paper presented initial results of the balanced POD-based gradient-based history matching algorithm. We have compared the results to the classical adjoint-based history matching algorithm and to the POD-based gradient-based history matching algorithm. We have obtained very good results for all three methods in the form of very good matches to the past data, low RMSEs, and good predictions of water production after breakthrough. The main difference between those methods is in the computational time. For the small example considered (882 states) our results from the first experiment indicate that the balanced POD-based approach gives computational improvements compared to the POD-based method. Assuming that the finite-difference approach converges as the presented adjoint approach both model-reduced approaches are more efficient than the finite difference approach. The POD-based and the balanced POD-based approach showed to be about twice and four times as efficient as finite difference approach, respectively. The most efficient is the adjoint-based method, as expected. However, in the second experiment the balanced POD-based approach has an efficiency very close to that of the adjoint-based method. In the situation that the adjoint model is available but not the sensitivity matrices of the system with respect to parameters of interest, the balanced POD-based method appeared to be a comparable alternative method to the adjoint-based approach.

ACKNOWLEDGMENT

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

REFERENCES

[1] D.S. Oliver, A.C. Reynolds and N. Liu, Inverse theory for petroleum reservoir characterization and history matching. Cambridge University Press, Cambridge (2008)

[2] M. P. Kaleta, R. G. Hanea, J. D. Jansen, A. W. Heemink, Model-reduced variational data assimilation for reservoir model updating. Proc. 11th European Conference on the Mathematics of Oil Recovery, Bergen, Norway, 8-11 September 2008

[3] M. P. Kaleta, R. G. Hanea, J. D. Jansen, A. W. Heemink, Model-reduced gradient-based history matching. Submitted to Computational Geosciences

[4] K. Karhunen, Zur spektral theorie stochasticher prozesse. Annales Academiae Scientiarum Fennicae, Ser. A1, 34, 1-7 (1946)

[5] M. Loéve, Functions aleatoire de second ordre. Revue Science, **84** 195-206 (1946)

[6] J. L. Lumley, The structure of inhomogeneous turbulence flows. Atmospheric Turbulence and Radio Wave Propagation, In ed. by A. M. Yaglom and V. I. Tatarsky (Nauka, Moskau), 166-178 (1967)

[7] C. T. Mullis and R. A. Roberts, Synthesis of minimum roundoff noise fixed point digital filters. Trans. Circuits Syst., **23** 551-562 (1976)

[8] B. C. Moore, Principal component analysis in linear systems: Controllability, observability and model reduction. IEEE Trans. Automat. Control, **26** 17-32 (1981)

[9] S. Lall, I. E. Marsden and S. Glavaski, A subspace approach to balanced truncation for model reduction of nonlinear control systems. Internal. J. Robust Nonlinear Control, **12** 519-535 (2002)

[10] J. M. A. Scherpen, Balancing for nonlinear systems. Syst. Contr. Lett., **21** 143-153 (1993)

[11] C. W. Rowley, Model reduction for fluids, using balanced proper orthogonal decomposition. International Journal of bifurcation and chaos, **15** 997-1013 (2005)

[12] M. J. Zandvliet, J. F. M. van Doren, O. H. Bosgra, J. D. Jansen and P.M.J. van den Hof, Controllability, observability and identifiability in single-phase porous media flow. Computational Geosciences, **12** 605-622 (2008)

[13] K. Aziz and A. Settari, Petroleum reservoir simulation. Applied Science Publishers, London, UK (1979)

[14] G. R. Gavalas, P. C. Shah and J. H. Seinfeld, Reservoir history matching by Bayesian estimation. SPE Journal, **16** 337-350 (1976)

[15] P. T. M. Vermeulen, A. W. Heemink, Model-reduced variational data assimilation. Monthly Weather Review, **134** 2888-2899 (2006)

[16] P. Courtier, J-N. Thépaut and A. Hollingsworth, A strategy for operational implementation of 4D-Var, using an incremental approach. Quarterly Journal of the Royal Meteorological Society, **120** 13671387 (1994)

[17] A.C. Antoulas, Approximation of Large-Scale Dynamical Systems. SIAM, Philadelphia (2005).

[18] S. Lall, I. E. Marsden and S. Glavaski, Empirical model reduction of controlled nonlinear systems. IFAC World Congress, **F** 473-478 (1999)

[19] K. Willcox and J. Peraire, Balanced model reduction via the proper orthogonal decomposition, AIAA J. 40, 2323-2330 (2002)