Ensemble-based Reservoir History Matching using Hyper-reduced-order Models

by

Seonkyoo Yoon

B.S. Civil and Environmental Engineering, Yonsei University, 2007
M.S. Civil and Environmental Engineering, Yonsei University, 2009

SUBMITTED TO THE DEPARTMENT OF
CIVIL AND ENVIRONMENTAL ENGINEERING
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2016

Signature redacted

Author ....
Department of Civil and Environmental Engineering
August 5, 2016
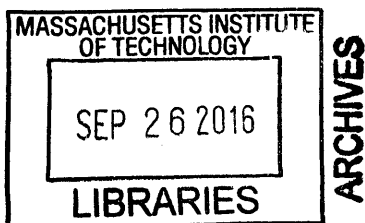
Signature redacted

Certified by ....
John R. Williams
Professor of Civil and Environmental Engineering
Thesis Supervisor

Signature redacted

Accepted by ....
Jesse Kroll
Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

1

(This page is intentionally left blank)

Ensemble-based Reservoir History Matching using Hyper-reduced-order Models

by

Seonkyoo Yoon

Abstract

Subsurface flow modeling is an indispensable task for reservoir management, but the associated computational cost is burdensome owing to model complexity and the fact that many simulation runs are required for its applications such as production optimization, uncertainty quantification, and history matching. To relieve the computational burden in reservoir flow modeling, a reduced-order modeling procedure based on hyper-reduction is presented. The procedure consists of three components: state reduction, constraint reduction, and nonlinearity treatment. State reduction based on proper orthogonal decomposition (POD) is considered, and the impact of state reduction, with different strategies for collecting snapshots, on accuracy and predictability is investigated. Petrov–Galerkin projection is used for constraint reduction, and a hyper-reduction that couples the Petrov–Galerkin projection and a 'gappy' reconstruction is applied for the nonlinearity treatment. The hyper-reduction method is a Gauss–Newton framework with approximated tensors (GNAT), and the main contribution of this study is the presentation of a procedure for applying the method to subsurface flow simulation. A fully implicit oil–water two-phase subsurface flow model in three-dimensional space is considered, and the application of the proposed hyper-reduced-order modeling procedure achieves a runtime speedup of more than 300 relative to the full-order method, which cannot be achieved when only constraint reduction is adopted.

In addition, two types of sequential Bayesian filtering for history matching are considered to investigate the performance of the developed hyper-reduced-order model to relive the associated computational cost. First, an ensemble Kalman filter (EnKF) is considered for Gaussian system and a procedure embedding the hyper-reduced model (HRM) into the EnKF is presented. The use of the HRM for the EnKF significantly reduces the computational cost without much loss of accuracy, but the combination

requires a few remedies such as clustering to find an optimum reduced-order model according to spatial similarity of geological condition, which causes an additional computation. For non-Gaussian system, an advanced particle filter, known as regularized particle filter (RPF), is considered because it does not take any distributional assumptions. Particle filtering has rarely been applied for reservoir history matching due to the fact that it is hard to locate the initial particles on highly probable regions of state spaces especially when large scale system is considered, which makes the required number of particles scale exponentially with the model dimension. To resolve the issues, reparameterization is adopted to reduce the order of the geological parameters. For the reparameterization, principal component analysis (PCA) is used to compute the reduced space of the model parameters, and by constraining the filtering analysis with the computed subspace the required number of initial particles can be reduced down to a manageable level. Consequently, a huge computational saving is achieved by embedding the HRM into the RPF. Furthermore, the additional cost of clustering required to identify the geospatially optimum reduced-order model is saved because the advanced particle filter allows to easily identify the groups of geospatially similar particles.

Thesis Supervisor: John R. Williams

Title: Professor of Civil and Environmental Engineering

(This page is intentionally left blank)

Acknowledgments

I am grateful to many people for being supportive through my life at MIT. First, I appreciate my advisor professor John R. Williams for his guidance, insight, and support through the past years. Without him I couldn't finish my doctoral research. I also appreciate my committee professors. I thank professor Ruben Juanes for serving as the chair in the committee. I learned from him computational modeling for subsurface flow and it was so enjoyable to refine my skill for viewing the world in the way of discretization. I am also deeply grateful to professor Youssef M. Marzouk for his guidance. It was lucky to learn the method of data assimilation from his class. Throughout my thesis research I have received a number of valuable and insightful comments from him regarding Bayesian filtering. I believe I am so privileged to be under the guidance from the three eminent scholars for my doctoral research. I am indebted to them for their valuable mentoring.

I am also very grateful to the former and current administrative staffs at the department of Civil and Environmental Engineering, especially Patricia Glidden, Kris Kipp, Rebecca R. Fowler, and Kiley Clapper for being kind, helpful, and supportive.

I would also like to thank my colleagues and friends at MIT. Especially I deeply appreciate Dr. Zeid M. Alghareeb for his guidance to the research of reservoir simulation and reduced-order modeling. He has been encouraging and supportive although my early-stage work was a bit chaotic at first. I also thank the members of MIT Geonumerics, Dr. Bruce Jones, Kai Pan, Abdulaziz Albaiz, and Mohamad Sindi for the meaningful research discussions and invaluable friendship. Special thanks to my officemates Linsen Chong,

Contents

Chapter 3 Ensemble-based History Matching using Reduced-order Models

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Literature Review

Oil and gas provide almost 60 percent of the current global energy demand. The dominance of hydrocarbon energy is expected to continue in the future, at least to 2040 (see Figure 1.1), while the global energy demand is expected to increase by about 35 percent from 2010 to 2040 (Exxon Mobil, 2014). In order to meet the rising demand, it will become increasingly important to maximize oil recovery from existing reservoirs, since most oilfields are already at a mature stage of their life cycle, and the discovery of large new oilfields is becoming rare.

In the last decade, the need to increase oil recovery from reservoirs has resulted in the development of a variety of technologies to better measure and control the production process through wells. A well equipped with such technology is called a "smart" (or

intelligent) well. Each smart well is equipped with multiple down-hole variable control valves, each for a different section of the well. Using these valves, the amount of water injected into and oil produced from a specific geologic layer can also be controlled. Compared to traditional wells, smart wells drastically increase monitoring capability and controllability. The enhanced monitoring capability and controllability of smart wells open a possibility of real-time reservoir management in an iterative way, which is known as closed-loop reservoir management (CLRM). The advantage of the CLRM is that it can deal with the uncertainty of model parameters. In the framework of CLRM the uncertain model parameters can be adjusted whenever new observations are available through history matching, and the well controls are optimized based on the newly updated model. This process can be carried out iteratively in a near-continuous way over the life of a reservoir, so that it is possible to significantly increase the life-cycle value of an oil reservoir (Jansen et al., 2005), as depicted in Figure 1.2.

Figure 1.1: World energy demand projected to 2040, from ExxonMobil (2014)

Figure 1.2: A closed-loop reservoir management process (from Capolei et al., 2013)

The problem of history matching can be treated as an optimization problem, in which objective function is formulated and iterative schemes of minimization/maximization are applied. To find the direction toward optimum controls in the iterative schemes, gradients of the objective function with respect to controls at every time step are required. The computation of the gradients can be based on the use of an adjoint formulation (Brouwer et al., 2004; Sarma et al., 2006; Sarma et al., 2008; Jansen, 2011). Although the adjoint method is a computationally efficient way of evaluating the gradient information, it has the defects that it requires access to the simulation code and enormous programming efforts.

Alternative methods that avoid the use of the adjoint formula are ensemble-based methods. For example, ensemble Kalman filter (EnKF) methods have gained enormous

popularity for history matching over the last decade in the field of reservoir engineering (Nævdal et al., 2002; Nævdal et al., 2005; Gu and Oliver, 2006, 2007; Aanonsen et al., 2009; Chen et al., 2009; Chen and Oliver, 2010). An advantage of the ensemble-based method is that they are adaptable to any reservoir simulator, because the simulator serves only as a black-box. In addition, a recent benchmark study, in which different history matching methods are tested on the synthetic "Brugge filed" by several research groups, showed that the three best results are all obtained by methods using an ensemble Kalman filter for history matching (Peters et al., 2010).

However, the ensemble-based approach to history matching is challenging to apply in practice due to high computational cost. A typical reservoir simulator may take up to several hours to complete a single run due to model complexity, and a large number of simulation runs are required for the ensemble-based methods. A promising approach for relieving the computational burden is to replace the highly resolved model with computationally efficient, reduced-order models that are able to reproduce the features of the full-order simulation results at a fraction of their computational burden.

A common way of constructing the reduced-order model is based on projection, which replaces the full-order states with a small set of reduced variables and reduces the computational complexity by solving a projection of the full-order model onto a reduced subspace. The use of the projection-based reduced-order reservoir models is supported by an important observation that although the state of reservoir systems is described in a large dimensional vector space, it is actually possible to reproduce the high dimensional description with a small set of reduced variables that lies in a lower-dimensional manifold. One of the most popular methods for constructing the lower-dimensional

manifold is proper orthogonal decomposition (POD), which has been actively studied in recent years in the field of petroleum reservoir engineering (Vermeulen et al., 2004; Heijn et al., 2004; van Doren et al., 2006; Markovinović and Jansen, 2006; Cardoso et al., 2009). The popularity is due to an appealing feature of POD: minimization of the reconstruction error, defined as the square distance between the original snapshot data and its projection approximations. In addition, the lower-dimensional manifold produced by POD can be used as a subspace on which the system equations are projected. This is referred to as Galerkin projection; on which, most projection-based reduced-order reservoir modeling methods are founded (van Doren et al., 2006; Cardoso et al., 2009).

A recent study of reduced-order subsurface reservoir modeling shows the advantage of Petrov–Galerkin projection compared to Galerkin projection in terms of stability and accuracy (He and Durlofsky, 2015). Petrov–Galerkin projection uses varying lower-dimensional subspaces updated according to the state of the system, which changes with each time point of a simulation, while the Galerkin projection uses a fixed lower-dimensional subspace. Although the method provides an improved solution, it cannot achieve significant computational savings when the system of interest is highly nonlinear. This occurs because, to compute the reduced nonlinear system to solve, one must first reconstruct the full-order state solution and evaluate the full-order nonlinear system before projecting it again onto a reduced subspace. Therefore, the system is still dependent on the dimension of the full-order model even though it solves a projection of the full-order system onto a lower-dimensional subspace. This is a common disadvantage of projection-based reduced-order modeling for nonlinear system.

For efficient approximation of nonlinear terms, the combination of POD and trajectory piecewise linearization (TPWL) has recently been suggested for typical nonlinear reservoir simulation models, so that independence from the full-order dimension is achieved (Cardoso and Durlofsky, 2010; He et al., 2011, 2013, He and Durlofsky, 2015). Its application has also been successfully expanded to more general cases such as compositional systems (He and Durlofsky, 2014) and thermal recovery processes (Rousset et al., 2014). The TPWL method, which was originally developed by Rewienski and White (2003), approximates the nonlinear term by a piecewise-linear function obtained by linearizing the system at selected points along its trajectory. However, for highly nonlinear problems, it is difficult to approximate the nonlinear term accurately with a piecewise-linear representation without involving a large number of linearized models along the trajectory. Furthermore, the fact that the method cannot search the nonlinear model away from the selected linearizing points may induce a lack of robustness for highly nonlinear problems.

Another approach for nonlinear model order reduction is based on hyper-reduction. For example, Chaturantabut and Sorensen (2010) proposed a discrete empirical interpolation method (DEIM) to deal with the nonlinearity. DEIM approximates the nonlinear terms of a model by a linear combination of its basis vectors, for which the expansion coefficients are determined using a small set of interpolation points. The set of basis vectors for the nonlinear terms can be generated using POD as well. Their work showed that nonlinear terms can be approximated by interpolating the evaluated values over a subset of spatial grid points, so that it avoids full evaluation of the nonlinear term, which improves efficiency. The application of DEIM to reservoir modeling has been

introduced by Suwartadi (2012), Glidin et al. (2013), Alghareeb and Williams (2013), and Ghommem et al. (2015). A localized variant of DEIM called LDEIM was proposed by Peherstorfer et al. (2014), and Yoon et al. (2014) presented a procedure of applying DEIM and LDEIM for reduced-order oil reservoir modeling. LDEIM computes several local subspaces, each tailored to a particular region of characteristic system behavior, and then selects an appropriate subspace on which the nonlinear term is projected. By using multiple local DEIM approximations, it is possible to reduce the number of interpolation points further, consequently reducing the computational costs even more than DEIM. However, the method of DEIM has not been considered in the framework of Petrov-Galerkin projection-based reduced-order model.

Coming back to the topic of ensemble-based reservoir history matching, EnKF, the most popular method, is known that it is not suitable for non-Gaussian system because it approximates state-space model Gaussian. As a remedy, a sequential Bayesian Monte Carlo method, known as Particle filter (PF), can be considered because it does not take any assumption of probability distribution. However, PF has rarely been applied for reservoir history matching due to the fact that it is hard to locate the initial particles on highly probable regions of state spaces especially when large scale system is considered, which makes the required number of particles scale exponentially with the model dimension (Aanonsen et al., 2009), well known as the so-called 'curse of dimensionality'. To resolve the issues, reparameterization can be adopted to reduce the order of the geological parameters. The order reduction via reparameterization for history matching has been applied in the framework of EnKF usually for atmospheric and oceanic applications (Cane et al., 1996; Pham, 2001; Hoteit and Pham, 2004; Rozier et al., 2007;

Solonen et al., 2016). For oil reservoir history matching, Zhang et al., (2007) applied Karhunen–Loève decomposition to represent the random filed, and Jafapour and McLaughlin (2008, 209) suggest the use of discrete Cosine transform (DCT) parameterization method, which is a Fourier related transform using orthogonal cosine basis functions.

1.2 Scope of Work

Based on the analysis of the literature in previous section, three areas of knowledge where new research is needed have been identified. First, the ensemble-based history matching needs to be improved in terms of efficiency, and it can be achieved by embedding a reduced-order model into the history matching framework. Second, for a typical nonlinear oil reservoir system, a new reduced-order modeling method, which is compatible with Petrov–Galerkin projection and does not require the full-order state variables for constructing the nonlinear terms, is needed. Third, for history matching of non-Gaussian reservoir system, the application of particle filtering need to be developed by solving the issue of the so-called 'curse of dimensionality'.

Therefore, this study first presents a hyper-reduction procedure that can be compatible with the use of Petrov–Galerkin projection (Yoon et al., 2016). The hyper-reduction method uses an approximated tensor in the Gauss–Newton method to estimate the reduced-order state. The method was originally developed by Carlberg et al. (2011), and they show that the computational savings are significant and cannot be achieved when using only Petrov–Galerkin projection. The main contribution of this study is the

presentation of a procedure for applying the method to subsurface flow simulation and a framework of embedding the reduced-order model into ensemble-based history matching. A fully implicit two-phase flow reservoir model consisting of oil and water in three-dimensional space is considered. In addition, I investigate a new way of defining a lower dimensional subspace of the system state, proposed by Carlberg et al. (2011, 2013), that constructs a snapshot matrix according to consistency. The implication of this reduced-order reservoir modeling procedure has wide applicability in multi-phase reservoir flow modeling even though it is presented for the simple two-phase flow case.

Then, this study presents the application of the hyper-reduced-order model for ensemble-based history matching to relieve the computational cost associated with forward simulations. Two types of filtering methods are considered: ensemble Kalman filtering for Gaussian system, and particle filtering for non-Gaussian system. To the best of my knowledge, there is no application of hyper-reduced-order reservoir models and particle filtering for reservoir history matching.

1.3 Dissertation Outline

In Chapter 2, the modeling procedure for the fully implicit two-phase flow reservoir model is presented, followed by a subsection that presents the reduced-order modeling procedure based on hyper-reduction. Then, the derived reduced-order models are tested in comparison with standard reduced-order models using synthetic geologic models in light of their ability to accurately simulate the state of a reservoir system and well responses with various well controls. Finally, discussions and conclusions are presented.

In Chapter 3, two Ensemble-based history matching algorithms are presented. First, the method of ensemble Kalman filter for history matching is presented, followed by a subsection presenting the method of particle filter. The two methods are compared and discussed in view of reservoir history matching. Then, procedures of embedding the proposed hyper-reduced-model into the filtering algorithms are presented, followed by case studies.

Finally, I conclude with discussion on future research directions for the reduced-order modeling technique developed in this work. This final chapter also discuss on the extension of the application of the proposed hyper-reduced-order model for reservoir management decision making.

Chapter 2[†]

Reduced-order models for Subsurface

Flow Simulation

2.1 Oil–water Flow Equations and Discretized System

Subsurface flow models are derived by coupling mass conservation equations with the multiphase version of Darcy's law. For the oil–water case, we will consider immiscible oil–water flow, neglecting the effects of fluid and rock compressibility as well as capillary pressure. The governing equations with the approximations can be written as

$$\frac{\partial}{\partial t}(\phi \rho_\alpha s_\alpha) - \nabla \cdot [\rho_\alpha \lambda_\alpha \mathbf{k}(\nabla p_\alpha - \rho_\alpha \hat{g})] + q_\alpha = 0,$$ (2.1)

where the subscript α designates the fluid phase as oil with o or water with w. Here, t is time, ϕ is porosity, $\mathbf{k}$ is permeability, λ is phase mobility, p is phase pressure, ρ is phase density, $\hat{g} = -g\nabla z$ is the gravitational pull-down force, s is phase saturation, and q is the phase source/sink term. The mobility is a function of water saturation such that $\lambda_\alpha(s_w) = k_{r\alpha}(s_w)/\mu_\alpha$, where $k_{r\alpha}(s_w)$ is the relative permeability of phase α and μ_α is the viscosity of phase α. The system is closed by the additional constraints $s_o + s_w = 1$ and $p_c = p_o - p_w$, where p_c is the capillary pressure, which is assumed zero in this study, such that $p_o = p_w$. The primary unknowns in the above equation are taken to be the oil phase pressure p_o and the water phase saturation s_w. Once the primary unknowns are estimated, the alternate phase quantities, i.e., p_w and s_o, are determined straightforwardly using the additional constraints. For clarity of notation, p_o and s_w will henceforth be denoted by p and s, omitting the subscripts relying on the closure constraints.

The fully implicit, discretized representations of the governing equations can be written as

$$\mathbf{r}^n = \mathbf{r}(\mathbf{x}^n, \mathbf{x}^{n-1}, \mathbf{u}^n) = 0.$$ (2.2)

Here, $\mathbf{r}$ is the residual vector, n denotes time step, and $\mathbf{u}$ is the set of well control parameters, which are taken to be the pressures of injection and production wells in this

study. The well pressures are often referred as bottom-hole pressures (BHPs). Finally, $\mathbf{x}$ designates the state vector that contains both primary variables in each grid block. The state vector $\mathbf{x}$ with n_c grid blocks can be written as

$$\mathbf{x} = \begin{bmatrix} \mathbf{p} \\ \mathbf{s} \end{bmatrix},$$ (2.3)

implying that the dimension of the state vector is $2n_c$. In Equation (2.2), $\mathbf{x}^{n-1}$ is the state vector solved at the previous time step $n-1$, and the goal is to compute the $\mathbf{x}^n$ that drives the residual to zero. For simplicity, the residual equation that we seek to solve for $\mathbf{x}$ at a time step can be rewritten with the unknown input vector as

$$\mathbf{r}(\mathbf{x}) = 0.$$ (2.4)

Owing to the non-linearity involved with Equation (2.4), it is typically solved using a non-linear solver such as Newton's method. For each iteration, this entails solving

$$\mathbf{J}\boldsymbol{\delta} = -\mathbf{r},$$ (2.5)

where $\mathbf{J}$ is the Jacobian matrix given by $\mathbf{J} = \partial \mathbf{r} / \partial \mathbf{x}$ and $\boldsymbol{\delta} = \mathbf{x}^{n,k} - \mathbf{x}^{n,k-1}$ with iteration level k. Solving Equation (2.5) is computationally expensive because the number of grid blocks for practical reservoir models is on the order of $10^4 - 10^6$.

More details of oil–water reservoir flow modeling can be found in the works by Aziz and Settari (1979) and Gerritensen and Durlofsky (2005). The model has also been considered in the context of reduced-order modeling based on POD-TPWL by Cardoso et al. (2009), Cardoso et al. (2010), and He et al. (2011).

2.2 Reduced-order Modeling for Subsurface Flow

In this section, the procedure of the proposed reduced-order modeling is presented. First, the projection-based reduced-order model using POD is described, followed by a section presenting hyper-reduction based on Gauss–Newton with approximate tensors (GNAT).

2.2.1 Projection-based Order Reduction

To reduce the dimensionality of Equation (2.4), a projection process based on POD is employed to compute a lower-dimensional subspace for an approximate solution $\tilde{\mathbf{x}}$. As discussed in van Doren et al. (2006) and Cardoso et al., (2009), we separately compute two lower-dimensional subspaces, $\mathcal{P}$ for the oil pressure and $\mathcal{S}$ for the water saturation. Consequently, we have the approximate solution $\tilde{\mathbf{x}}$ to Equation (2.4) written as

$$
\begin{aligned}
\mathbf{x} \approx \tilde{\mathbf{x}} = \mathbf{x}^0 + \mathbf{\Phi}_x \mathbf{z} \\
= \begin{bmatrix} \mathbf{p}^0 \\ \mathbf{s}^0 \end{bmatrix} + \begin{bmatrix} \mathbf{\Phi}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_s \end{bmatrix} \begin{bmatrix} \mathbf{z}_p \\ \mathbf{z}_s \end{bmatrix},
\end{aligned}
\tag{2.6}
$$

where $\mathbf{x}^0$ denotes the initial state condition, $\mathbf{\Phi}_p \in \mathbb{R}^{n_c \times l_p}$ and $\mathbf{\Phi}_s \in \mathbb{R}^{n_c \times l_s}$ are l_p-dimensional and an l_s-dimensional basis matrices for $\mathcal{P}$ and $\mathcal{S}$, respectively, $\mathbf{z}_p \in \mathbb{R}^{l_p}$

and $z_s \in \mathbb{R}^{l_s}$ denote the coordinates of the state vector in the corresponding basis, and

$\mathbf{\Phi}_x \in \mathbb{R}^{n_c \times l}$, where $l = l_p + l_s$, is the basis matrix for the direct sum of the subspaces

$\mathcal{D} = \mathcal{P} \oplus \mathcal{S}$ that designates the entire subspace of the state vector $\mathbf{z}$. The designation of

l_p and l_s can be done by 'energy criteria' (Cardoso et al., 2009; Cardoso et al., 2010; He

et al., 2011). The computational benefits of the reduced-order representation of the state

vector comes from the fact that $l \ll 2n_c$. To construct the basis matrices $\mathbf{\Phi}_p$ and $\mathbf{\Phi}_s$, a

simulation of the full-order model is performed, and the solved states are saved. The

states consist of solutions for p and s in all grid blocks. Once a snapshot matrix is

assembled from manipulating the saved states, a projection matrix is constructed by POD.

As an example, the procedure of constructing the POD basis $\mathbf{\Phi}_p \in \mathbb{R}^{n_c \times l_p}$ for the

subspace $\mathcal{P}$ using singular value decomposition (SVD) is described in Algorithm 1. SVD

is the factorization of a matrix $\mathbf{S}_p$ into the form $\mathbf{S}_p = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U}$ is orthogonal, $\mathbf{\Sigma}$

is diagonal, and $\mathbf{V}$ is orthogonal. The fundamental basis for the use of SVD is its

inherent feature that the reconstruction error $\sum_{i=1}^{n_t} \left\| \mathbf{s}_p^i - \overline{\mathbf{s}} - \mathbf{\Phi}_p \mathbf{z}_p^i \right\| / n_t$ is minimized when

the projection matrix $\mathbf{\Phi}_p$ is computed by SVD.

Algorithm 1. Basis construction via POD (from Carlberg et al., 2011).

Input: Snapshot matrix $\mathbf{S}_p = [\mathbf{s}_p^1, ..., \mathbf{s}_p^{n_t}] \in \mathbb{R}^{n_c \times n_t}$ (n_t denotes the number of snapshots)
Output: $\mathbf{\Phi}_p \in \mathbb{R}^{n_c \times l_p}$
    1. Compute the thin SVD: $\mathbf{S}_p = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$
    2. Choose the basis dimension $n \in [1, 2, ..., n_t]$

3. Truncate the left singular matrix to get the basis $\boldsymbol{\Phi}_p = [u^1,...,u^{l_p}]$, where

   $$\mathbf{U} \equiv [u^1,...,u^{n_t}]$$

In most studies of reduced-order reservoir modeling, the collection of the states is used as a snapshot matrix, and the basis matrices are constructed by applying POD to the snapshot matrices. However, during the inline stage where attempts are made to solve Equation (2.5), the increment in the state $\mathbf{x}^n - \mathbf{x}^{n(0)}$—not the state itself—is sought in the subspace $\mathcal{D}$, where $\mathbf{x}^{n(0)} = \mathbf{x}^{n-1}$ denotes the initial guess for the Newton solver, which is usually the state vector solved at the previous time step $n-1$. This implies that the snapshot matrices are consistent when they comprise the state increments in the form of $\mathbf{x}^n - \mathbf{x}^{n-1}$, not the state $\mathbf{x}^n$ itself. The consistency is defined in the context that an approximation is consistent if, when the truncation of POD basis vectors is not performed, it does not introduce any additional error in the solution of the same simulation from which the snapshot data were acquired. A rigorous proof of the consistent projection from using snapshots in the form of state increments can be found in Appendix A of Carlberg et al. (2013). Their way of defining consistency implies that consistency is exactly met when the reduced-order simulation is conducted using the same model control as the training simulation from which the snapshots are acquired. Therefore, an investigation on the suitability of the new strategy is necessary because most applications of reduced-order modeling for reservoir simulations require predictability under control changes, and this study performs a case study comparing the common strategy of collecting state vectors and the new strategy of collecting snapshots in the form of state increments (see Section 4.1).

Substituting Equation (2.6) into Equation (2.4) gives $\mathbf{r}(\tilde{\mathbf{x}}) = 0$, which is an overdetermined system of $2n_c$ equations with l unknowns. Consequently, the reduced-order system seeks to compute $\tilde{\mathbf{x}}$ as the solution to a minimization problem:

$$\tilde{\mathbf{x}} = \underset{\tilde{\mathbf{X}} \in \mathbf{x}^0 + \mathcal{D}}{\arg\min} \left\| \mathbf{r}(\tilde{\mathbf{X}}) \right\|_2 . \tag{2.7}$$

An equivalent function to be minimized for the above minimization problem is $0.5 \left\| \mathbf{r}(\tilde{\mathbf{x}}) \right\|_2^2$, and its gradient, when equated to zero, can be written as

$$\mathbf{\Phi}_x^T \frac{\partial \mathbf{r}}{\partial \mathbf{x}} (\mathbf{x}^0 + \mathbf{\Phi}_x \mathbf{z})^T \mathbf{r}(\mathbf{x}^0 + \mathbf{\Phi}_x \mathbf{z}) = 0 . \tag{2.8}$$

Then, the minimization problem can be solved by the Gauss–Newton method, which is equivalent to applying Newton's method to solve Equation (2.8) for $\mathbf{z}$, ignoring the second-order derivative terms in the Hessian matrix. For a single Gauss–Newton iteration, since the approximation to the Hessian matrix results in a normal equation, the problem can be viewed as a least-squares solution of

$$\boldsymbol{\delta}_z^k = \underset{\Delta \in \mathbb{R}^l}{\arg\min} \left\| \mathbf{J}^k \mathbf{\Phi}_x \Delta + \mathbf{r}^k \right\|_2 . \tag{2.9}$$

Here, k denotes the level of Gauss–Newton iteration. At iteration k, the reduced-order state vector is iterated forward by $\mathbf{z}^{k+1} = \mathbf{z}^k + \boldsymbol{\delta}_z^k$. The residual $\mathbf{r}^k$ and its Jacobian $\mathbf{J}^k$ are

defined as $\mathbf{r}^k \equiv \mathbf{r}(\mathbf{x}^0 + \mathbf{\Phi}_x \mathbf{z}^k)$ and $\mathbf{J}^k \equiv \dfrac{\partial \mathbf{r}}{\partial \mathbf{x}}(\mathbf{x}^0 + \mathbf{\Phi}_x \mathbf{z}^k)$, respectively. This is

mathematically analogous to performing a Petrov–Galerkin projection onto a subspace

whose basis vectors are defined by the columns of $\mathbf{J}^k \mathbf{\Phi}_x$, meaning that the basis is

variable and state-dependent. From the projection, l constraints are introduced, and the

order of the problem is reduced from $2n_c$ to l, resulting in the following iterations:

$$\mathbf{\Phi}_x^T (\mathbf{J}^k)^T \mathbf{J}^k \mathbf{\Phi}_x \boldsymbol{\delta}_z^k = -\mathbf{\Phi}_x^T (\mathbf{J}^k)^T \mathbf{r}^k , \tag{2.10}$$

$$\mathbf{z}^{k+1} = \mathbf{z}^k + \boldsymbol{\delta}_z^k . \tag{2.11}$$

Carlberg et al. (2011) demonstrated the outperformance of Petrov–Galerkin projection

over Galerkin projection, which is a more commonly used method. The Galerkin

projection method is equivalent to solving $\mathbf{\Phi}_x^T \mathbf{r}(\mathbf{x}^0 + \mathbf{\Phi}_x \mathbf{z}) = \mathbf{0}$, and previous research on

reduced-order reservoir modeling has been mostly based on this method. Recently, He et

al. (2014) and He and Durlofsky (2015) applied Petrov–Galerkin projection to reduced-

order reservoir modeling in the framework of TPWL and demonstrated its superior

accuracy and numerical stability over Galerkin projection.

2.2.2 Hyper-Reduction Method

Although the reduced-order Gauss–Newton problem of Equation (2.7) is of much

lower dimension than the original full-order system, the computational cost is still a

function of the full-order model dimension n_c. This dependence remains because the

projection-based procedure described in the preceding section still requires, at every

Newton iteration, to evaluate the full-order Jacobian and residual, to construct the

variable and state-dependent left subspace basis—through the matrix-matrix product

$\mathbf{J}^k\mathbf{\Phi}_\mathbf{x}$—and to project the full-order residual and Jacobian onto the reduced subspace.

Consequently, the computational savings obtained by this projection-based model

reduction method can be limited, which is an obstacle confronted by most projection-

based nonlinear model reduction methods.

One way to mitigate the computational cost involved with the nonlinearity in

Equation (2.9) is to approximate the nonlinear residual $\mathbf{r}^k$ and the left subspace basis

$\mathbf{J}^k\mathbf{\Phi}_\mathbf{x}$ with surrogates $\tilde{\mathbf{r}}^k$ and $\widetilde{\mathbf{J}^k\mathbf{\Phi}_\mathbf{x}}$, respectively, that are constructed using a 'gappy'

POD technique (Everson and Sirovich, 1995; Willcox, 2006). It is a hyper-reduction

method referred to as GNAT, which was introduced by Carlberg et al. (2011). We use the

method because of its consistency and optimality properties that were demonstrated by

Carlberg et al. (2011, 2013).

The numerical formulation for the oil–water model in Equation (2.2) is based on a

fully implicit solution technique, solving the oil and water conservation equations

simultaneously. It implies that the convergence behavior in solving the mass conservation

of the two phases may differ, so we apply the GNAT procedure separately for the oil and

water phases. The GNAT framework approximates the nonlinear residual and the left

subspace basis by computing only a small subset of their rows. Given the subset of $n_{i,\alpha}$

sample indices $\mathscr{I}_\alpha = \{\mathbf{i}_{1,\alpha}, \mathbf{i}_{2,\alpha}, \ldots, \mathbf{i}_{n_{i,\alpha},\alpha}\} \subset \{1, \ldots, n_c\}$ for phase α, the partially evaluated

residual and the left subspace basis can be denoted by $\mathbf{P}^T \mathbf{r}^k$ and $\mathbf{P}^T \mathbf{J}^k \mathbf{\Phi}_x$, respectively. Here, $\mathbf{P}$ is the restriction matrix denoted by

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_w & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_o \end{bmatrix}, \tag{2.12}$$

where $\mathbf{P}_w = [\mathbf{e}_{i_{1,w}}, \mathbf{e}_{i_{2,w}}, ..., \mathbf{e}_{i_{n_{i,w},w}}]$ and $\mathbf{P}_o = [\mathbf{e}_{i_{1,o}}, \mathbf{e}_{i_{2,o}}, ..., \mathbf{e}_{i_{n_{i,o},o}}]$ are the restriction matrices to the columns associated with $\mathscr{I}_w$ and, $\mathscr{I}_o$ respectively, and $\mathbf{e}_{i_{j,\alpha}} = [0, ..., 0, 1, 0, ..., 0]^T \in \mathbb{R}^{n_c}$ is the $\mathbf{i}_{j,\alpha}$-th column of the identity matrix $\mathbf{I}_{n_c} \in \mathbb{R}^{n_c \times n_c}$ for $j = 1, ..., n_{i,\alpha}$. A greedy algorithm for computing the sample indices $\mathscr{I}_\alpha$ by inductively minimizing the error associated with the residual reconstruction through 'gappy' POD is described in Algorithm 2. In the algorithm, $n_{r,\alpha}$ and $n_{J,\alpha}$ are the numbers of basis vectors for the residual $\mathbf{r}^k$ and the left subspace basis $\mathbf{J}^k \mathbf{\Phi}_x$ corresponding to phase α, respectively, and the terms e_R and e_J can be viewed as the error between the input basis and its gappy reconstruction. Therefore, in line 4 of the algorithm, the sample index is chosen in such a way that the mean squared gappy reconstruction error for the residual and Jacobian POD bases is minimized. In addition, p_r and p_J in line 7 are introduced to constrain the reconstruction in lines 8 and 9 such that it only be conducted with the basis vectors that will actually be used online. As the final outcome, sample indices are selected to minimize the upper bound of the approximation error when the approximation is required to be exact at the sample indices (Carlberg et al., 2011, 2013). The algorithm also has a feature enabling the inclusion of 'seeded nodes' that are pre-

assigned to be the grid blocks containing wells. As discussed by Carlberg et al. (2013), the grid blocks containing wells are strategic locations because the values of the residuals corresponding to them are under direct influence of the well controls. In a similar context, Cardoso and Durlofsky (2010) presented a POD-TPWL procedure for reduced-order modeling in which they exclude the grid blocks containing wells from the reduced subspace and solve Equation (2.7) for those blocks in the original space.

Algorithm 2. A greedy algorithm for constructing sample indices.

Input: $\mathbf{\Phi}_{r,\alpha}, \mathbf{\Phi}_{J,\alpha}, n_{r,\alpha}, n_{J,\alpha}, n_{i,\alpha}$, seeded sample-node set $\mathscr{S}_{\alpha}$

Output: sample-node set $\mathscr{S}_{\alpha}$

   1. $m = 1$

   2. $e_R \leftarrow \phi_{r,\alpha}^1$, $e_J \leftarrow \phi_{J,\alpha}^1$

   3. while $|\mathscr{S}_{\alpha}| < n_{i,\alpha}$

   4. $i \leftarrow \underset{j \in \{1,\ldots,n_c\}/\mathscr{S}_{\alpha}}{\arg\max} \left[e_R(j)^2 + e_J(j)^2 \right]$

   5. $\mathscr{S}_{\alpha} \leftarrow \mathscr{S}_{\alpha} \cup i$

   6. $m = m + 1$

   7. $p_r = \min(m-1, n_{r,\alpha})$, $p_J = \min(m-1, n_{J,\alpha})$

   8. $e_R \leftarrow \phi_{r,\alpha}^m - \begin{bmatrix} \phi_{r,\alpha}^1 & \cdots & \phi_{r,\alpha}^{p_r} \end{bmatrix} \begin{bmatrix} \mathbf{P}^T \phi_{r,\alpha}^1 & \cdots & \mathbf{P}^T \phi_{r,\alpha}^{p_r} \end{bmatrix}^\dagger \mathbf{P}^T \phi_{r,\alpha}^m$

   9. $e_J \leftarrow \phi_{J,\alpha}^m - \begin{bmatrix} \phi_{J,\alpha}^1 & \cdots & \phi_{J,\alpha}^{p_J} \end{bmatrix} \begin{bmatrix} \mathbf{P}^T \phi_{J,\alpha}^1 & \cdots & \mathbf{P}^T \phi_{J,\alpha}^{p_J} \end{bmatrix}^\dagger \mathbf{P}^T \phi_{J,\alpha}^m$

 10. end while

With the sample matrix $\mathbf{P}$, the surrogate approximations of the nonlinear residual and the left subspace basis can be respectively written as

$$\tilde{\mathbf{r}}^k = \underset{\mathbf{R} \in range(\mathbf{\Phi}_r)}{\arg\min} \left\| \mathbf{P}^T \mathbf{r} - \mathbf{P}^T \mathbf{R} \right\|_2, \tag{2.13}$$

$$\widetilde{\mathbf{J}^k \mathbf{\Phi}}_{\mathbf{x}} = \underset{\mathbf{L} \in range(\mathbf{\Phi}_{\mathbf{J}})^l}{\arg\min} \left\| \mathbf{P}^T \mathbf{J}^k \mathbf{\Phi}_{\mathbf{x}} - \mathbf{P}^T \mathbf{L} \right\|_F, \tag{2.14}$$

where

$$\mathbf{\Phi}_{\mathbf{r}} = \begin{bmatrix} \mathbf{\Phi}_{\mathbf{r}.w} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_{\mathbf{r}.o} \end{bmatrix}, \tag{2.15}$$

$$\mathbf{\Phi}_{\mathbf{J}} = \begin{bmatrix} \mathbf{\Phi}_{\mathbf{J}.w} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_{\mathbf{J}.o} \end{bmatrix}. \tag{2.16}$$

Here, $\mathbf{\Phi}_{\mathbf{r},\alpha} = [\phi_{\mathbf{r},\alpha}^1 \ \cdots \ \phi_{\mathbf{r},\alpha}^{n_{\mathbf{r},\alpha}}] \in \mathbb{R}^{n_c \times n_{\mathbf{r},\alpha}}$ and $\mathbf{\Phi}_{\mathbf{J},\alpha} = [\phi_{\mathbf{J},\alpha}^1 \ \cdots \ \phi_{\mathbf{J},\alpha}^{n_{\mathbf{J},\alpha}}] \in \mathbb{R}^{n_c \times n_{\mathbf{J},\alpha}}$ are basis matrices whose spans are assumed to contain the parts of $\mathbf{r}^k$ and $\mathbf{J}^k \mathbf{\Phi}_{\mathbf{x}}$, respectively, according to the phase α. Each of the matrices is computed by applying Algorithm 1 to a set of snapshots; several procedures for collecting the snapshots can be found in Carlberg et al. (2011). The solutions to the least-square problems in Equations (2.13) and (2.14) can be respectively written as

$$\tilde{\mathbf{r}}^k = \mathbf{\Phi}_{\mathbf{r}} (\mathbf{P}^{\mathbf{T}} \mathbf{\Phi}_{\mathbf{r}})^{\dagger} \mathbf{P}^T \mathbf{r}^k, \tag{2.17}$$

$$\widetilde{\mathbf{J}^k \mathbf{\Phi}}_{\mathbf{x}} = \mathbf{\Phi}_{\mathbf{J}} (\mathbf{P}^T \mathbf{\Phi}_{\mathbf{J}})^{\dagger} \mathbf{P}^T \mathbf{J}^k \mathbf{\Phi}_{\mathbf{x}}, \tag{2.18}$$

where the superscript † designates the left pseudoinverse. Therefore, $\tilde{\mathbf{r}}^k$ and the columns

of $\widetilde{\mathbf{J}^k\boldsymbol{\Phi}_{\mathbf{x}}}$ are computed to minimize the approximation errors at the sample indices

$\| \mathbf{P}^T\mathbf{r}^k - \mathbf{P}^T\tilde{\mathbf{r}}^k \|_2$ and $\| \mathbf{P}^T\mathbf{J}^k\boldsymbol{\Phi}_{\mathbf{x}} - \mathbf{P}^T\widetilde{\mathbf{J}^k\boldsymbol{\Phi}_{\mathbf{x}}} \|_F$, respectively. The approximations of $\tilde{\mathbf{r}}^k$ and

$\widetilde{\mathbf{J}^k\boldsymbol{\Phi}_{\mathbf{x}}}$ replace the reduced Equation (2.9) with

$$
\begin{aligned}
\boldsymbol{\delta}_{\mathbf{z}}^k &= \arg\min_{\Delta\in\mathbb{R}^l} \| \widetilde{\mathbf{J}^k\boldsymbol{\Phi}_{\mathbf{x}}} \Delta + \tilde{\mathbf{r}}^k \|_2 \\
&= \arg\min_{\Delta\in\mathbb{R}^l} \| \boldsymbol{\Phi}_{\mathbf{J}}(\mathbf{P}^T\boldsymbol{\Phi}_{\mathbf{J}})^\dagger \mathbf{P}^T\mathbf{J}^k\boldsymbol{\Phi}_{\mathbf{x}}\Delta + \boldsymbol{\Phi}_{\mathbf{r}}(\mathbf{P}^T\boldsymbol{\Phi}_{\mathbf{r}})^\dagger \mathbf{P}^T\mathbf{r}^k \|_2 \\
&= \arg\min_{\Delta\in\mathbb{R}^l} \| (\mathbf{P}^T\boldsymbol{\Phi}_{\mathbf{J}})^\dagger \mathbf{P}^T\mathbf{J}^k\boldsymbol{\Phi}_{\mathbf{x}}\Delta + \boldsymbol{\Phi}_{\mathbf{J}}^T\boldsymbol{\Phi}_{\mathbf{r}}(\mathbf{P}^T\boldsymbol{\Phi}_{\mathbf{r}})^\dagger \mathbf{P}^T\mathbf{r}^k \|_2,
\end{aligned}
\tag{2.19}
$$

transforming the Petrov–Galerkin iterations of Equations (2.10) and (2.11) into the following GNAT iterations:

$$
(\mathbf{P}^T\boldsymbol{\Phi}_{\mathbf{J}})^\dagger \mathbf{P}^T\mathbf{J}^k\boldsymbol{\Phi}_{\mathbf{x}}\boldsymbol{\delta}_{\mathbf{z}}^k = -\boldsymbol{\Phi}_{\mathbf{J}}^T\boldsymbol{\Phi}_{\mathbf{r}}(\mathbf{P}^T\boldsymbol{\Phi}_{\mathbf{r}})^\dagger \mathbf{P}^T\mathbf{r}^k,
\tag{2.20}
$$

$$
\mathbf{z}^{k+1} = \mathbf{z}^k + \boldsymbol{\delta}_{\mathbf{z}}^k.
\tag{2.21}
$$

Here, the matrices $(\mathbf{P}^T\boldsymbol{\Phi}_{\mathbf{J}})^\dagger \in \mathbb{R}^{(n_{J,w}+n_{J,o})\times(n_{i,w}+n_{i,o})}$ and $\boldsymbol{\Phi}_{\mathbf{J}}^T\boldsymbol{\Phi}_{\mathbf{r}}(\mathbf{P}^T\boldsymbol{\Phi}_{\mathbf{r}})^\dagger \in \mathbb{R}^{(n_{J,w}+n_{J,o})\times(n_{i,w}+n_{i,o})}$

are precomputed in the offline stage. Additionally, for each grid point, the nonlinear

residual only depends on the state solutions for nodes in its immediate neighborhood,

rendering the Jacobian matrix very sparse. Consequently, the sampled residuals $\mathbf{P}^T\mathbf{r}^k$

depend only on a small subset of the state vector $\mathbf{x}$ according to the nodal connectivity, and the GNAT method updates only these entries of the state vector. As a result, the computation for the GNAT iteration in Equations (2.20) and (2.21) can be rewritten as

$$(\mathbf{P}^T \mathbf{\Phi_J})^\dagger \mathbf{P}^T \mathbf{J}^k \overline{\mathbf{P}} \overline{\mathbf{P}}^T \mathbf{\Phi_x} \delta_z^k = -\mathbf{\Phi_J}^T \mathbf{\Phi_r} (\mathbf{P}^T \mathbf{\Phi_r})^\dagger \mathbf{P}^T \mathbf{r}^k, \tag{2.22}$$

$$\mathbf{z}^{k+1} = \mathbf{z}^k + \delta_z^k. \tag{2.23}$$

where $\overline{\mathbf{P}}$ is the restriction of the identity matrix to its columns associated with the set that influences the residual entries corresponding to the sample-index set $\mathscr{I} = \mathscr{I}_w \oplus \mathscr{I}_o$. Consequently, the partial computation of the state does not scale with the large full-order dimension n_c, reducing the CPU cost significantly in solving the minimization problem of Equation (2.7).

Note that the actual computation in the GNAT iteration does not make the literal use of the restriction matrices $\mathbf{P}$ and $\overline{\mathbf{P}}$ lest the full evaluation of the residual and Jacobian be conducted. Instead, the sample indices corresponding to the restriction matrices are used to specify the entries to be evaluated. This renders the GNAT procedure intrusive in terms of implementation into a reservoir simulator in that it requires access to the residual and Jacobian. For the oil–water model in this study, the residual and Jacobian in the full-order simulator are modified so that only a small subset of their rows and columns is evaluated according to the sample indices.

2.3 Case Study

Two case studies were performed to investigate several aspects in constructing and using reduced-order models. In the first case study, a two-dimensional oil–water reservoir model is considered, and we investigate the appropriateness of the strategy of collecting snapshots in the form of the state increments $\mathbf{x}^n - \mathbf{x}^{n-1}$. The strategy is compared to collecting snapshots in the form of states, which is the common approach for projection-based reduced-order reservoir modeling. In the second case study, a three-dimensional oil–water reservoir model is considered to investigate the numerical efficiency of using the proposed hyper-reduction method, GNAT, compared to when only the projection-based order reduction method is applied. In addition, the gravitational effects on the reduced-order model are investigated. The two case studies are sequentially presented in the following two subsections.

2.3.1 Case 1: Two-dimensional SPE10

The oil–water model in this work has the states consisting of solutions for p and s in all grid blocks. In most studies of reduced-order reservoir modeling, The collection of the states at particular times in the simulation forms the snapshot matrix, and applying POD to the snapshot matrix constructs the basis matrix. However, as indicated in Section 3.1, the unknowns sought by a Newton solver are the increments in the state, implying that the snapshots need to be collected in the form of the state increments, $\mathbf{x}^n - \mathbf{x}^{n-1}$ not the state $\mathbf{x}^n$ itself.

To assess the two snapshot collecting strategies in terms of predictability with respect to changes in well controls, a case study using a two-dimensional geologic model was

performed. The geologic model is the top layer of the SPE-10 benchmark permeability field (Christie and Blunt, 2001). The model contains 13,200 grid blocks (with $n_x = 60$, $n_y = 220$, and $n_z = 1$, where n_k is the number of grid blocks in direction k). The physical dimension of each grid block is $20\,ft \times 10\,ft \times 2\,ft$. There are six production wells and two injection wells, i.e., eight total well blocks, in a five-spot pattern, i.e., four production wells surrounding an injection well. All of the wells are under the BHP controls. The permeability in the x-direction and the well locations are shown in Figure 2.1. Permeability is taken to be a diagonal tensor, with $k_x = k_y$. The initial oil and water saturations are 0.8 and 0.2, respectively and the residual oil (s_{or}) and water (s_{wr}) saturations are 0.2. For oil, we set $\rho_o = 55\ \text{lb/ft}^3$ and $\mu_o = 5\,\text{cP}$; for water, $\rho_w = 64\ \text{lb/ft}^3$ and $\mu_w = 1\,\text{cP}$. The system is assumed incompressible, and capillary pressure and gravity effects are neglected. The relative permeability relationships are specified as

$$k_{ro}(s_w) = k_{ro}^0 \left(\frac{1 - s_w - s_{or}}{1 - s_{wr} - s_{or}} \right)^a , \tag{2.24}$$

$$k_{rw}(s_w) = k_{rw}^0 \left(\frac{s_w - s_{wr}}{1 - s_{wr} - s_{or}} \right)^b , \tag{2.25}$$

where k_{ro}^0 and k_{rw}^0 are the endpoint relative permeabilities. Here, we set $k_{ro}^0 = k_{rw}^0 = 1$ and $a = b = 2$.
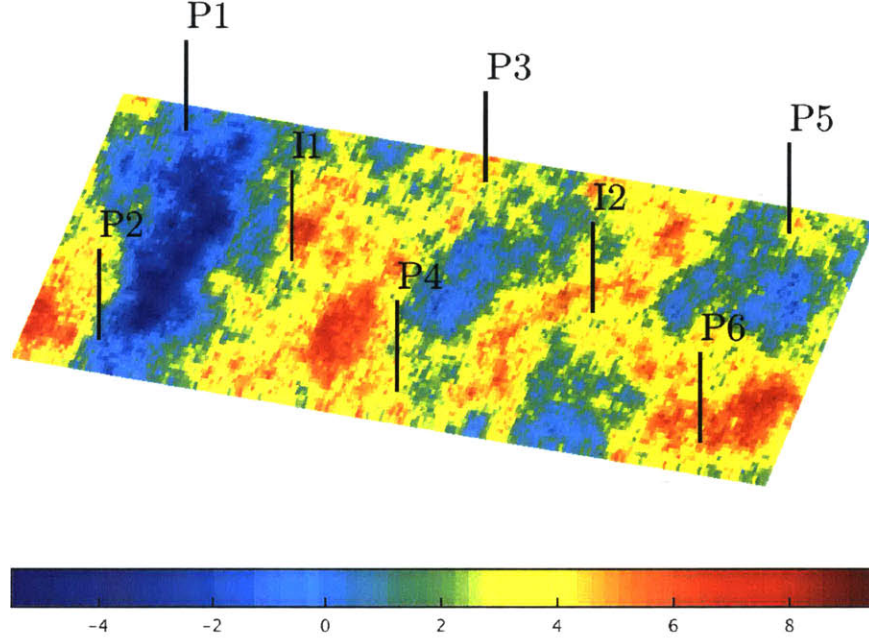
Figure 2.1: Log-permeability in the x-direction of the SPE 10 reservoir model (the top layer) with two injectors and six producers.

The BHPs of a training set of wells are devised to create a training run for constructing the reduced-order models. We specify the two injection wells to inject water at a constant BHP of 6500 psi, while the six production wells constantly produce at 3500 psi. The training case was simulated for 1500 days, collecting the states $\mathbf{x}^n$ at every time step. Then, the two snapshot matrices were assembled from the state increments $\mathbf{x}^n - \mathbf{x}^{n-1}$ and the states $\mathbf{x}^n$ themselves. By applying POD—described in Algorithm 1—to the snapshot matrices, the basis matrices $\mathbf{\Phi}_\mathbf{x}$ were constructed with $l_p = 120$ and $l_s = 240$. The number of basis vectors was selected by limited numerical experimentation. The basis matrices $\mathbf{\Phi}_\mathbf{x} \in \mathbb{R}^{2n_c \times l}$, where $l = l_p + l_s$, have $120 + 240 = 360$ basis vectors, meaning that the number of unknowns in the model $2n_c$ decreased from 26,400 to 360. To compare the basis matrices in terms of predictability with respect to changes in well control

parameters, a time-varying test set of BHPs for the wells, which is shown in Figure 2.2 with the training schedule, was randomly generated within the range 6000–7000 psi for the injection wells and 3000–4000 psi for the production wells. The schedule for each well is perturbed every 150 days. Figure 2.2 shows that the training schedule, taken to be constant for all wells, is completely different from the test schedule.

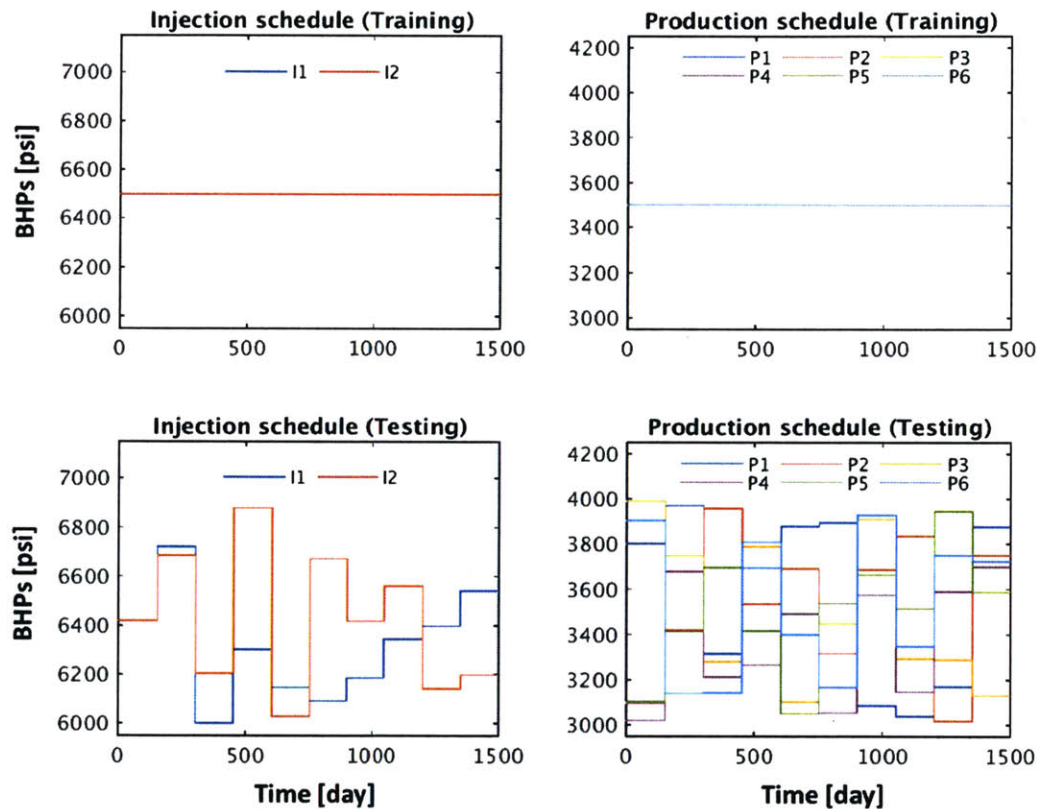**Figure 2.2: Schedule of well controls for the two injection and four production wells in Case 1.**

The performance of the reduced-order models with the two different snapshot collection strategies is measured by quantifying the mismatch of the well flow rates for both oil and water between the full-order solution and reduced-order solutions. The mismatch measure for oil is defined as (He et al., 2013)

$$E_o = \frac{1}{n_{pw}} \sum_{j=1}^{n_{pw}} \frac{\int_0^T \left| Q_{o,FOM}^j - Q_{o,ROM}^j \right| dt}{\int_0^T Q_{o,FOM}^j dt} , \qquad (2.26)$$

where n_{pw} is the number of production wells, T is the total simulation time, and $Q_{o,FOM}$ and $Q_{o,ROM}$ are the oil production rates simulated by the full-order and reduced-order model, respectively. This manner of measuring the mismatch is also applied for evaluating the water production mismatch E_w and the water injection mismatch E_{iw}.

The error measurement of the state solution averaged over all time steps and all grid blocks is also evaluated. The average error measurement for pressure is defined as (He et al., 2015)

$$E_p = \frac{1}{n_t n_c} \sum_{i=1}^{n_t} \sum_{j=1}^{n_c} \left| p_{j,FOM}^i - p_{j,ROM}^i \right| , \qquad (2.27)$$

where n_t is the number of time steps in the simulation, $p_{k,FOM}^i$ is the full-order pressure solution for block j at time step i, and $p_{j,ROM}^i$ is the analogous quantity for the reduced order models. Equation (2.27) can be applied to evaluate the error for saturation.

The Petrov–Galerkin projection method, which outperforms Galerkin projection (He and Durlofsky, 2015), is adopted, and the two strategies of collecting snapshots are compared in terms of predictability with respect to changes in well controls. Figure 2.3 shows the final saturation fields obtained from the full-order model and the two Petrov–Galerkin projection-based models: one with raw state snapshots and one with state

increment snapshots. Good agreement is observed between the full-order simulation result and the two reduced-order simulation results. The average errors in the final saturation field and the pressure field defined in Equation (2.27) are shown Table 2.1. While both strategies of collecting snapshots show good agreement, the error measures indicate that the snapshots consisting of state increments provide slightly better accuracy.

(a) Full order simulation   (b) Petrov–Galerkin (raw snapshots)   (c) Petrov–Galerkin (increment snapshots)

Figure 2.3: The final saturation simulated by the full-order simulation and the two Petrov–Galerkin projection-based reduced-order model with different types of snapshots.

Table 2.1: Summary of mismatch measurements of Petrov–Galerkin projection-based reduced-order models (Case 1).

Snapshots	E_o	E_w	E_{iw}	E_p [psi]	E_s	Speedup
Raw State	0.0284	0.0378	0.0232	13.3139	0.0086	7.5
State Increment	0.0160	0.0161	0.0103	6.7228	0.0054	7.7

The injected water rates in the two injection wells are shown in Figure 2.4. Figure 2.5 shows the simulation solutions of the oil and water production rates for the six production wells using the training and the test schedules when the snapshots consist of the raw states. Figure 2.6 shows the results when the snapshots are collected in the form of state increments. Overall, the two snapshot collecting strategies again show general agreements with the full-order simulation results. However, in the first injection well and the third production well, the water injection and production solutions show poorer agreement with the full-order method solution when the snapshots comprise the raw states than the other strategy, while the rate predictions in other wells are in overall agreement. The mismatch measurement calculated from Equation (2.26) is shown in Table 2.1, and all the mismatch measurements for when the snapshot matrix consists of the state increments are less than when the raw states are used for the snapshot matrix. In particular, the water injection mismatch measurement E_{iw} and the production mismatch measurement E_w nearly double when constructing the snapshot matrix with the raw states than with the state increments. The performance analysis using a different number of basis vectors also supports the state increment collection over raw state collection as a snapshot in terms of accuracy. Clearly, collecting the state increments as snapshots shows better performance in terms of predictability than collecting the states. However, we note that the use of state increments as snapshots does not give radical improvements of the reduced-order simulation, as shown in Figures 2.4–2.6.

Figure 2.4: Water injection simulated with the Petrov–Galerkin projection-based reduced-order model with different types of snapshots.

Figure 2.5: Oil and water production simulated with the Petrov–Galerkin projection-based reduced-order model when the snapshots consist of raw states.

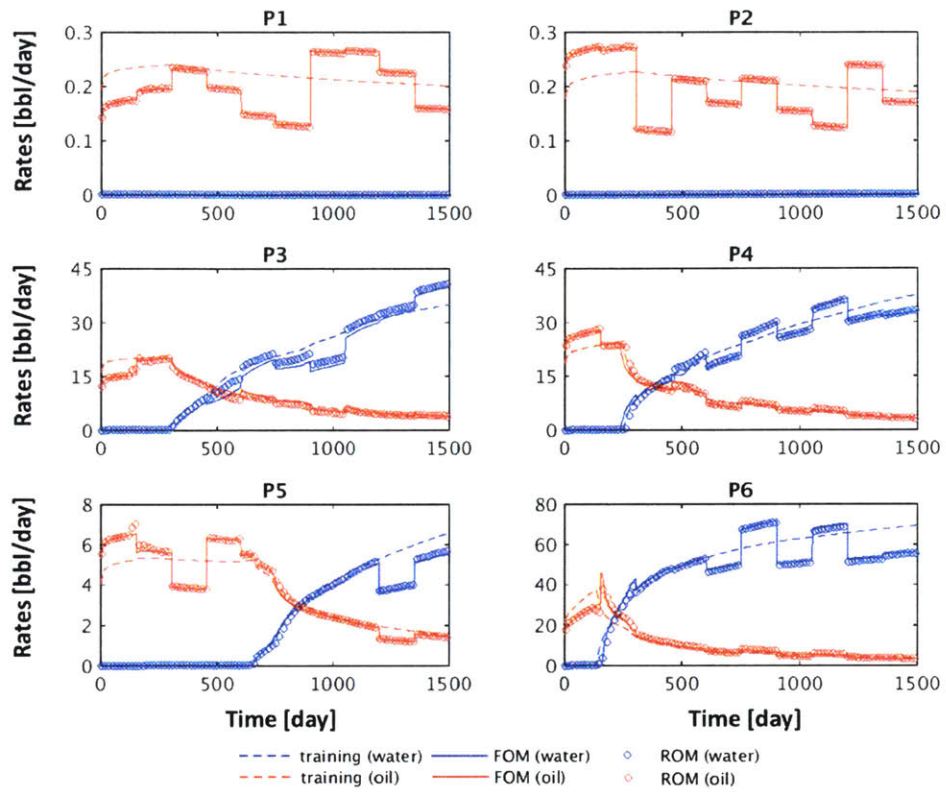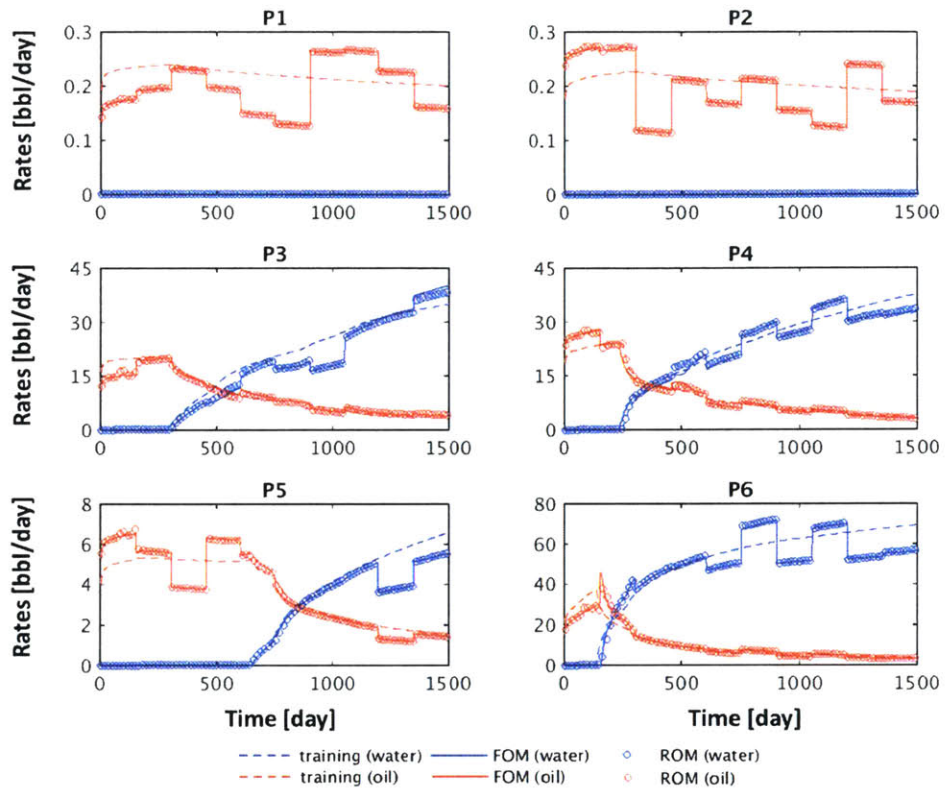**Figure 2.6: Oil and water production simulated with the Petrov–Galerkin projection-based reduced-order model when the snapshots consist of state increments.**

As mentioned earlier, order reduction based only on the projection is limited for lowering the computational cost because it is still a function of the full-order model dimension n_c. Furthermore, order reduction based on Petrov–Galerkin projection requires an additional matrix multiplication compared to Galerkin projection-based order reduction because the basis vectors of its low-order subspace $\mathbf{J}^k \mathbf{\Phi}_x$ in the least-square problem of Equation (2.9) are variable and state-dependent over simulation, which causes additional computational time. This is demonstrated by the fact that the speedup factor achieved by the two aforementioned reduced-order models is approximately 10 when compared with the full-order model as shown in Table 2.1. The speedup factor is computed by comparing only the computation time spent in the online stage, and the preprocessing overhead required for the Petrov–Galerkin projection-based models is approximately equal to the time required to simulate one full-order model. In the following section, a case study applying the hyper-reduction method based on GNAT is presented and its performance in lowering the computational cost is demonstrated.

2.3.2 Case2: Three-dimensional SPE10

To investigate the computational efficiency and the predictability of the GNAT-based reduced-order model under gravitational effects, a three-dimensional geologic model is considered: a portion of the top five layers of the SPE10 benchmark model as shown in Figure 2.7. The model contains 33,000 grid blocks ($n_x = 60$, $n_y = 110$, and $n_z = 5$), whereas a practical industrial models typically contains approximately 10^5–10^6 grid blocks (Margot and Durlofsky, 2005). The well placement is the same as in Case 1 of the

previous section, but the two injection wells also have a single perforation at the bottom layer and the six production wells also have a single perforation at the top layer.
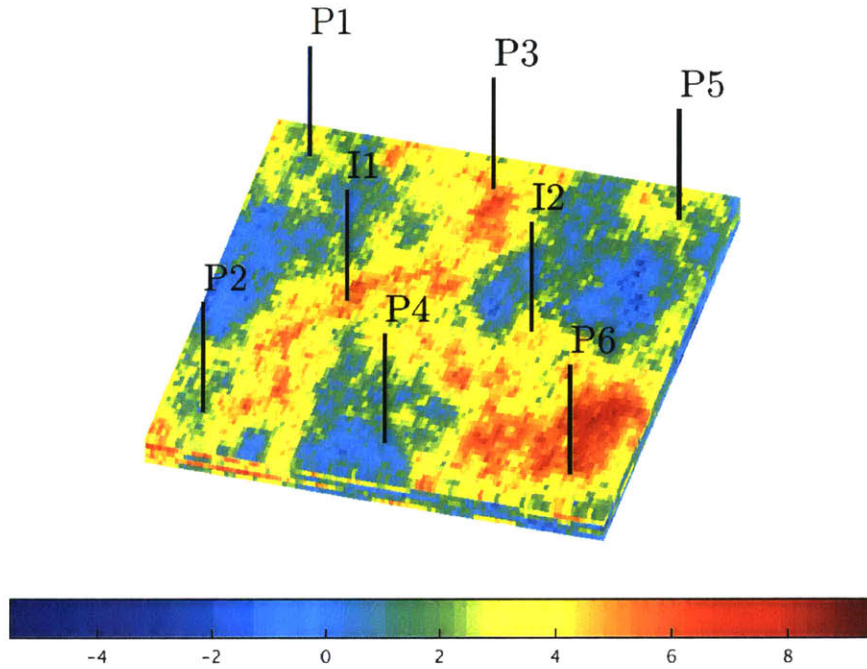

Figure 2.7: Log-permeability in the x-direction of a portion of the SPE 10 reservoir model (5 layers) with two injectors and six producers.

Two reduced-order models are considered: the Petrov–Galerkin projection-based reduced-order model and the GNAT-based hyper-reduced model. To construct the POD basis matrix $\mathbf{\Phi}_x$, the training set of BHPs for the wells from Case 1 is used again as shown in Figure 2.8, and snapshots in the form of state increments $\mathbf{x}^n - \mathbf{x}^{n-1}$ are collected over 1500 days. Then, the basis matrix $\mathbf{\Phi}_x$ is constructed by applying POD to the snapshot matrix with $l_p = 150$ and $l_s = 225$, meaning that the number of model unknowns is reduced from $2n_c = 66,000$ to $l_p + l_s = 375$. By limited numerical

experimentation, we set $n_{r,o} = n_{r,w} = 360$, $n_{J,o} = n_{r,w} = 360$, and $n_{i,w} = n_{i,o} = 720$. Note that to construct the GNAT-based reduced-order model, an additional training simulation with Petrov–Galerkin projection-based reduced-order simulation needs to be conducted (Carlberg et al., 2011). Through the additional training simulation, the snapshots for constructing the basis matrices Φ_r and Φ_J are constructed. More details of the snapshot collection procedures can be found in the work by Carlberg et al. (2011). The sampling indices $\mathcal{I}_w$ and $\mathcal{I}_o$ selected by applying Algorithm 2 on Φ_r and Φ_J are plotted in Figure 2.9. Note that we do not use the scheme of 'seeded nodes' in this case study.
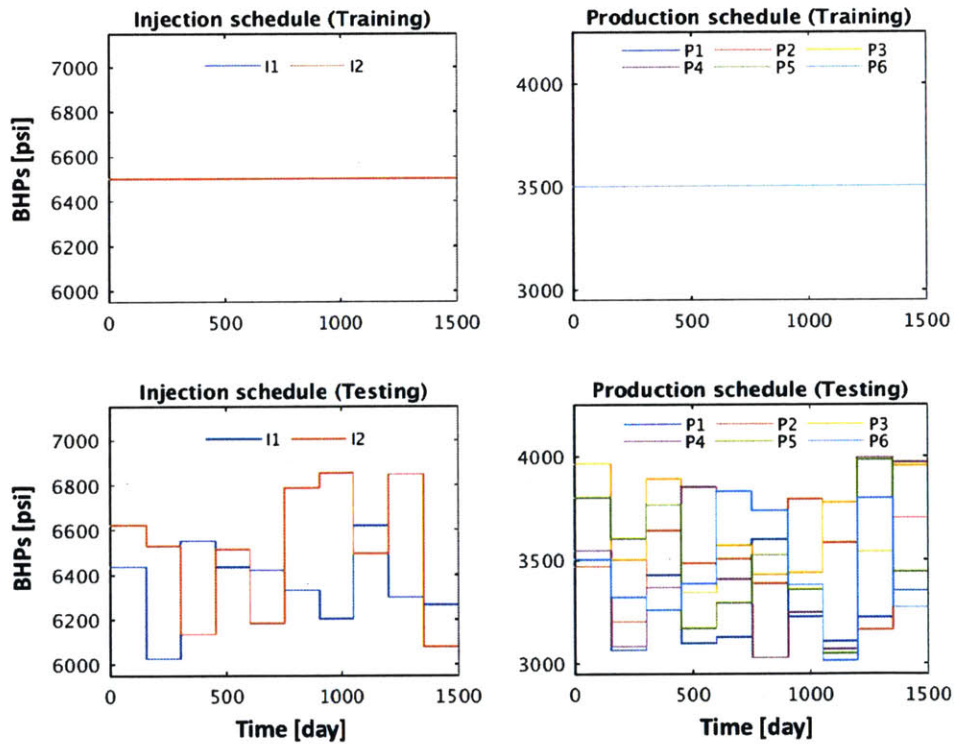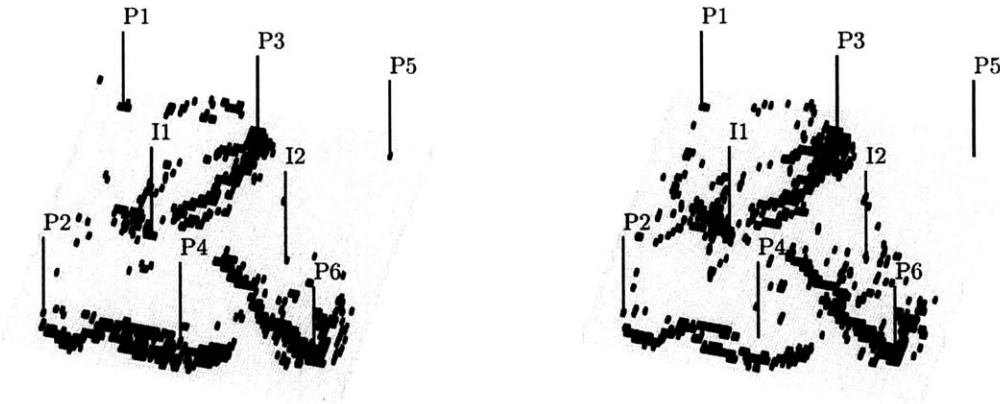

Figure 2.8: Schedule of well controls for the two injection and four production wells in Case 2.

A new time-varying test set of BHPs for the well was randomly generated using the same range as Case 1; this schedule is shown in Figure 2.8. The final saturation fields obtained from the two reduced-order models are in good agreement as shown in Figure 2.10, while the error estimates summarized in Table 2.2 show better accuracy from using the Petrov–Galerkin projection-based reduced-order model. The injected water rates in the two injections wells are shown in Figure 2.11. Figure 2.12 shows the simulation solutions of the oil and water production rates for the six production wells when only the Petrov–Galerkin projection is used, whereas Figure 2.13 shows the result when the proposed hyper-reduction method, GNAT, is adopted. The mismatch measures are summarized in Table 2.2. Overall, both reduced-order models show general agreement with the full-order simulation results, but the GNAT-based reduced-order model shows slightly less accuracy because it replaces the nonlinear residual and the action of the Jacobian with surrogates. However, the computational time speedup factor for this method (~300) is significantly larger than for the Petrov–Galerkin projection method (~10). Such high speedup cannot be achieved by applying only the Petrov–Galerkin projection method because of its dependency on the full-order model. The preprocessing overhead required for the Petrov–Galerkin projection-based model is approximately equal to the time required to simulate one full-order model as in Case 1, whereas the overhead for constructing the GNAT-based hyper-reduced order model is approximately 1.5 times that spent by a full-order model. The additional overhead computation time compared to the overhead for the Petrov–Galerkin projection-based model is used for the additional training simulation to construct the basis matrices Φ_r and Φ_J, and to precompute the matrices in Equation (2.20).

Table 2.2: Summary of mismatch measurements of Petrov–Galerkin projection-based reduced-order model and GNAT-based reduced-order model (Case 2).

	E_o	E_w	E_{iw}	E_p (psi)	E_s	Speedup
Petrov–Galerkin	0.0128	0.0439	0.0097	4.1166	0.0037	10.7
GNAT	0.0171	0.0573	0.0124	6.5960	0.0043	335.3

A noteworthy feature of the sampling indices is that while the sampling indices for oil, $\mathscr{S}_o$, are chosen to contain all of the well grid cells, the indices for water, $\mathscr{S}_w$, do not contain the grid cells of the fourth and the fifth production wells as shown in Figure 2.9(b). This is because there is no water production in the two wells; consequently, the data stored in the snapshot matrix does not indicate any significant variability of water saturation in the two grid cells, so they do not need to be selected by Algorithm 2.

(a) Sampling indices for oil conservation ($n_{i,w} = 720$)   (b) Sampling indices for oil conservation ($n_{o,w} = 720$)

Figure 2.9: Sampling indices for oil and water mass balance equations.

(a) Full order simulation   (b) Petrov–Galerkin   (c) GNAT

Figure 2.10: The final saturation simulated by the full-order model, the Petrov–Galerkin projection-based reduced-order model, and the GNAT-based reduced-order model.

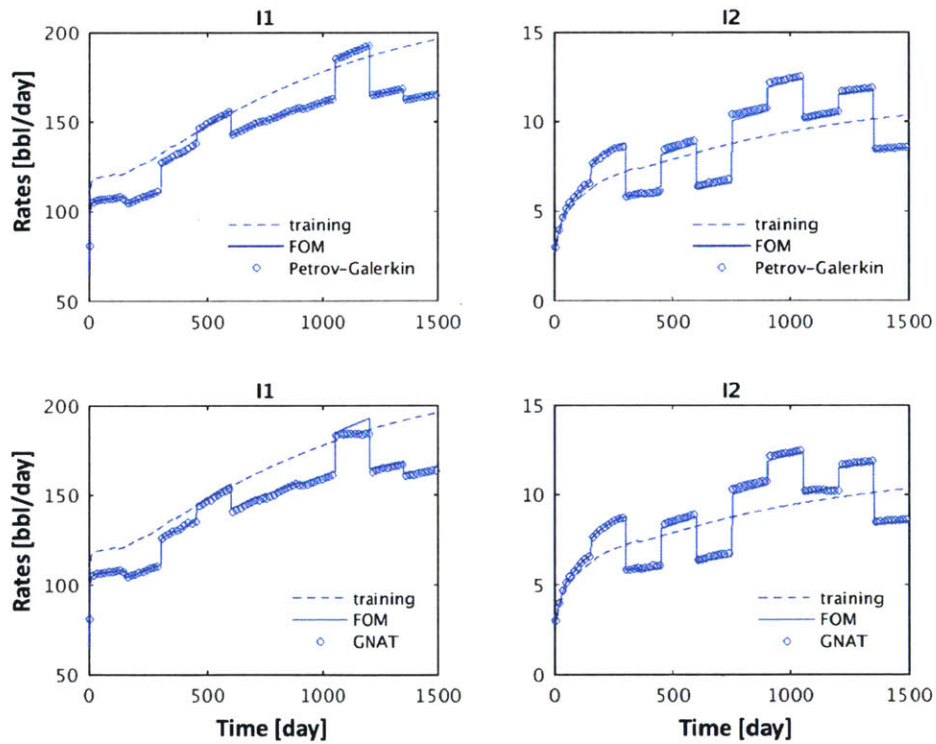Figure 2.11: Water injection rates simulated with the Petrov–Galerkin projection-based reduced-order model and the GNAT-based hyper-reduced-order model.
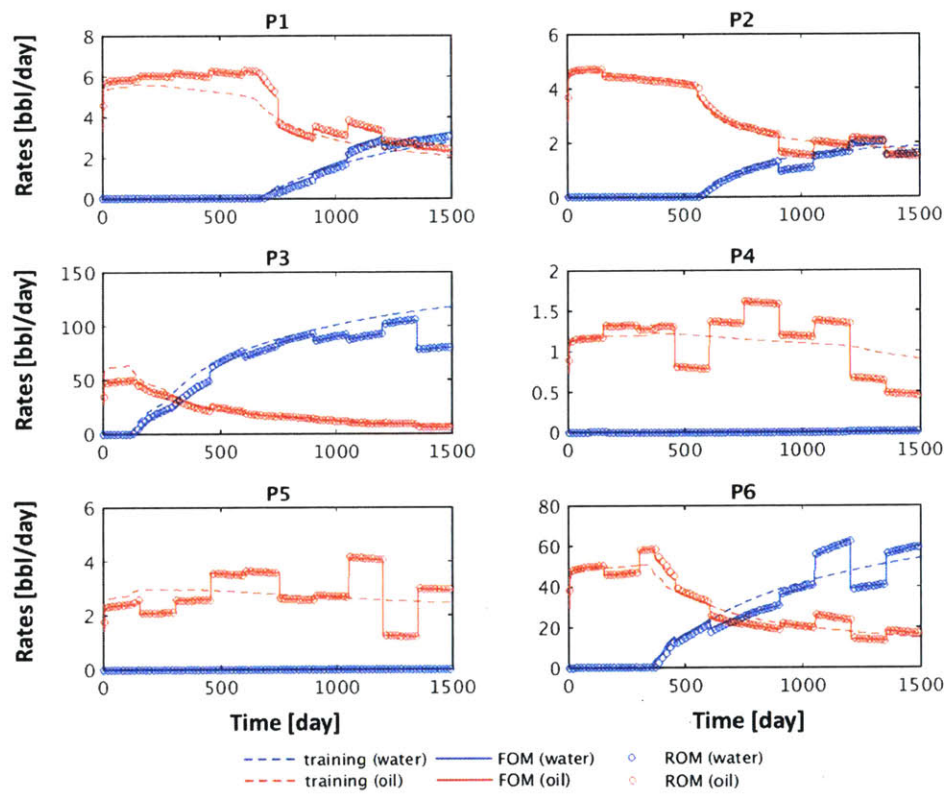
Figure 2.12: Oil and water production simulated with the Petrov–Galerkin projection-based reduced-order model.
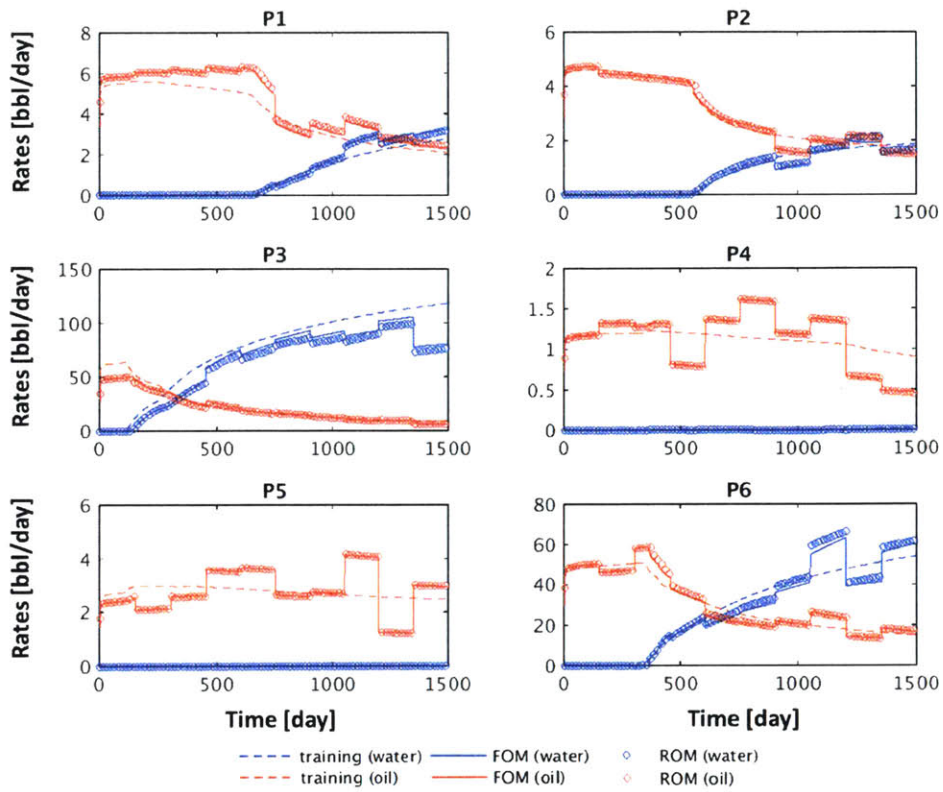
Figure 2.13: Oil and water production simulated with the GNAT-based hyper-reduced-order model.

In addition, an issue of inaccuracy in reduced-order subsurface flow modeling has been reported when gravitational effects are considered. Under gravitational effects, counter-current flow can occur in the vertical direction according to the local competition between convective and gravitational effects, and this causes more localized variability in the saturation field (Cardoso and Durlofsky, 2010). However, since we apply the GNAT procedure separately for the oil and water phases, more flexibility is available to handle the local variability and to retain the accuracy as demonstrated in the above results.

The oil–water model in this study includes the assumption of incompressible fluid and rock. If compressibility is introduced into the model, it increases the nonlinearity of the system because it causes density and porosity variance with pressure and possibly temperature as in a thermal recovery system (Rousset et al., 2014). Consequently, in view of GNAT procedure, the accuracy and stability in the gappy reconstruction of the nonlinear residual and the left subspace basis will be poor unless more sample indices are introduced, which will degrade the speedup. A potential remedy for this problem is the use of several local reduced-order bases, each tailored to a particular region of characteristic nonlinear behavior caused by compressibility. The concept of local reduced-order bases was introduced by Amsallem et al., (2012) for GNAT and Peherstorfer et al. (2014) for DEIM. By using multiple local GNAT approximations, we can expect that the hyper-reduced-order models can handle the local variability of the nonlinear behavior. The performance of this localized approach will be of interest for evaluation in future work.

2.4 Summary

In this study, we developed a reduced-order modeling procedure based on GNAT for oil–water two-phase subsurface flow simulation, and the strategy of collecting snapshots in the form of state increments was investigated in terms of model predictability and accuracy. The result shows that the proposed order reduction method based on GNAT enables the simulation to be conducted in a manageable time without much loss of accuracy, even when a large three-dimensional geologic model with gravitational effects is considered. However, it should be noted that the predictability of the reduced-order model can be retained when the discrepancy between training and test well controls is within a tolerable range. In addition, the snapshots comprised of state increments show less error compared to the snapshots comprised of raw states. The hyper-reduction method is more than 300 times faster than the full-order method, whereas the speedup factor achieved when only constraint reduction is adopted is approximately 10. Future work should cover the further development of the proposed hyper-reduced-order modeling procedure for more general subsurface flow simulations such as compressible flow with capillary effects, idealized thermal simulation cases, and compositional systems. The performance of GNAT in nonlinearity treatment also needs to be compared to previously proposed methods such as TPWL and DEIM. In addition, the application of the hyper-reduced-order model to production optimization and history matching should also be conducted in future work.

Chapter 3

Ensemble-based History Matching

using Reduced-order Models

History matching is a procedure to adjust model parameters and states so that the history of model responses can reproduce the real observations. In the field of oil reservoir engineering, the history matching has been conducted manually although it is less efficient. Recently, a large amount of well log data has become available due to technological development in well sensors, and this poses a need of computer assisted history matching that can handle such big data. The method of Bayesian filter provides a well-suited framework for the history matching problem, and Figure 3.1 shows a flowchart of sequential Bayesian filter-based history matching.
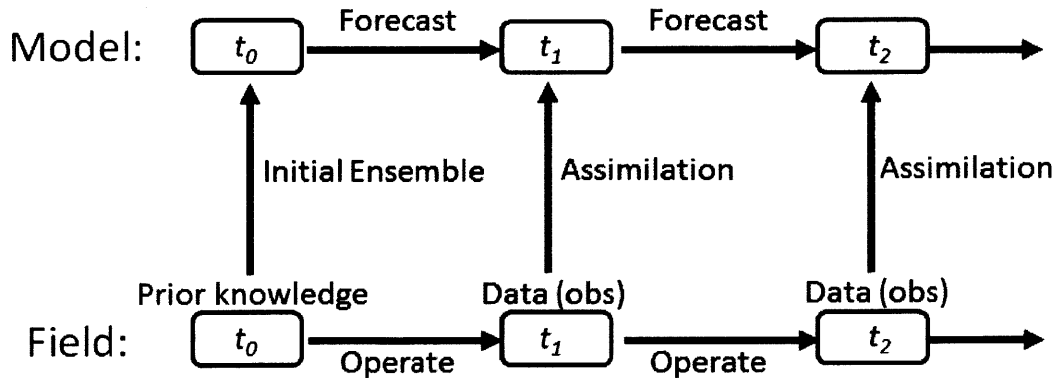
Figure 3.1: Sequential Bayesian Filter-based History Matching.

As an example of Bayesian filters, the method of ensemble Kalman filter (EnKF) has gained enormous popularity for history matching over the last decade in the field of reservoir engineering (Nævdal et al., 2002; Nævdal et al., 2005; Gu and Oliver, 2006, 2007; Aanonsen et al., 2009; Chen et al., 2009; Chen and Oliver, 2010). The EnKF is a sequential Bayesian filter assuming all the probability distributions involved with a state-space model as Gaussian. For non-Gaussian systems, the method of particle filter can be considered since it does not require any assumption of probability distribution on the model (Leeuwen, 2009). However, the method of particle filtering has rarely been considered for reservoir history matching since it is difficult to locate initial particles on highly probable region on the space of model parameters (Aanonsen et al., 2009).

In this section, the theory of Bayesian filters based on Kalman filter and particle filter are presented. Then, the embedment of the hyper-reduced-order reservoir models into the two filters are presented in the subsequent subsections. They can be divided into two parts. First, the ensemble Kalman filter method embedded by the developed reduced-order model is devised. Second, an advanced particle filter with regularization embedded by the reduced-order model is be presented.

59

3.1 Kalman Filtering for History Matching

The theory of Kalman filter is introduced in Section 3.1.1, and then a subsection describing ensemble Kalman filter (EnKF) method is followed. The way of embedding hyper-reduced-order models into the EnKF is proposed in Section 3.1.3.

3.1.1 Kalman Filter

By letting $Y \in \mathbb{R}^{n_Y}$ to denote the system state vector and assuming that it has Gaussian probability distribution with mean μ and covariance $\mathbf{C}_Y$, the pdf of the state vector can be written as

$$P(Y) \propto \exp\left(-\frac{1}{2}(Y-\mu)^T \mathbf{C}_Y^{-1}(Y-\mu)\right). \tag{3.1}$$

In sequential data assimilation, the pdf is called the prior, and it is evolved in time by running the reservoir simulation model and updated by assimilating new observation data. The data vector $\mathbf{d}$ is computed from state vector Y as $\mathbf{d} = HY + \varepsilon_\mathbf{d}$, where H is the so-called the observation matrix so that HY is what the value of the data would be for the state Y in the absence of data error $\varepsilon_\mathbf{d}$. The error is usually assumed Gaussian as $\varepsilon_\mathbf{d} \sim N(0, \mathbf{C}_\mathbf{d})$, where $\mathbf{C}_\mathbf{d}$ denotes the covariance matrix of measurement noise. Then, the probability density $p(\mathbf{d}|Y)$ of the data $\mathbf{d}$ conditioned by the system state Y, commonly called the data likelihood, is

$$P(\mathbf{d}\,|\,Y) \propto \exp\left(-\frac{1}{2}(\mathbf{d}-HY)^T \mathbf{C}_Y^{-1}(\mathbf{d}-HY)\right). \tag{3.2}$$

The Bayes theorem tells that, by combining the pdf of the state and the data likelihood, the posterior density of the system state can be derived as

$$P(Y\,|\,\mathbf{d}) \propto p(\mathbf{d}\,|\,Y)p(Y). \tag{3.3}$$

This posterior is considered as new prior density for the system state in view of sequential data assimilation. That is, by denoting the posterior state by $\hat{Y}$ instead of $Y\,|\,\mathbf{d}$ and the posterior pdf by $P(\hat{Y})$, the following posterior pdf is derived which is also Gaussian,

$$P(\hat{Y}) \propto \exp\left(-\frac{1}{2}(\hat{Y}-\hat{\mu})^T \hat{\mathbf{C}}_Y^{-1}(\hat{Y}-\hat{\mu})\right), \tag{3.4}$$

with the posterior mean $\hat{\mu}$ and covariance $\hat{\mathbf{C}}_Y$ given by the Kalman update formulas

$$\hat{\mu} = \mu + K(d - H\mu), \tag{3.5}$$

$$\hat{\mathbf{C}}_Y = (I - KH)\mathbf{C}_Y, \tag{3.6}$$

where

$$K = \mathbf{C}_Y H^T (H\mathbf{C}_Y H^T + \mathbf{C_d})^{-1} \qquad\qquad (3.7)$$

is commonly known as Kalman gain.

3.1.2 Ensemble Kalman Filter

The Ensemble Kalman filter (EnKF) is a Monte-Carlo implementation of Kalman filter. The original Kalman filter, which was developed for linear state-space system, takes the Gaussian assumption on all pdfs, and provides optimal formulas for updating the mean and covariance by the Bayesian theorem as described in the previous section.

The EnKF starts from an ensemble of initial reservoir models generated along with a priori geostatistical assumptions. The method then seeks to drive ensemble members to match production data at every assimilation step by correcting the models based on data mismatch and on the statistical correlation between members. Usually all the variables of interest are collected into a state vector Y. A typical state vector of a two-phase flow problem is composed of porosity, permeability, pressure, and water saturation at each grid-block and all the available production data as

$$Y = [\phi, \ln k, P, S_w, WPR, OPR]^T . \qquad\qquad (3.8)$$

The components of Y are containing either the static or dynamic variables at all the grid blocks or production data from all the wells. For the EnKF application, an ensemble of state vectors are collected in a matrix $\mathbf{Y}$:

$$\mathbf{Y} = \left[Y_1, Y_2, Y_3, \ldots, Y_{N_c} \right], \tag{3.9}$$

where N_e denotes the total number of ensemble members. The statistics needed for the analysis step are approximated from the ensemble matrix. At the analysis step, each ensemble member is updated by

$$Y_j^u = Y_j^f + \underbrace{\mathbf{C}_Y^f H^T \left(H \mathbf{C}_Y^f H^T + \mathbf{C_d} \right)^{-1}}_{\textit{Kalman gain}} \left(\mathbf{d}_{obs,j} - H Y_j^f \right) \quad \text{for } j = 1, \ldots, N_e, \tag{3.10}$$

where j is the index of the individual ensemble member and $\mathbf{d}_{obs,j}$ is the perturbed observation, where the perturbation is for maintaining the correct variance after updating. The superscript u indicates update and superscript f indicates forecast. $\mathbf{C_d}$ is the covariance matrix of the measurement noise. H is a matrix operator that selects measured variables from the state vector. $\mathbf{C}_Y H^T$ is the cross-covariance between all the state variables and the predicted observations, and $H\mathbf{C}_Y H^T$ is the auto-covariance of the predicted observations. The covariance of the state vector is approximated using the standard statistical formula:

$$\mathbf{C}_Y = \frac{1}{N_e - 1} \sum \left(Y_j^f - \overline{Y}^f \right) \left(Y_j^f - \overline{Y}^f \right)^T, \tag{3.11}$$

where $\bar{Y}^f$ denotes the mean of the state vectors. If there are additional data to assimilate after the update, the state vectors of Equation (3.9) are evolved to the next assimilation point using the forward simulation model, which is corresponding to the forecast step.

3.1.3 Hyper-reduced-order models for EnKF

In order to imbed the GNAT-based hyper-reduced-order model into the EnKF, certain members of the ensemble should be evolved by full-order simulation so that the corresponding reduced-order models are trained. For this, the unsupervised machine learning algorithm of k-means clustering is used.

Letting $N_{FOM} < N_e$ is the number of ensemble members to be evolved by full-order simulation, the k-means algorithm is used, at the beginning of each forecast step, in order to cluster the N_e realizations into N_{FOM} clusters according to geospatial similarity. The measure of the similarity can be defined by Euclidean distance (He et al., 2013). Then, the ensemble member closest to the centroid of each cluster is evolved by full-order siulation. These near-centroid ensemble members are expected to be reasonably well distributed over the static geological parameter space. Using these N_{FOM} simulations, N_{FOM} GNAT-based hyper-reduced-order models are trained and the rest of realizations in each cluster are simulated using the near-centroid reduced-order model. As a result, there are N_{FOM} full-order simulation runs and $(N_e - N_{FOM})$ reduced-order runs in a forecast step. Then, the newly observed data are assimilated with the predicted well responses from the both of full-order runs and the reduced-order runs as an analysis step to update state vectors. This process is repeated until there is no available observation data.

3.2 Particle Filtering for History Matching

An essential assumption in Kalman filters described in previous sections is that the prior and posterior pdfs of the system states follow Gaussian distributions. However, this assumption is often invalid in practice. In this section, a couple of sequential Bayesian filters including sequential importance resampling (SIR) and regularized particle filter (RPF), all of which are based on sequential importance sampling (SIS), is presented to deal with non-Gaussian systems. The SIS is first introduced in the next subsection followed by the subsections describing the SIR and the RPF, respectively. Finally, the framework of embedding the HRM into the RPF is presented.

3.2.1 Sequential Importance Sampling (SIS)

A general dynamic state-space model can be described as

$$Y_t \sim P(Y_t \mid Y_{t-1}), \tag{3.12}$$

$$D_t \sim P(D_t \mid Y_t), \tag{3.13}$$

$$Y_0 \sim P(Y_0), \tag{3.14}$$

where t is time step, Y_t is unknown state vector, and D_t is observation. $P(Y_t \mid Y_{t-1})$ describes the system transition in the state and implies that the state at previous time step is enough information to move forward to next time step. This is based on Markovian assumption of $P(Y_t \mid Y_{0:t-1}, D_{1:t-1}) = P(Y_t \mid Y_{t-1})$. In addition, the observations are assumed that they are conditionally independent only on the current state as

$P(D_t | Y_{0:t}, D_{1:t-1}) = P(D_t | Y_t)$. By using the Markov properties of the model and data, we get a recursive form of posterior distribution as

$$P(Y_{0:t} | D_{1:t}) \propto P(D_t | Y_{0:t}, D_{1:t-1})P(Y_{0:t} | D_{1:t-1})$$

$$= P(D_t | Y_t)P(Y_t | Y_{0:t-1}, D_{1:t-1})P(Y_{0:t-1} | D_{1:t-1})$$

$$= P(D_t | Y_t)P(Y_t | Y_{t-1})P(Y_{0:t-1} | D_{1:t-1}) \,. \qquad (3.15)$$

Since it is usually impossible to sample directly from the posterior due to its complicated formal appearance, an importance distribution $q(Y_{0:t})$ is introduced so that we can have the following decomposition of the expectation of an arbitrary function $g(Y_{0:t})$ over the posterior density $P(Y_{0:t} | D_{1:t})$

$$\int g(Y_{0:t})P(Y_{0:t} | D_{1:t})dY = \int \left[g(Y_{0:t}) \frac{P(Y_{0:t} | D_{1:t})}{q(Y_{0:t})} \right] q(Y_{0:t})dY \,. \qquad (3.16)$$

The key idea of importance sampling is to form Monte Carlo approximation to the above expectation by drawing N samples from the importance density as

$$Y_{0:t}^{(i)} \sim q(Y_{0:t}), \quad i = 1, \ldots, N, \qquad (3.17)$$

and by forming the approximation as

$$E[g(Y_{0:t}) \mid D_{1:t}] \approx \sum w_t^{(i)} g(Y_{0:t}^{(i)}) \, , \qquad\qquad (3.18)$$

where the importance weights $w_t^{(i)}$ is defined as

$$w_t^{(i)} \propto \frac{P(Y_{0:t} \mid D_{1:t})}{q(Y_{0:t})} \, . \qquad\qquad (3.19)$$

The numerator $P(Y_{0:t} \mid D_{1:t})$ can be decomposed as follows

$$P(Y_{0:t} \mid D_{1:t}) = P(Y_{0:t} \mid D_{1:t-1}, D_t)$$

$$\propto P(D_t \mid Y_{0:t}, D_{1:t-1}) P(Y_{0:t} \mid D_{1:t-1})$$

$$= P(D_t \mid Y_t) P(Y_t \mid Y_{0:t-1}, D_{1:t-1}) P(Y_{0:t-1} \mid D_{1:t-1})$$

$$= P(D_t \mid Y_t) P(Y_t \mid Y_{t-1}) P(Y_{0:t-1} \mid D_{1:t-1}) \, . \qquad (3.20)$$

In addition, the importance distribution can be decomposed as $q(Y_{0:t}) = q(Y_t \mid Y_{0:t-1}) q(Y_{0:t-1})$. Consequently, the importance weights can be rewritten as

$$w_t^{(i)} \propto \frac{P(D_t \mid Y_t^{(i)}) P(Y_t^{(i)} \mid Y_{t-1}^{(i)}) P(Y_{0:t-1}^{(i)} \mid D_{1:t-1})}{q(Y_{0:t}^{(i)})}$$

$$= \frac{P(D_t \mid Y_t^{(i)}) P(Y_t^{(i)} \mid Y_{t-1}^{(i)})}{q(Y_t^{(i)} \mid Y_{t-1}^{(i)})} \frac{P(Y_{0:t-1}^{(i)} \mid D_{1:t-1})}{q(Y_{0:t-1})} \, . \qquad (3.21)$$

Assuming that we have already drawn the samples $Y_{0:t-1}^{(i)}$ from the importance distribution $q(Y_{0:t-1})$ and computed the corresponding importance weights $w_{t-1}^{(i)}$, we can now draw samples $X_{0:t}^{(i)}$ from the importance distribution $q(Y_{0:t})$ by drawing the new state samples for the step t as $Y_t^{(i)} \sim q(Y_t \mid Y_{0:t-1}^{(i)})$. Since the importance weights from the previous step are proportional to the last term in the above equation as

$$w_t^{(i)} \propto \frac{P(Y_{0:t-1}^{(i)} \mid D_{1:t-1})}{q(Y_{0:t-1})}, \tag{3.22}$$

we can have a recursive formula for the particle weights as

$$w_t^{(i)} \propto \frac{P(D_t \mid Y_t^{(i)}) P(Y_t^{(i)} \mid Y_{t-1}^{(i)})}{q(Y_t^{(i)} \mid Y_{t-1}^{(i)})} w_{t-1}^{(i)}. \tag{3.23}$$

Note that it is convenient to select the importance distribution to be Markovian in the sense that $q(Y_t \mid Y_{1:t-1}^{(i)}) = P(Y_t \mid Y_{t-1}^{(i)})$. This means that the proposal distribution proposes from the prior, which is simply moving the particles forward with the dynamic forward model. Consequently, we have a recursive formula for the particle weights as $w_t^{(i)} \propto P(D_t \mid Y_t^{(i)}) w_{t-1}^{(i)}$. Therefore, with this multiplicative formula of updating the particle weights, it is not necessary to track all the history the states during the analysis step, and this approach is widely used due to the ease of implementation. A pseudo-code description of this algorithm is given in Algorithm 3.1.

Algorithm 3.1. SIS Particle Filter (Arulampalam et al., 2002)

1. Draw N samples $Y_0^{(i)}$ from the prior

$$Y_0^{(i)} \sim P(Y_0), \quad i = 1,...,N,$$

   and set $w_0^{(i)} = 1/N$, for $i = 1,...,N$.

2. For each $t = 1,...,T$ do the following:

   a. Draw samples $Y_t^{(i)}$ from the importance distributions (model dynamics)

   $$Y_t^{(i)} \sim P(Y_t \mid Y_{t-1}), \qquad i = 1,...,N.$$

   b. Calculate new weights according to

   $$w_t^{(i)} \propto P(D_t \mid Y_t^{(i)}) w_{t-1}^{(i)}$$

   and normalize them to sum to unity.

3.2.2 Sequential Importance Resampling (SIR)

The sequential importance sampling (SIS) algorithm described in the previous section

is a Monte Carlo (MC) method that forms the basis for most advanced particle filters. A

common problem with the SIS algorithm is the degeneracy phenomenon, also known as

collapse of the weights, which describes a situation where all but one particle has

negligible weights after a few iterations. A schematic example of the weight collapse is

depicted in Figure 3.2. To determine collapse of the weights, the effective sample size is

usually introduced as

$$N_{eff} = \frac{1}{\sum_{i=1}^{N} (w_t^{(i)})^2}, \qquad (3.24)$$

which approaches $N_{eff} = 1$ as the weight becomes zero for all but one particle. The collapse of the weights will be eventually resulted in by running the model forward, which causes the tracking performance of the particles to quickly degrade. As a remedy for this, one should resample the particles. However, it is better to resample less often because resampling will often cause a loss of information. A common approach is to resample only when the effective sample size of Equation (3.24) falls below a certain threshold, for example, $N_{eff} < N/10$, where N is the total number of particles.

Figure 3.2: Collapse of the particle weights (van Leeuwen, 2009).

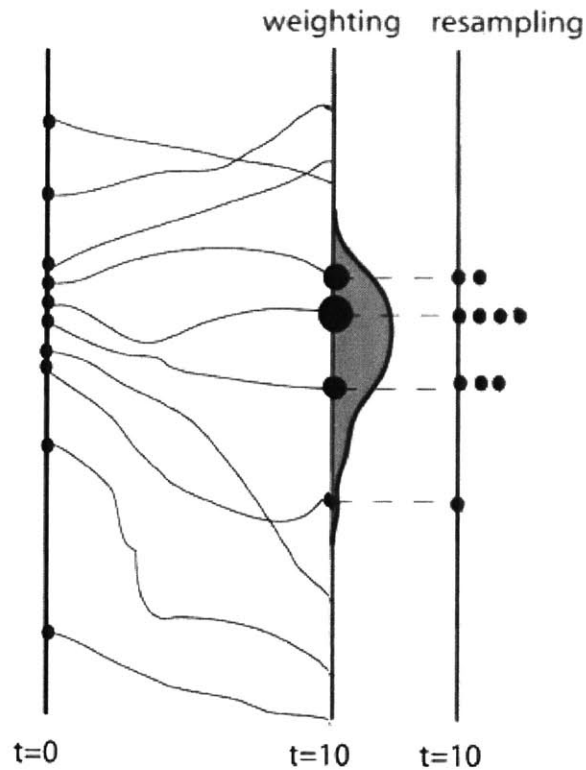weighting resampling

t=0   t=10   t=10

Figure 3.3: Resampled particles according to weights (van Leeuwen, 2009).

The idea of resampling procedure is to abandon particles with very small weights and make multiple copies of particles with large weights as shown in Figure 3.3. Although the theoretical distribution represented by the weighted set of samples does not change, resampling introduces additional variance to estimates. This variance introduced by the resampling procedure can be reduced by proper choice of the resampling method. A comparative analysis and review of resampling approaches can be found in Doucet et al. (2000) and van Leeuwen (2009). In this study, systematic resampling, also known as stochastic universal sampling, is used due to its computational simplicity and good empirical performance. It is shown that the method has the lowest sampling noise

(Kitagawa, 1996; Arulampalam, 2002). In the resampling method all weights are put after each other on a line [0,1]. Then a random number is drawn from a uniform density on $[0, N^{-1}]$. Then N line pieces starting from the random number with interval length N^{-1} are laid on the line [0,1]. A particle is chosen when one of the end points of these line pieces falls in the weight bin of that particle. Clearly, particles with high weights span an interval larger than N^{-1} and will be chosen a number of times. While small weight particles have a negligible chance of being chosen. An algorithm of the method is shown in Algorithm 3.2.

Algorithm 3.2. Resampling Algorithm

---

1. Draw a random number as
$$u_1 \sim U[0, N^{-1}]$$

2. For $i = 2 : N$
   - construct N line pieces: $u_i = u_{i-1} + N^{-1}$

3. Initialize the CDF: $c_0 = 0$

4. For $i = 1 : N$
   - construct CDF: $c_i = c_{i-1} + w_t^{(i)}$

5. For $j = 1 : N$
   a. estimate $n_j = $ the number of $u_k \in (c_{j-1}, c_j]$ for $k = 1, ..., N$
   b. allocate n_j copies of the particle $Y_t^{(j)}$ to the new distribution

6. Assign uniform weights: $w_t^{(i)} = N^{-1}$ for all $i = 1, ..., N$

3.2.3 Regularized Particle Filter (RPF)

A potential problem with sequential importance resampling is that several particles will be identical after the resampling step. Especially, since we assume that the model has

no error, these particles will remain identical during forward integration. Consequently the sample diversity will be lost, and this is known as sample impoverishment. To avoid the issue, some "jitter" has to be applied to the resampled particles as shown in Figure 3.4 (van Leeuwen, 2009). The particle filter with this treatment has been termed regularized particle filter (RPF). The idea of the RPF is to change the discrete approximation of a posterior distribution in SIR to a continuous approximation as shown in Figure 3.5. Kernel density smoothers are usually used to approximate the continuous posterior distribution, and consequently the posterior particles are drawn from the following approximation

$$P(Y_t \mid D_{1:t}) \approx \sum w_t^{(i)} K_h(Y_t - Y_t^{(i)}) , \qquad (3.25)$$

where

$$K_h(Y) = \frac{1}{h^{n_Y}} K\left(\frac{Y}{h}\right) \qquad (3.26)$$

is the rescaled kernel density, $h > 0$ is the bandwidth and n_Y is the dimension of the state vector Y. The kernel K and bandwidth h can be chosen to minimize the mean integrated square error (MISE) between the true posterior density and the corresponding regularized weighted empirical measure. For equal weight particles, it can be shown that the Epanechnikov kernel is the optimal choice in that it minimizes the MISE (Silverman, 1986). The Epanechnikov kernel is given by

$$
K_{opt} = \begin{cases} \dfrac{n_Y + 2}{2c_{n_Y}} (1 - \|Y\|^2) & \text{if } \|Y\| < 1 \\[2ex] 0 & \text{otherwhise} \end{cases} ,
\tag{3.27}
$$

where c_{n_Y} is the volume of the unit sphere in $\mathbb{R}^{n_Y}$

$$
c_{n_Y} = \frac{\pi^{n_Y/2}}{\Gamma(n_Y/2 + 1)},
\tag{3.28}
$$

where Γ = gamma function. The bandwidth should be as large as possible to avoid degeneracy, and as small as possible to avoid extra statistical noise. With unit covariance matrix the optimal bandwidth can be estimated by

$$
h_{opt} = A \cdot N^{-1/(n_Y+4)} \text{ with } A = [8c_{n_Y}^{-1}(n_Y + 4)(2\sqrt{\pi})^{n_Y}]^{1/(n_Y+4)}.
\tag{3.29}
$$

In comparison with SIR, the RPF additionally includes a regularization step as

$$
Y_t^{(i)*} = Y_t^{(i)} + h_{opt} D_t \varepsilon^{(i)},
\tag{3.30}
$$

where $Y_t^{(i)*}$ is a new particle, D_t is estimated from S_t, which is the empirical covariance matrix such that $D_t D_t^T = S_t$. Note that the calculation of the empirical covariance matrix S_t is carried out prior to the resampling and is therefore a function of both the $Y_t^{(i)}$ and

$w_k^{(i)}$. In addition, $\varepsilon^{(i)}$ is the random noise drawn from the kernel. Generating from the Epanechnikov kernel consists of generating $\sqrt{\beta}U$, where β follows a beta distribution with parameters $(n_y/2, 2)$, and U is uniformly distributed over the unit sphere of $\mathbb{R}^{n_y}$ (Musso et al., 2001; Devroye and Györfi, 1985).
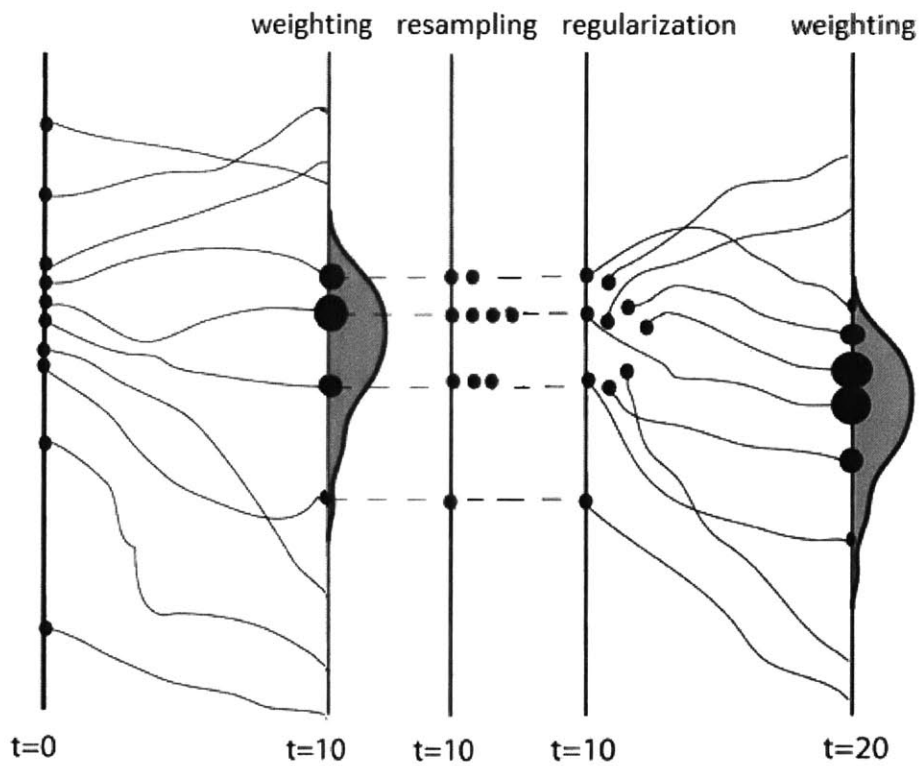

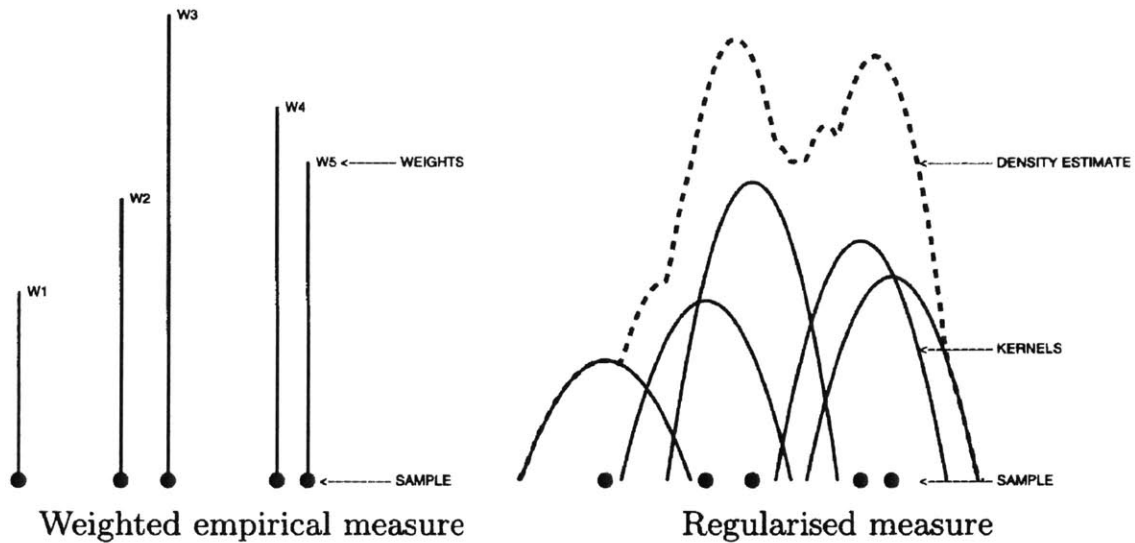**Figure 3.4: Scattered particles according to weighted empirical measure (van Leeuwen, 2009).**

Figure 3.5: Regularization of an empirical measure (Musso et al., 2001).

3.2.4 Order reduction for Geological Representation

For a large scale system such as oil reservoir simulation models, particle filtering has rarely been considered since it is difficult to locate the initial particles on highly probable region on the space of model parameters (Aanonsen et al., 2009). A remedy for the issue of system scale is to reparameterize the model parameters using projection methods as used in Chapter 2 in order to reduce the dimension of dynamic model states. For example, Jafarpour and McLaughlin (2008, 2009) suggested discrete cosine transform (DCT) to reparameterize model parameters in reduced-order space for the ensemble Kalman filter. An advantage of DCT is that it uses fixed bases vectors derived from Fourier transformation, but it cannot outperform data-driven methods such as Karhunen–Loève transform (KLT) which is identical to proper orthogonal decomposition (POD) introduced in Chapter 2. In this study, POD is used to reparameterize the model parameters in a way that the state vector Y is approximated as

$$Y_t = \mathbf{\Phi}_Y \, \upsilon_t \, , \tag{3.31}$$

where $\mathbf{\Phi}_Y \in \mathbb{R}^{n_Y \times n_\upsilon}$ is the basis matrix computed by applying POD, described in Algorithm 1, to the matrix which contains the state vectors of all the particles. Then, the regularization step is carried out in terms of υ_t. That is, the above optimal bandwidth is replaced by

$$h_{opt} = A \cdot N^{-1/(n_\upsilon+4)} \text{ with } A = [8c_{n_\upsilon}^{-1}(n_\upsilon+4)(2\sqrt{\pi})^{n_\upsilon}]^{1/(n_\upsilon+4)}, \tag{3.32}$$

and the regularization step draws new particles from kernel density as

$$\upsilon_t^{(i)*} = \upsilon_t^{(i)} + h_{opt} \mathbf{D}_t \varepsilon^{(i)} \, , \tag{3.33}$$

where $\upsilon_t^{(i)*}$ is a new reduced-order particle, $\mathbf{D}_t$ is estimated from $\mathbf{S}_t$, which is the empirical covariance matrix of υ_t such that $\mathbf{D}_t \mathbf{D}_t^T = \mathbf{S}_t$.

3.2.5 Hyper-reduced-order Models for RPF

Contrary to embedding the GNAT-based reduced-order model into the framework of EnKF, the embedment of the reduced-order model into the RPF can be carried out without the clustering procedure like k-means algorithm. Based on the fact that the regularization step is conducted for the identical particles resulted from resampling, we

can just select a particle in the group of identical particles to be evolved by full-order simulation, and then the simulation run is used to train a GNAT-based reduced-order model that will be used for evolving the rest of particles within the group, which are regularized by Equation (3.33). Figure 3.6 shows the selected particles for training in the framework of regularized particle filtering.
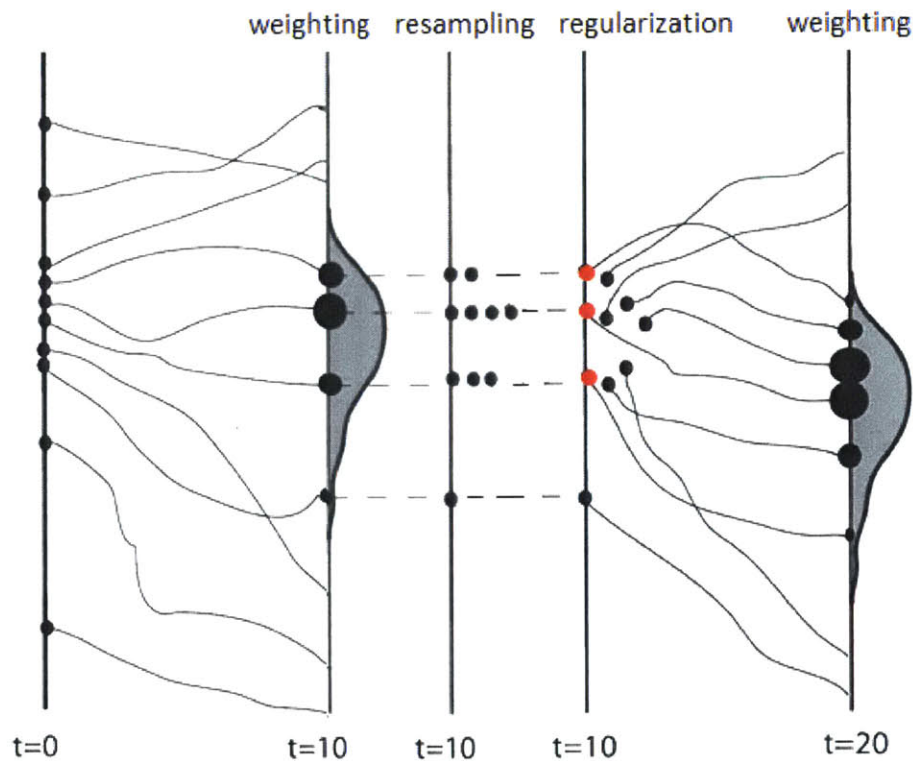

Figure 3.6: Scattered particles and selected particles (in red color) for training.

3.3 Case Study

In this section, the two frameworks of ensemble-based history matching combined with hyper-reduced-order models are tested by applying them to two-dimensional reservoir fields. In the following subsection, the EnKF combined with the hyper-reduced-

order model is applied to a 2D Gaussian permeability field, followed by a subsection in which the RPF combined with the HRM is applied to a 2D non-Gaussian field. The pdf of the non-Gaussian filed is bimodal.

3.3.1 EnKF-based History Matching using Hyper-reduced-order Models

In order to evaluate the performance EnKF combined with the HRM, a 2D Gaussian geologic model is considered in a history matching problem. Its permeability filed is the unknown parameter that will be sequentially tracked by assimilating the well responses. The log-permeability in x-direction and its histogram are shown in Figure 3.7.

(a) Log-permeability field   (b) Histogram

Figure 3.7: Log-permeability in the x-direction of the 2D Gaussian reservoir model with eight injectors (triangles) and eight producers (circles).

The log-permeability field is assumed Gaussian process with mean $E[\log k] = 3.9$, which is corresponding to $k = 50$ in millidarcy (md). The variance of the log-permeability is one. The realization is defined on 45×45 grid cells of size $30\,ft \times 30\,ft \times$

$30 ft$. There are eight injection wells and eight productions wells, assuming that all of the wells are under bore hole pressure control (BHP). In addition, the permeability values in the cells where the wells are located are assumed known and the mean value $k = 50$ is assigned on the cells. The spatial correlation structure is modeled by a spherical variogram with range of 330 ft. The permeability field is generated through sequential Gaussian simulation within SGEMS (Remy et al., 2008).

We specify the injection wells to inject water at a constant BHP of 5300 psi, while the production wells produce at a constant BHP of 4800 psi. The well locations and a realization of permeability field in the x-direction are shown in Figure 3.10. Permeability is taken to be a diagonal tensor, with $k_x = k_y$. The initial oil and water saturations are 0.85 and 0.15, respectively and the residual oil (s_{or}) and water (s_{wr}) saturations are 0.15. For oil, we set $\rho_o = 55 \ lbm/ft^3$, $\mu_o = 5 \ cp$; for water, $\rho_w = 65 \ lbm/ft^3$, $\mu_w = 1 \ cp$. The system is assumed incompressible, and capillary effects are neglected. The relative permeability relationships are specified as

$$k_{ro}(s_w) = k_{ro}^0 \left(\frac{1 - s_w - s_{or}}{1 - s_{wr} - s_{or}} \right)^a, \tag{3.34}$$

$$k_{rw}(s_w) = k_{rw}^0 \left(\frac{s_w - s_{wr}}{1 - s_{wr} - s_{or}} \right)^b, \tag{3.35}$$

where k_{ro}^0 and k_{rw}^0 are the endpoint relative permeabilities. Here, we set $k_{ro}^0 = k_{rw}^0 = 1$ and $a = b = 2$.

The simulations are run for 2000 days. The first half of the simulation period is considered as assimilation period, and the rest of the period is for prediction. The assimilation is carried out at every 100 days, so there are 10 times of assimilation step. The well responses are assumed available only at every 100 days, and the response consists of water injection rates at injection wells, oil/water production rates at production wells. Therefore, 24 dimensional data observation vector is assimilated in each assimilation step at every 100 days. The ensemble members of $N_e = 200$ are generated by sequential Gaussian simulation using the geostatistical information same as the base model shown in Figure 3.7. Note that the base realization is not a member of the ensemble realizations. Two cases are considered. One is to carry out the data assimilation only using full-order simulation runs. The other one is to run only 25% of the ensemble members by full-order simulation, while the rest of ensembles by reduced-order simulation. This means that 50 near-centroid ensemble members are selected through k-means clustering and are run by full-order simulation for training. Then, the 50 trained hyper-reduced order models are used to evolve the rest of ensembles members (150) in reduced-order. Using the initial ensemble realizations and the finally assimilated permeability fields, the water and oil production rates for the whole period of 2000 days are simulated and shown in Figure 3.8–3.15.

In general, the production rate mismatches from initial ensemble permeability realizations are improved significantly for both of 100% FOM simulation runs and 25% ROM + 75% ROM simulation runs. In addition, the span of predicted production rates, which can be viewed as prediction uncertainty is also reduced encompassing the 'true' well productions.

The computational cost required for a single full-order simulation run is almost 106 times of a single hyper-reduced-order simulation run, which means that the 75% of computational cost required for the full-order history matching is saved.
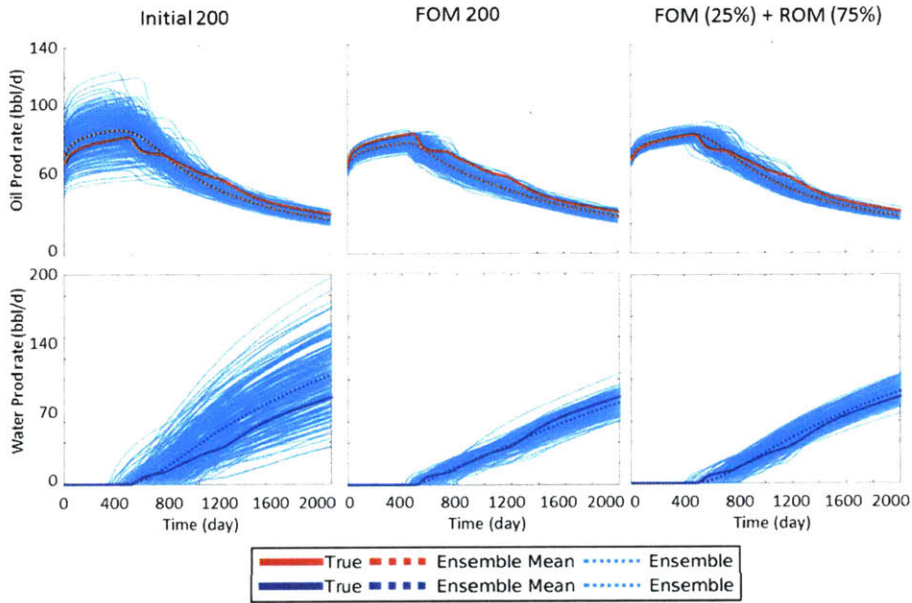

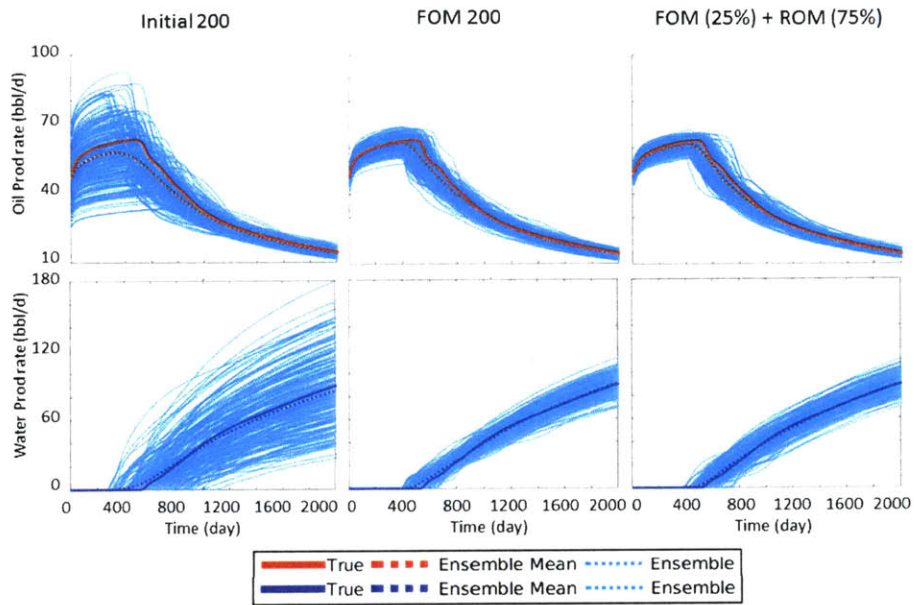**Figure 3.8: Oil/Water production rates at production well P1.**

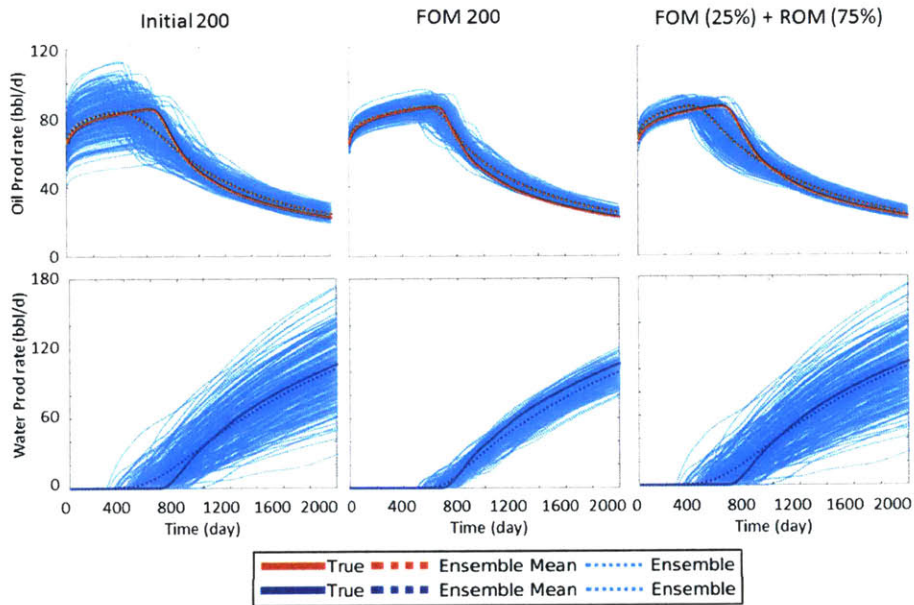**Figure 3.9: Oil/Water production rates at production well P2**

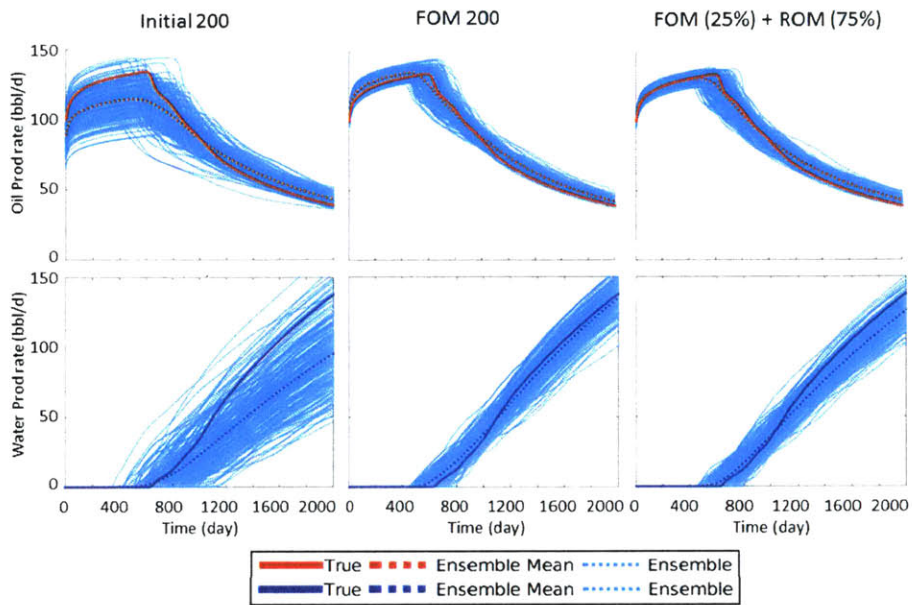**Figure 3.10: Oil/Water production rates at production well P3**

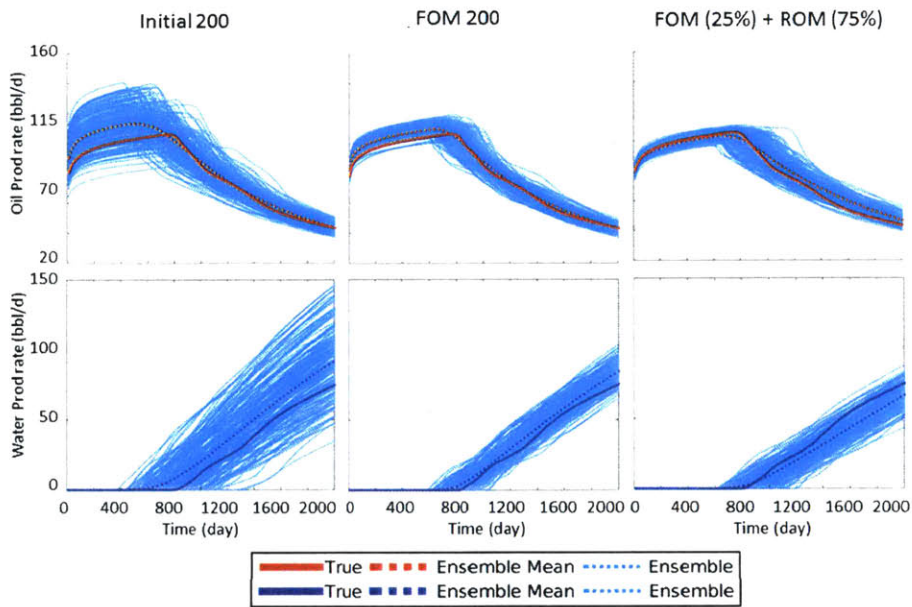**Figure 3.11: Oil/Water production rates at production well P4**
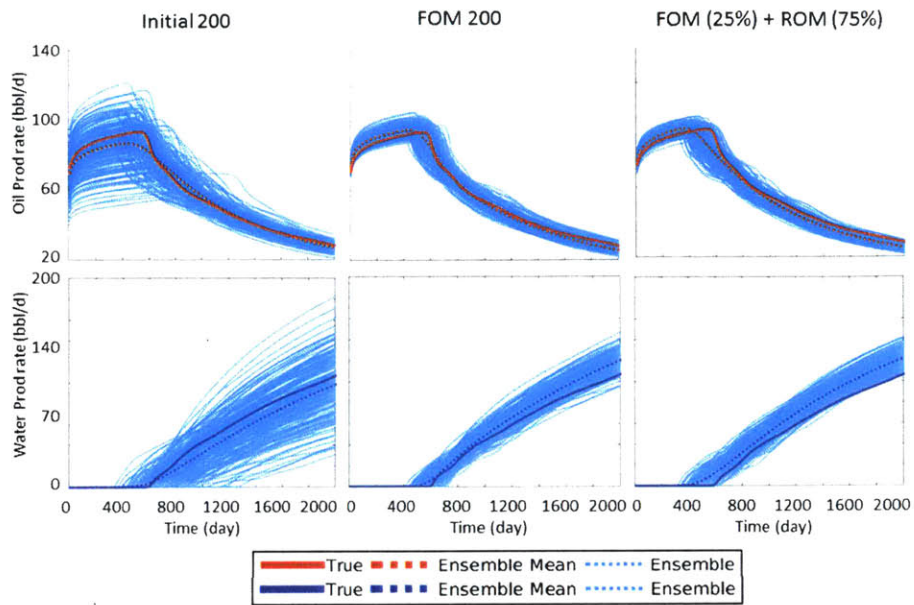

Figure 3.12: Oil/Water production rates at production well P5

Figure 3.13: Oil/Water production rates at production well P6
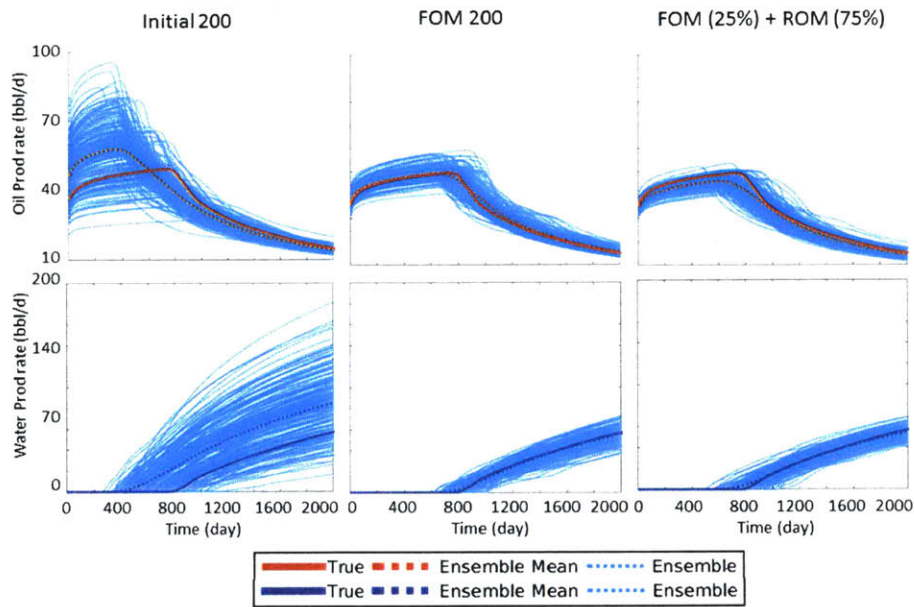

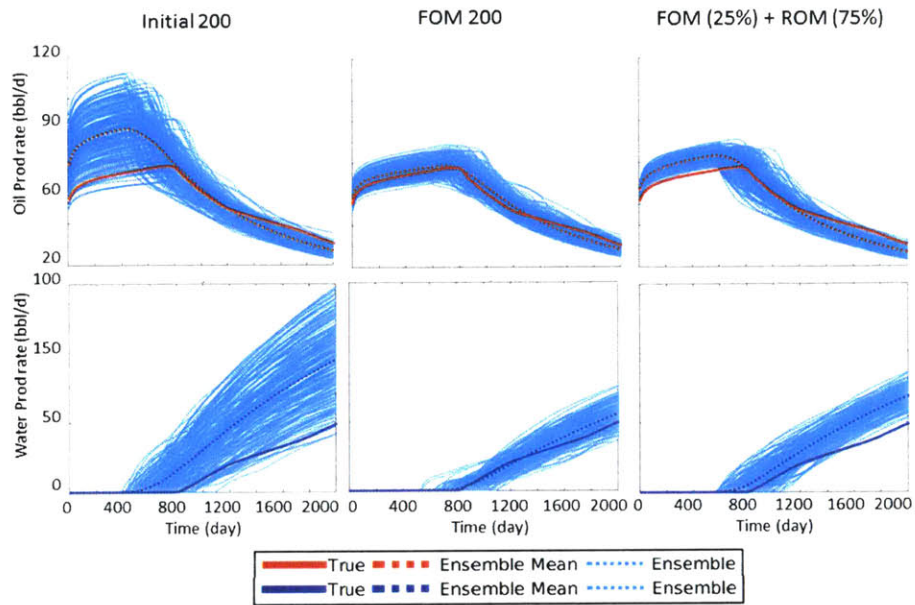**Figure 3.14: Oil/Water production rates at production well P7**

Figure 3.15: Oil/Water production rates at production well P8

The mean of history matched ensembles of permeability fields are shown in Figure 3.16. Again, the averaged permeability ensembles show general agreements with the base permeability model although it is not included in the initial ensemble members.
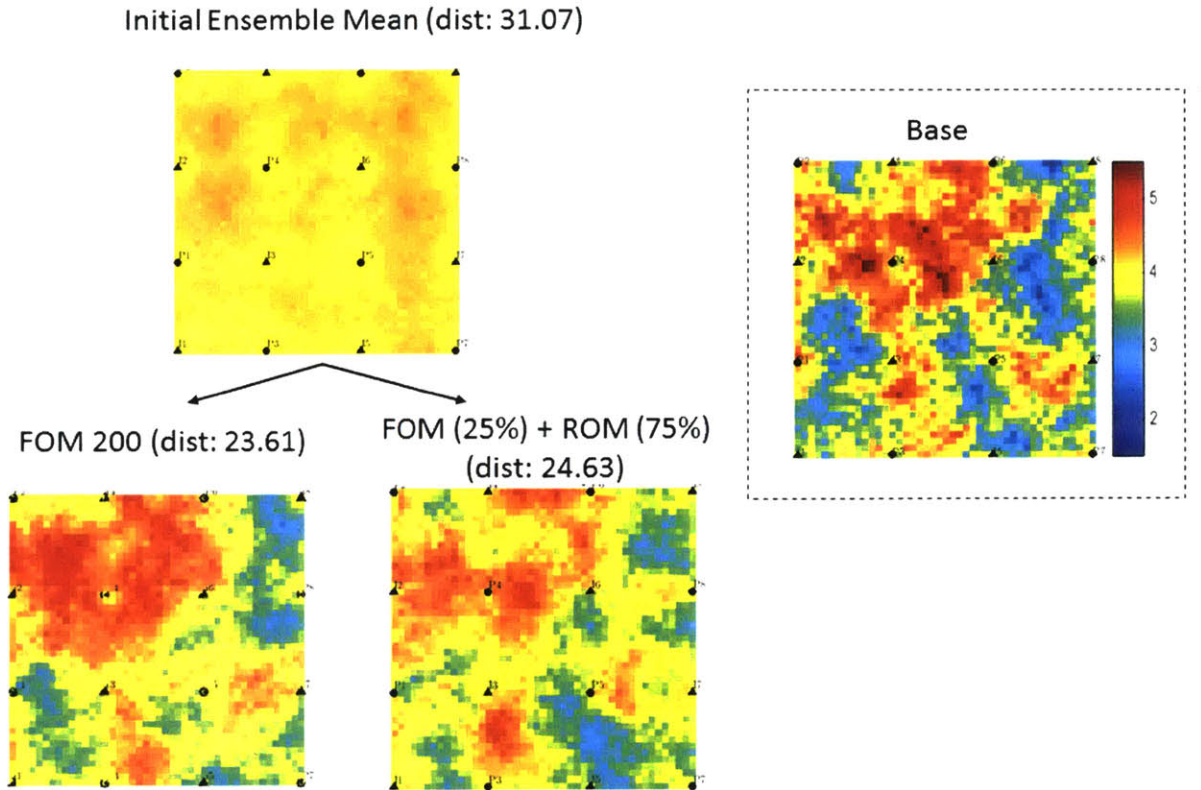
Initial Ensemble Mean (dist: 31.07)

Base

FOM 200 (dist: 23.61)

FOM (25%) + ROM (75%)
(dist: 24.63)

Figure 3.16: History matched permeability fields.

3.3.2 RPF-based History Matching using Hyper-reduced-order Models

The advantage of using PF is based on the fact that the method does not take any assumption of probability distribution on the static and dynamic parameters, while EnKF is suitable only when Gaussian assumption is appropriate. However, it has been shown that PF requires a large number of particles (ensembles) as system scale increases as in the case of reservoir history matching. The disadvantage induces much computational burden, and consequently PF has rarely been applied for reservoir history matching. Therefore, this study introduces two remedies: HRM for computational cost and reparameterization of geologic parameters to reduce the required number of particle.s.

In addition, to investigate the performance of RPF combined with HRM for non-Gaussian geologic model, the 2D Gaussian permeability field used in the previous section are converted to follow a bimodal distribution as shown in Figure 3.17.
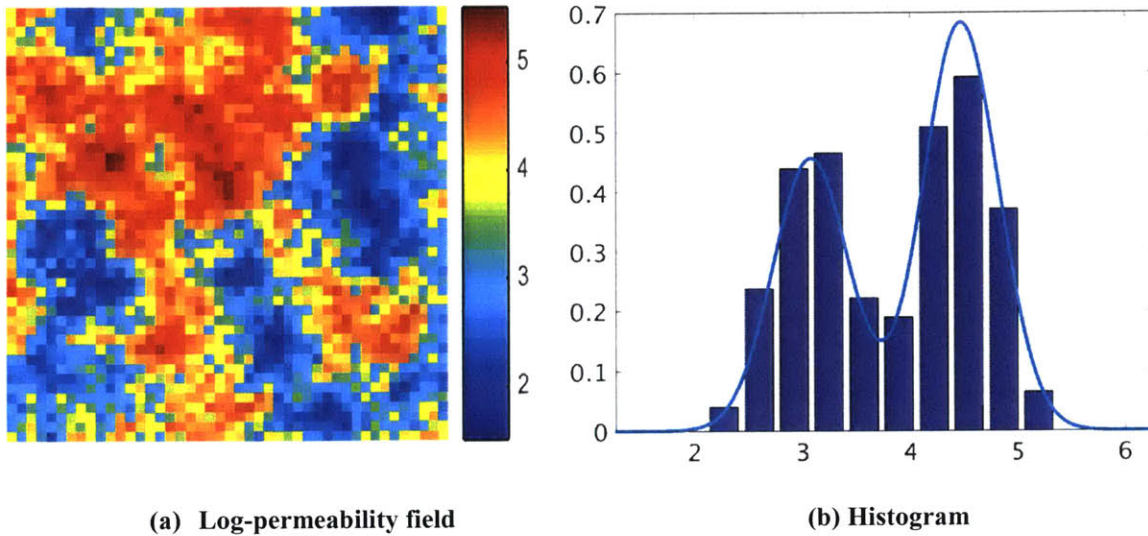

(a) **Log-permeability field**   (b) **Histogram**

Figure 3.17: Log-permeability in the x-direction of the 2D Gaussian reservoir model with eight injectors (triangles) and eight producers (circles).

To construct 1000 particles of permeability fields, log-permeability fields of Gaussian process are first generated using the same geostatistical conditions as the EnKF case study in the previous section. Then, the geologic realizations are converted to follow a bimodal distribution, which consists of two Gaussian distributions: first, $E[\log k] = 3.2$ and $Var[\log k] = 0.35$ with probability of 0.4; second, $E[\log k] = 4.6$ and $Var[\log k] = 0.35$ with probability of 0.6. There are again eight injection wells and productions wells, assuming that all of the wells are under BHP control.

Well locations and controls of BHPs are scheduled the same as the EnKF case study in the previous section. Permeability is taken to be a diagonal tensor, with $k_x = k_y$. The

initial oil and water saturations are 0.85 and 0.15, respectively and the residual oil (s_{or})

and water (s_{wr}) saturations are 0.15. For oil, we set ρ_o = 55 lbm/ft^3, μ_o = 5 cp; for

water, ρ_w = 65 lbm/ft^3, μ_w = 1 cp. The system is assumed incompressible, and

capillary effects are neglected.

The simulations are run for 2000 days as in the case study of EnKF. The specification

of time schedule of assimilation and prediction, and the types of well observation and its

availability are set the same as the EnKF case study. Again, note that the base realization

is not a member of the 1000 particles. Using the initial particle realizations and the finally

assimilated permeability fields, the water and oil production rates for the whole period of

2000 days are simulated and shown in Figure 3.18–3.25.

As discussed in Section 3.2.5, RPF does not require to predesign the ratio of full-

order runs and reduced-order runs because resampling steps in RPF will specify the close

particles. In this case study, 17% of particles in average are evolved by full-order runs,

while 83% of particles are evolved by reduced-order runs. The computational cost

required for a single full-order simulation run is almost 113 times of a single hyper-

reduced-order simulation run, which means that the 83% of computational cost required

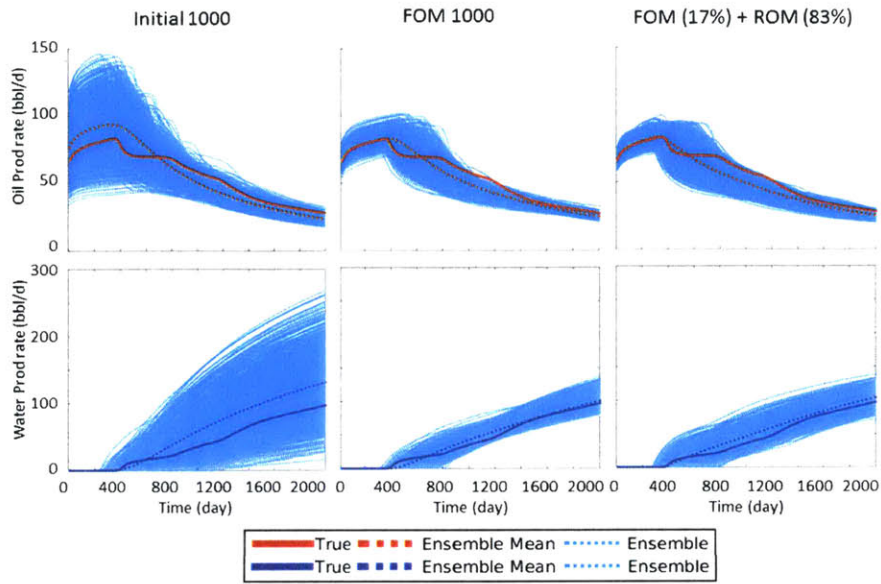for the full-order history matching is saved.

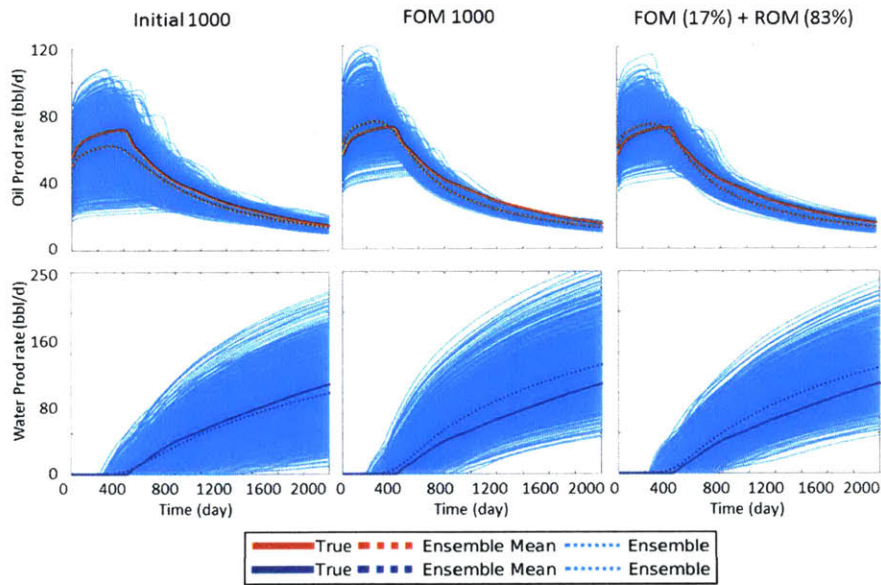**Figure 3.18: Oil/Water production rates at production well P1**

Figure 3.19: Oil/Water production rates at production well P2

Figure 3.20: Oil/Water production rates at production well P3

Figure 3.21: Oil/Water production rates at production well P4

Figure 3.22: Oil/Water production rates at production well P5
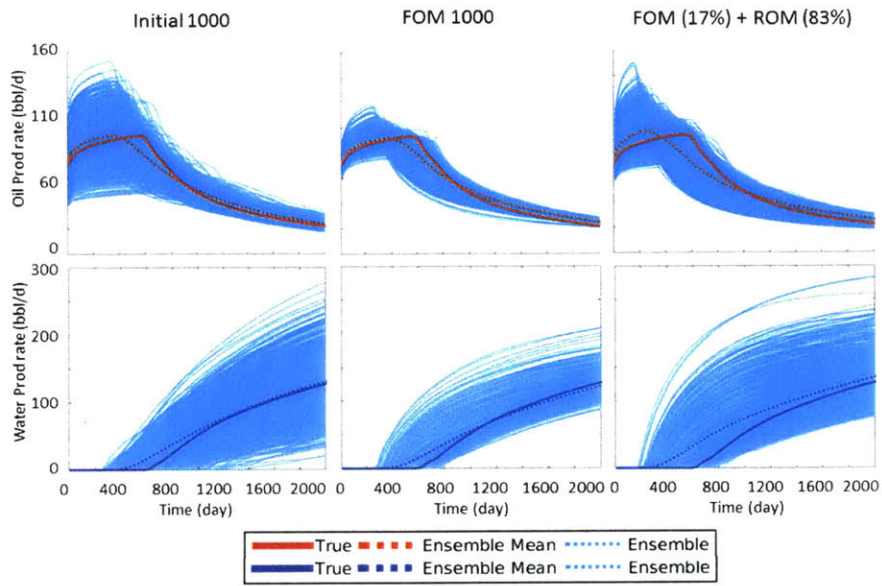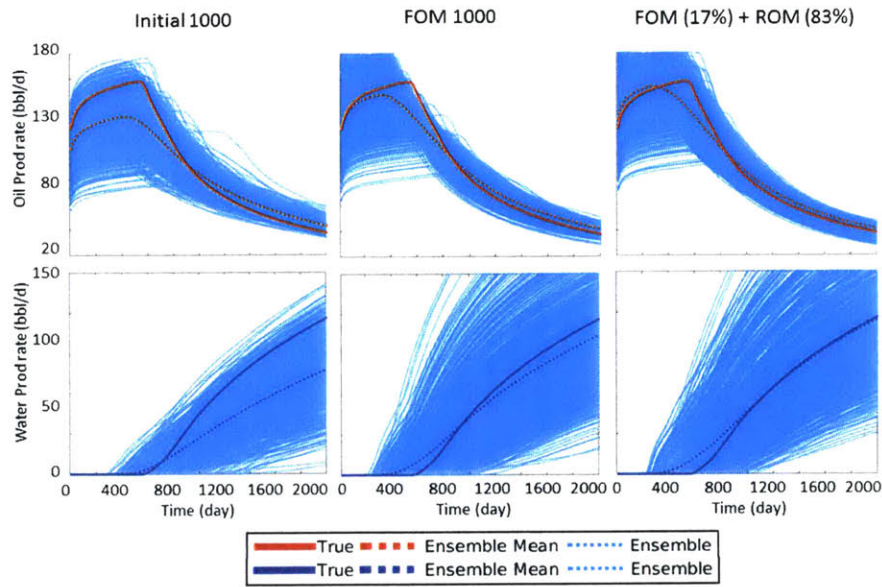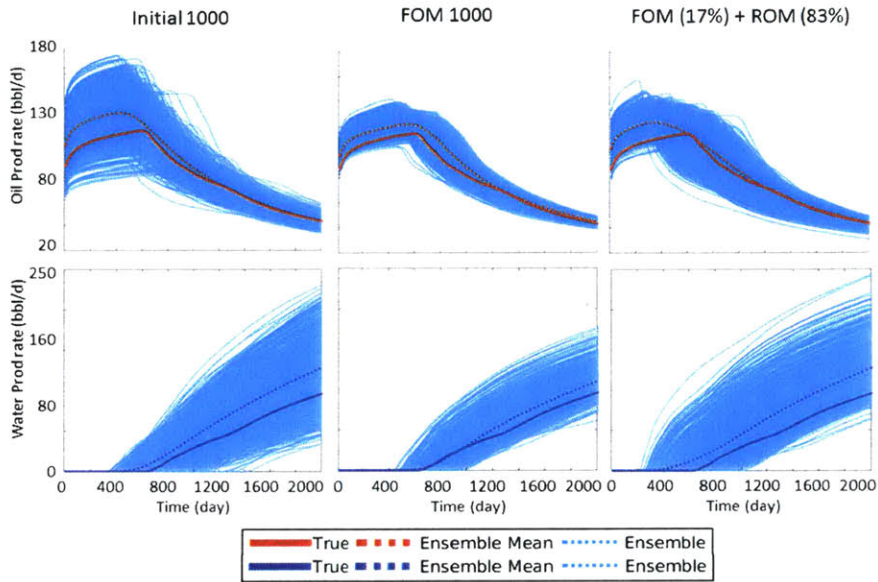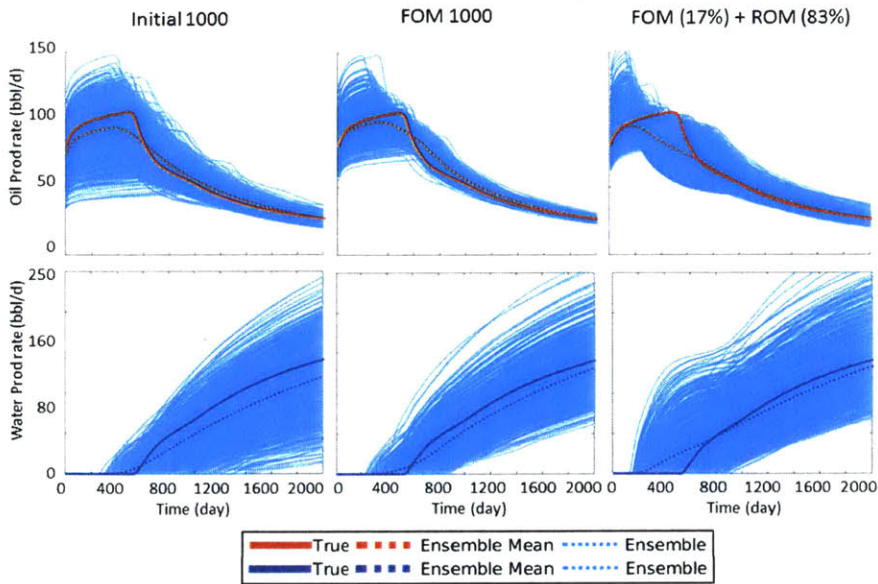

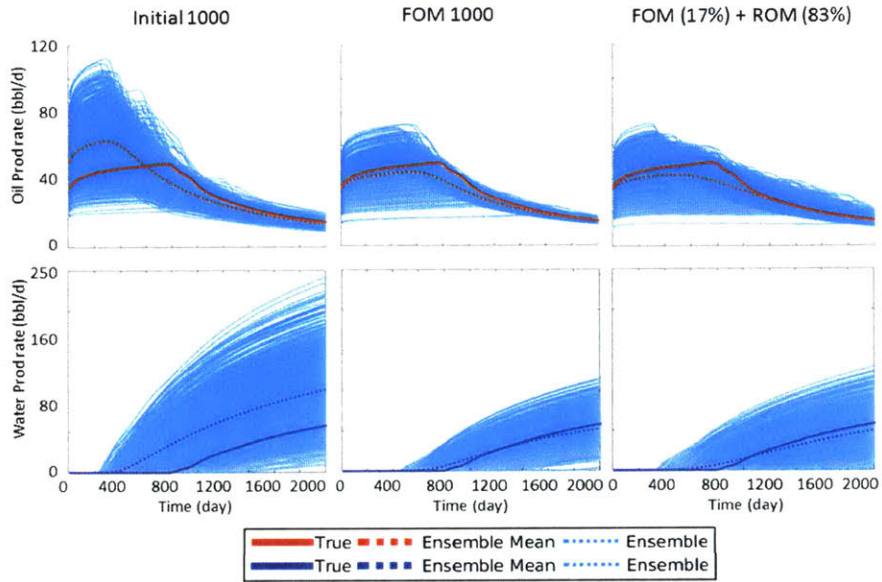**Figure 3.23: Oil/Water production rates at production well P6**

Figure 3.24: Oil/Water production rates at production well P7

Figure 3.25: Oil/Water production rates at production well P8

In general, the production rate mismatches from initial ensemble permeability realizations are improved significantly for both of 100% FOM simulation runs and 17% ROM + 83% ROM simulation runs. Note that the percentage is not predesigned, and it is an outcome of the resampling steps, in which identical particles generated according to the empirical probability measure as described in Algorithm 3.2. In addition, the span of predicted production rates, which can be viewed as prediction uncertainty is also reduced encompassing the 'true' well productions.

The mean of history matched ensembles of permeability fields are shown in Figure 3.26. Again, the averaged permeability of the finally updated particles shows a general agreement with the base permeability although it is not included in the initial particles.

Figure 3.26: History matched permeability fields.

3.4 Summary

In this chapter, two types of sequential Bayesian filtering are considered in view of using hyper-reduced-order models. The purpose of usi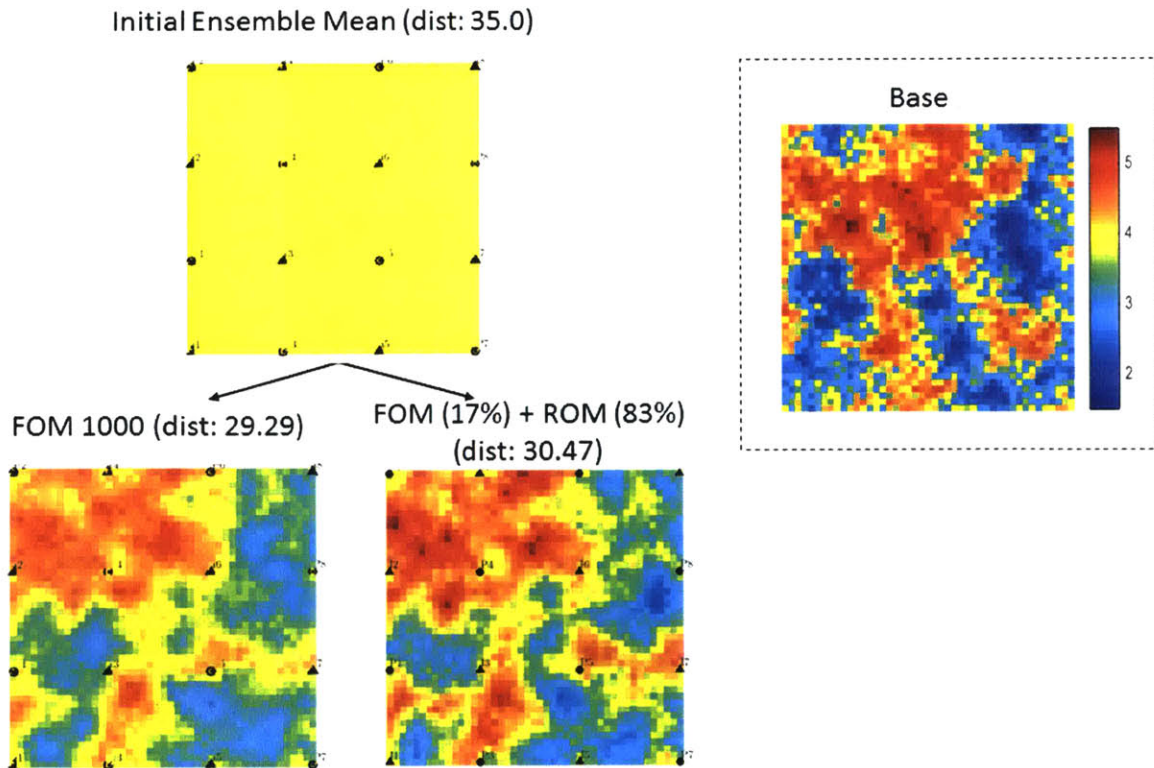ng the HRM is to expedite the computation required for the sequential procedure of data assimilation using ensemble (particle) approaches. For Gaussian system, EnKF has been the most popular method for reservoir history matching, and it is demonstrated through this study that combining the EnKF and the HRM can significantly reduce the computational cost without much loss of accuracy in comparison with full-order model-based history matching. However, the combination requires a few remedies such as clustering to find an optimum reduced-order models according to spatial similarity of geologic parameters, which causes an additional computation. For non-Gaussian system, EnKF is not guaranteed to be optimum because it is derived from Gaussian assumptions. Therefore, this study investigated the use of particle filtering. However, PF has rarely been applied for reservoir history matching due to the fact that it is hard to locate the initial particles on highly probable regions of state spaces, which makes the required number of particles exponentially scale with the model dimension. To resolve the issues, the geological parameters are reparameterized by principal component analysis (PCA), which is another name of POD method so that the required number of initial particles can be reduced down to a manageable level. An advanced particle filter known as regularized particle filter is considered, and a huge computational saving is achieved by the embedment of HRM. Furthermore, the additional cost of clustering required in the framework of combining EnKF and HRM is saved because the advanced particle filter contains resampling steps which do not require the procedure of finding the similar particles.

Chapter 4

Conclusions and Future Work

4.1 Summary and Conclusions

The main contributions of this work consists of two parts. First, GNAT-based hyper-reduced-order models for multiphase subsurface flow simulation are developed to relieve the computational burden in reservoir flow modeling. Second, frameworks of combining the developed HRM and sequential Bayesian Filters are suggested. For the development of HRMs, three components of order reduction are presented: state reduction, constraint reduction, and nonlinearity treatment. State reduction based on proper orthogonal decomposition (POD) is considered, and the impact of state reduction, with different strategies for collecting snapshots, on accuracy and predictability is investigated. Petrov–Galerkin projection is used for constraint reduction, and a hyper-reduction that couples the Petrov–Galerkin projection and a 'gappy' reconstruction is applied for the

nonlinearity treatment. The hyper-reduction method is a Gauss–Newton framework with approximated tensors (GNAT), and the main novelty of this study is the procedure for applying the method to subsurface flow simulation. A fully implicit oil–water two-phase subsurface flow model in three-dimensional space is considered, and the application of the proposed hyper-reduced-order modeling procedure achieves a runtime speedup of more than 300 relative to the full-order method, which cannot be achieved when only constraint reduction is adopted.

For history matching with the use of HRMs, two types of sequential Bayesian filtering are considered. First, EnKF is considered for Gaussian system and a procedure of embedding the HRM into the EnKF is presented. The use of the HRM for the EnKF can significantly reduce the computational cost without much loss of accuracy, but the combination requires a few remedies such as clustering to find an optimum reduced-order models according to geological condition, which causes an additional computation. For non-Gaussian system, an advanced particle filter, known as regularized particle filter, is considered because it does not take any distributional assumptions. PF has rarely been applied for reservoir history matching due to the fact that it is hard to locate the initial particles on highly probable regions of state spaces, which makes the required number of particles exponentially scale with the model dimension. To resolve the issues, the geological parameters are reparameterized by principal component analysis (PCA), which is another name of POD method so that the required number of initial particles can be reduced down to a manageable level. A huge computational saving is achieved by embedding the HRM into the framework of PF. Furthermore, the additional cost of

clustering required in the framework of combining EnKF and HRM is saved because the advanced particle filter does not require the procedure of finding the close particles.

4.2 Future Work

Future work should cover the further development of the proposed hyper-reduced-order modeling procedure for more general subsurface flow simulations such as compressible flow with capillary effects, idealized thermal simulation cases, and compositional systems. The performance of GNAT in nonlinearity treatment also needs to be compared to previously proposed methods such as TPWL and DEIM. In addition, the application of the hyper-reduced-order model to production optimization and history matching should also be conducted in future work.

For the use of the reduced-order model in history matching, the predictability of the reduced-order model under the change of model parameter such as permeability and porosity should be investigated to evaluate the tolerable range of the change in view of using reduced-order models. The basic EnKF considered in this study is using perturbed observations, which can cause additional variability on ensemble-based statistics. To resolve the issue, there has been researches using ensemble square root filters (Sakov and Oke, 2008a; 2008b), and this will be of interest for future work. For particle filtering with the use of HRM, the performance of history matching is demonstrated just for 45 by 45 dimensional model. Compared to the scales common in practice, the result is limited since the PF with HRM is not investigated for large scale systems. This can be important because PF is sensitive to the scale of the system, and therefore the performance of the

proposed framework of combining PF and HRM should be tested for real scale reservoir system.

Finally, the use of the HRM for reservoir design optimization should also be considered, and the optimization procedure can be combined with history matching in the framework of closed-loop reservoir management as shown in Figure 4.1.
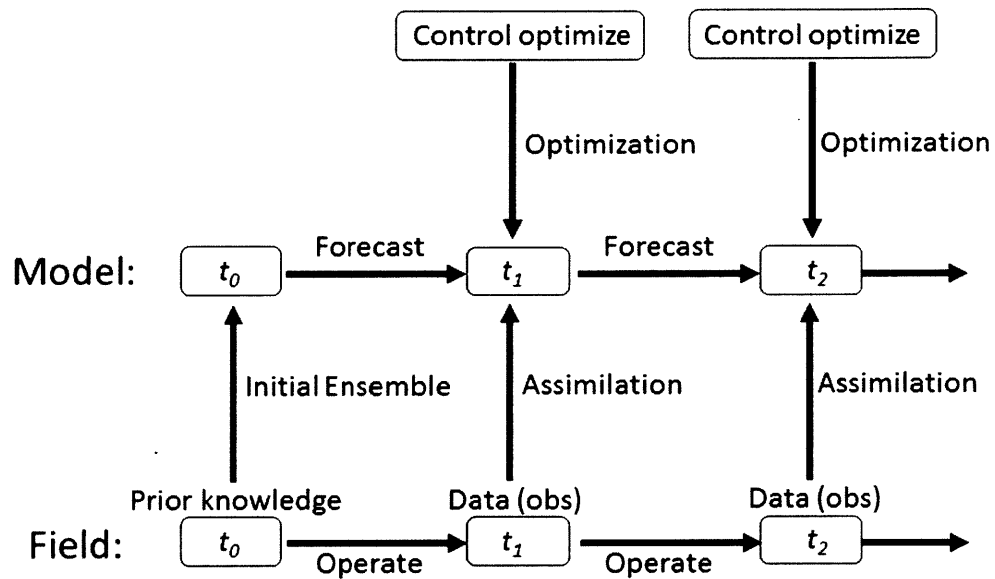
Figure 4.1: Sequential procedure for closed-loop reservoir management.

Bibliography

Aanonsen, S. I., Nævdal, G., Oliver, D. S., Reynolds, A. C., & Vallès, B. 2009. The Ensemble Kalman

   Filter in Reservoir Engineering–a Review. *SPE Journal*, 14(03), 393-412.

Alghareeb, Z. M., and Williams, J. 2013. Optimum Decision-Making in Reservoir Management Using

   Reduced-order Models. In *SPE Annual Technical Conference and Exhibition*. Society of

   Petroleum Engineers.

Amsallem, D., Zahr, M. J., and Farhat, C. 2012. Nonlinear model order reduction based on local reduced-

   order bases. *International journal for numerical methods in engineering*, 92(10), 891-916.

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. 2002. A tutorial on particle filters for online

   nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on signal processing*, 50(2), 174-

   188.

Aziz, K., and Settari, A. 1979. *Petroleum reservoir simulation*. Chapman & Hall.

Brouwer, D. R., Nævdal, G., Jansen, J. D., Vefring, E. H., & Van Kruijsdijk, C. P. J. W. 2004. Improved

   reservoir management through optimal control and continuous model updating. In *SPE Annual

   Technical Conference and Exhibition*. Society of Petroleum Engineers.

Cane, M.A., Kaplan, A., Miller, R.N., Tang, B., Hackert, E.C. and Busalacchi, A.J., 1996. Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter. *Journal of Geophysical Research*, 101, 599-617.

Cardoso, M. A., Durlofsky, L. J., and Sarma, P. 2009. Development and application of reduced-order modeling procedures for subsurface flow simulation. *International journal for numerical methods in engineering*, 77(9), 1322-1350.

Cardoso, M. A. and Durlofsky, L. J. 2010. Linearized reduced-order models for subsurface flow simulation. *Journal of Computational Physics*, *229*(3), 681-700.

Carlberg, K., Bou-Mosleh, C., and Farhat, C. 2011. Efficient non-linear model reduction via a least-squares Petrov—Galerkin projection and compressive tensor approximations. *International Journal for Numerical Methods in Engineering*, 86(2), 155-181.

Carlberg, K., Farhat, C., Cortial, J., and Amsallem, D. 2013. The GNAT method for nonlinear model reduction: effective implementation and application to computational fluid dynamics and turbulent flows. *Journal of Computational Physics*, 242, 623-647.

Chaturantabut, S. and Sorensen, D. C. 2010. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, *32*(5), 2737-2764.

Chen, Y., Oliver, D. S., & Zhang, D. 2009. Data assimilation for nonlinear problems by ensemble Kalman filter with reparameterization. *Journal of Petroleum Science and Engineering*, 66(1), 1-14.

Chen, Y., & Oliver, D. S. 2010. Ensemble-based closed-loop optimization applied to Brugge field. *SPE Reservoir Evaluation and Engineering*, 13(01), 56-71.

Christie, M. A., and Blunt, M. J. 2001. Tenth SPE comparative solution project: A comparison of upscaling techniques. *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers.

Devroye, Luc, and Laszlo Gyorfi. *Nonparametric density estimation: the L1 view.* Vol. 119. John Wiley & Sons Incorporated, 1985.

Doucet, A., Godsill, S. and Andrieu, C., 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and computing*, 10(3), pp.197-208.

Everson, R., and Sirovich, L. 1995. Karhunen–Loeve procedure for gappy data. Journal of the Optical Society of America A, 12(8), 1657-1664.

Exxon Mobil Corporation (2014), The Outlook for Energy: A View to 2040, *Technical report from www.exxonmobil.com/energyoutlook*.

Gerritsen, M. G., and Durlofsky, L. J. 2005. Modeling fluid flow in oil reservoirs. *Annu. Rev. Fluid Mech.*, 37, 211-238.

Gildin, E., Ghasemi, M., Romanovskay, A., and Efendiev, Y. 2013. Nonlinear complexity reduction for fast simulation of flow in heterogeneous porous media. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers.

Ghommem, M., Gildin, E., and Ghasemi, M. (2015). Complexity Reduction of Multiphase Flows in Heterogeneous Porous Media. *SPE Journal*.

Gu, Y. and Oliver, D.S., 2006. The ensemble Kalman filter for continuous updating of reservoir simulation models. *Journal of Energy Resources Technology*, 128(1), pp.79-87.

Gu, Y. and Oliver, D.S., 2007. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE Journal*, 12(04), pp.438-446.

He, J., Sætrom, J. and Durlofsky, L. J. 2011. Enhanced linearized reduced-order models for subsurface flow simulation. *Journal of Computational Physics*, 230(23), 8313-8341.

He, J., Sarma, P. and Durlofsky, L. J. 2013. Reduced-order flow modeling and geological parameterization for ensemble-based data assimilation. *Computers & Geosciences*, 55, 54-69.

He, J., and Durlofsky, L. J. 2014. Reduced-order modeling for compositional simulation by use of trajectory piecewise linearization. *SPE Journal*, 19(05), 858-872.

He, J., and Durlofsky, L. J. 2015. Constraint reduction procedures for reduced-order subsurface flow models based on POD–TPWL. *International Journal for Numerical Methods in Engineering*, 103(1), 1-30.

Heijn, T., Markovinović, R., and Jansen, J. D. 2004. Generation of low-order reservoir models using system-theoretical concepts. *SPE Journal*, 9(2), 202-218.

Hoteit, I. and Pham, D.T., 2004. An adaptively reduced-order extended Kalman filter for data assimilation in the tropical Pacific. *Journal of marine systems*, 45(3), pp.173-188.

Jafarpour, B., & McLaughlin, D. B. 2008. History matching with an ensemble Kalman filter and discrete cosine parameterization. *Computational Geosciences*, 12(2), 227-244.

Jafarpour, B., & McLaughlin, D. B. 2009. Estimating channelized-reservoir permeabilities with the ensemble Kalman filter: The importance of ensemble design. *SPE Journal*, 14(02), 374-388.

Jansen, J. D., Brouwer, D. R., Nævdal, G., & Van Kruijsdijk, C. P. J. W. 2005. Closed-loop reservoir management. *First Break*, 23(1).

Jansen, J. D. 2011. Adjoint-based optimization of multi-phase flow through porous media–a review. *Computers and Fluids*, 46(1), 40-51.

Kitagawa, G., 1996. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of computational and graphical statistics*, 5(1), pp.1-25.

Markovinović, R. and Jansen, J. D. 2006. Accelerating iterative solution methods using reduced-order models as solution predictors. *International journal for numerical methods in engineering*, 68(5), 525-541.

Musso, C., Oudjane, N. and Le Gland, F., 2001. Improving regularised particle filters. *In Sequential Monte Carlo methods in practice* (pp. 247-271). Springer New York.

Nævdal, G., Mannseth, T. and Vefring, E.H., 2002, April. Near-well reservoir monitoring through ensemble Kalman filter. *In SPE/DOE Improved Oil Recovery Symposium* (No. 1).

Nævdal, G., Johnsen, L. M., Aanonsen, S. I., and Vefring, E. H. 2005. Reservoir monitoring and continuous model updating using ensemble Kalman filter. *SPE Journal*, 10(01), 66-74.

Peters, L., Arts, R., Brouwer, G., Geel, C., Cullick, S., Lorentzen, R.J., Chen, Y., Dunlop, N., Vossepoel, F.C., Xu, R. and Sarma, P., 2010. Results of the Brugge benchmark study for flooding optimization and history matching. *SPE Reservoir Evaluation & Engineering*, 13(03), pp.391-405.

Peherstorfer, B., Butnaru, D., Willcox, K., and Bungartz, H. J. 2014. Localized discrete empirical interpolation method. *SIAM Journal on Scientific Computing, 36*(1), 168-192.

Pham, D.T., 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly weather review*, 129(5), pp.1194-1207.

Remy, N., Boucher, A. and Wu, J., 2009. *Applied geostatistics with SGeMS: A user's guide*. Cambridge University Press.

Rewienski, M. and White, J. 2003. A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and System, 22*(2), 155-170.

Rousset, M. A., Huang, C. K., Klie, H. and Durlofsky, L. J., 2014. Reduced-order modeling for thermal recovery processes. *Computational Geosciences*, 18(3-4), 401-415.

Rozier, D., Birol, F., Cosme, E., Brasseur, P., Brankart, J.M. and Verron, J., 2007. A reduced-order Kalman filter for data assimilation in physical oceanography. *SIAM review*, 49(3), pp.449-465.

Sakov, P. and Oke, P.R., 2008. A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. Tellus A, 60(2), pp.361-371.

Sakov, P. and Oke, P.R., 2008. Implications of the form of the ensemble transformation in the ensemble square root filters. *Monthly Weather Review*, 136(3), pp.1042-1053.

Sarma, P., Durlofsky, L.J., Aziz, K. and Chen, W.H., 2006. Efficient real-time reservoir management using adjoint-based optimal control and model updating. *Computational Geosciences*, 10(1), pp.3-36.

Sarma, P., Durlofsky, L.J. and Aziz, K., 2008. Computational techniques for closed–loop reservoir modeling with application to a realistic reservoir. *Petroleum Science and Technology*, 26(10-11), pp.1120-1140.

Silverman, B.W., 1986. The kernel method for univariate data. *In Density Estimation for Statistics and Data Analysis* (pp. 34-74). Springer US.

Solonen, A., Cui, T., Hakkarainen, J. and Marzouk, Y., 2016. On dimension reduction in Gaussian filters. *Inverse Problems*, 32(4), p.045003.

Suwartadi, E. 2012. *Gradient-based methods for production optimization of oil reservoirs* (Doctoral dissertation, Norwegian University of Science and Technology).

van Doren, J. F., Markovinović, R., and Jansen, J. D. 2006. Reduced-order optimal control of water flooding using proper orthogonal decomposition. *Computational Geosciences*, *10*(1), 137-158.

van Leeuwen, P.J., 2009. Particle filtering in geophysical systems. *Monthly Weather Review*, 137(12), pp.4089-4114.

Vermeulen, P. T. M., Heemink, A. W., and Te Stroet, C. B. M. 2004. Reduced models for linear groundwater flow models using empirical orthogonal functions. *Advances in water resources*, 27(1), 57-69.

Willcox, K. 2006. Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition. *Computers and fluids*, 35(2), 208-226.

Yoon, S., Alghareeb, Z. M., and Williams, J. 2014. Development of Reduced-order Oil Reservoir Models using Localized DEIM. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.

Yoon, S., Alghareeb, Z. M., and Williams, J. 2016. Hyper-reduced-order Models for Subsurface Flow Simulatoin. *SPE Journal*.

Zhang, D., Lu, Z. and Chen, Y., 2007. Dynamic reservoir data assimilation with an efficient, dimension-reduced Kalman filter. *SPE Journal*, 12(01), pp.108-117.