UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

# DATA ASSIMILATION USING THE ENSEMBLE KALMAN FILTER WITH EMPHASIS ON THE INEQUALITY CONSTRAINTS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

HEMANT A. PHALE
Norman, Oklahoma
2010

DATA ASSIMILATION USING THE ENSEMBLE KALMAN FILTER WITH EMPHASIS ON THE INEQUALITY CONSTRAINTS

A DISSERTATION APPROVED FOR THE
MEWBOURNE SCHOOL OF PETROLEUM AND GEOLOGICAL ENGINEERING

BY

Dr. Dean S. Oliver, Chair

Dr. Faruk Civan

Dr. Deepak Devegowda

Dr. Chandra Rai

Dr. S. Lakshmivarahan

# ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to Dr. Dean Oliver. I feel extremely fortunate and blessed to have the opportunity of having him as my advisor and mentor during my graduate studies. There is not a shred of doubt in my mind that he is, by far, the best advisor, both professionally and personally, that anyone could ever wish for. He accepted me as one of his graduate students when my first advisor decided to leave OU to join another school and there are no words to describe how grateful I am to Dr. Oliver for accepting me unconditionally and supporting me during those uncertain times. His passion for research, his depth of knowledge, and his drive for excellence in the research and industrial community never stop amazing me. I thank him for his innovative and brilliant ideas, for his patience, and for his caring, loving, and thoughtful nature and his faith in me which immensely helped me to get to this point in my life. My special thanks to Mary Oliver for her kindness and hospitality and for always preparing delicious food whenever I was at their house.

I would like to thank Dr. S. Lakshmivarahan, Dr. Chandra Rai, Dr. Deepak Devegowda, and Dr. Faruk Civan for serving on my committee as well as for their continuous encouragement throughout my graduate studies. I would also like to extend my thanks to all other faculty and staff of the Mewbourne School of Petroleum & Geological Engineering. I would like to extend my deep appreciation to Dr. Dongxiao Zhang, my first advisor at OU, who brought me to Norman. I thank him for giving me the opportunity and for his guidance and support during the early days of my graduate life.

I would like to acknowledge the OU Consortium on Ensemble Methods (OUCEM) and all its member-companies for the financial support during my Ph.D. studies at

OU. I would also like to acknowledge the computational resources provided by the OU Supercomputing Center for Education and Research (OSCER). I thank Dr. Henry Neeman (director of OSCER) and the entire support staff for their instantaneous and uninterrupted help with the supercomputer-related issues. Without their generous support, it would have been difficult to complete this study in a timely manner. I would also like to thank Schlumberger for their generous donation of multiple licenses of ECLIPSE.

I thank Chevron for providing me with internship opportunities for two summers which made me realize the value of working together as a team and gave me an excellent experience of working with the professionals from the oil and gas industry. I am grateful to the excellent mentors I worked with during my two internships at Chevron; Masroor Chaudhri in Houston and Gopi Nalla in San Ramon, California. I would also like to express my gratitude to Ning Liu, my unofficial mentor at Chevron Houston, for her assistance and support during my summer intern projects.

I would also like to thank all my colleagues who are/were part of this research group. I do not have enough words to thank Yan Chen who was always there to help solve my research-related issues, always asked me some "tough" questions to which I had no answers, and always encouraged me to complete my studies. Very special thanks to Yanfen Zhang for all the lengthy discussions we had on research issues and for her thoughtful suggestions and comments as well as critical reviews of my work. Very special thanks to Yao Tong and Yanfen Zhang for their hospitality in inviting me to taste some authentic Chinese food on countless occasions. As I run out of the space here, I would like to thank all my friends and colleagues, whom I could not mention personally, who added different shades of color to my stay here in Norman.

Most importantly, I wish to thank my parents, my brother, and my entire family for their unconditional love, support, and encouragement which gave me the strength and motivation needed to make this happen.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

The reliability of reservoir models generally increases as more data are included during their construction. In the recent past, the ensemble Kalman filter (EnKF) technique has established itself as a viable data assimilation method for models with large numbers of variables. The standard implementation of the EnKF and its variants are being used in a number of disciplines including reservoir engineering.

The standard formulation of the EnKF, however, does not take into account the physical constraints on state variables during the data assimilation step so, the constraint violations are often handled heuristically. When the standard implementation of the EnKF is used for history matching, the resulting updates to reservoir properties sometimes exceed physical bounds, especially when the problem has highly nonlinear model dynamics, the state variables are non-Gaussian, or the ensemble size is small. The physical constraints often contain valuable information about the system which are critical for the valid initialization and the forecast. A successful enforcing of these constraints during the data assimilation step of the EnKF is necessary for valid estimation of reservoir properties.

In the conventional EnKF approach, phase saturations (in case of black-oil models) and molar densities (in case of compositional models) are often included in the state vector. Unfortunately, the analysis step of the EnKF sometimes results in non-physical values of phase saturations or molar densities. In this dissertation, we illustrate the problem of using the standard EnKF with a compositional model in which the updated $CO_2$ molar density in some regions of the model space is observed

to take negative values while molar densities of the remaining components are increased. A number of solutions to the problem of constraint violation have been proposed in the past including the iterative EnKF, transformation of state variables, reparameterization, and truncation. The standard truncation schemes avoid negative values of molar densities, but do not address the problem of increased molar densities of other components. The results can include a spurious increase in reservoir pressure with a subsequent inability to maintain injection. In this dissertation, we present two different methods for incorporating the inequality constraints into the EnKF methodology.

In the first part of this dissertation we present a method for the constrained EnKF (CEnKF) which takes into account the physical constraints on the plausible values of state variables during the data assimilation such that the resulting solution is as close as possible to the unconstrained solution obtained from the standard EnKF, and at the same time, it lies in the feasible region. The proposed method can be implemented in two different approaches, both of which convert the inequality constraints to a small number of equality constraints. The first approach uses Lagrange multipliers to enforce the active constraints. In the second approach, the active constraints are used as virtual observations for calibrating the model parameters within plausible ranges. Applying the CEnKF technique in an iterative manner ensures that the resulting solution is within the limits set by the constraints.

The application of the proposed CEnKF method is successfully demonstrated on a synthetic 1D linear problem, on a synthetic 2D compositional model, and on a highly heterogeneous three-phase flow reservoir model. The effect of the constraints on mass conservation is illustrated using a 1D Buckley-Leverett flow example. Results show that the CEnKF technique is able to enforce the non-negativity constraints on molar densities and the bound constraints on phase saturations (all phase saturations must be between a lower and an upper bound), and achieve a better estimation of reservoir

properties than is obtained using only truncation with EnKF.

An interior-point method for incorporating the inequality constraints into the EnKF methodology (IPCEnKF) is presented in the second part of this dissertation. In this approach, the objective function for data assimilation is reformulated by adding a barrier function to penalize proximity of the state variables to the boundaries of the feasible region and to set a barrier on the state variables against leaving the feasible region. By doing so, the original constrained optimization problem is transformed into an unconstrained optimization problem. We present the solution of the resulting unconstrained problem in the form of a new iterative EnKF scheme which implicitly contains inequality constraints on the state variables. The proposed IPCEnKF method is efficient compared to the CEnKF as it does not require the identification of the active constraints. Although the method is iterative to reduce the effect of the barrier term at each assimilation step, the iterations do not require running the simulator.

The IPCEnKF method is successfully tested first on a 1D linear example to illustrate the performance when nonlinearity is not an issue, then on a more realistic 3D, three-phase reservoir flow assimilation problem based on the modified SPE9 model. Results from the reservoir problem show the effectiveness of the newly proposed IPCEnKF method in matching the observations and honoring the inequality constraints on phase saturations. The proposed method is able to achieve a better estimate of reservoir properties than is obtained using only truncation with the standard EnKF.

# CHAPTER I

# INTRODUCTION

Uncertainty quantification is one of the most important factors in a successful reservoir management process. The uncertainty in reservoir properties such as porosity, permeability, and fluid contacts can have a significant impact on the reservoir performance and the production strategy. One way to quantify the uncertainty in the future reservoir performance is to generate a suitable collection of reliable geological models of the reservoir. The reliability of reservoir models increases as more data are included in their construction. Traditionally, conditional geostatistical simulation (Deutsch and Journel, 1998) has been used to incorporate static data (e.g. geological, geophysical, and well-log or core data) into the construction of reservoir geological models. Although the well-logs or core data can provide a direct measurement of reservoir properties (porosity and permeability), the process of acquiring these data is often highly expensive. Therefore, the static data are available in limited quantity and are often sparsely located in the reservoir field. Geostatistical methods are typically constrained by the requirement of linearity in relationships between model variables and observations. On the other hand, the dynamic production data (e.g. oil production rate, water production rate, gas-oil ratio, and cumulative fluid production rates, etc.) are available in abundant quantity, but are nonlinearly related to model variables over a long distance in the reservoir.

History matching provides a way of incorporating both static data (e.g. well-log or core data) as well as dynamic data (e.g. production data) in the geological reservoir model construction. The traditional approach to history matching often consists of

manually adjusting local reservoir properties (e.g. porosity, permeability, or transmissibility, etc.) so that the flow simulation results using the adjusted parameters match the recorded observations. Traditional history matching methods, including manual history match, simultaneously utilize all the available production data obtained during the entire production history period and often involve heuristic trial-and-error techniques which become impractical for achieving well-by-well history match for reservoir models with large numbers of gridblocks and a large number of wells. On the contrary, data assimilation approach uses available observations in a sequential manner for estimating reservoir properties. Since data are assimilated sequentially as they become available, it provides a way of obtaining the estimate of reservoir properties that honors the production data up to the most recent time. As a direct consequence of this feature of data assimilation, the method is highly suitable for real-time reservoir management. Data assimilation is often called as an assisted or automatic history matching problem and can be considered as an optimization problem when it is formulated under the Bayesian framework (Oliver et al., 2008). Data assimilation is usually formulated as a minimization problem in which the objective is to minimize the mismatch between the actual measurements and their model counterparts with a regularization term (Tarantola, 1987).

Among the possible methods for solving this minimization problem, gradient-based approach is one of the most widely applied techniques. A number of methods including numerical perturbation and sensitivity equation have been developed in this area. The numerical perturbation method is easy to implement, however, as the necessary computations increase proportionally with the number of unknown parameters, the method often breaks down for reservoir models with large numbers of gridblocks with a few numbers of unknown parameters per gridblock to be estimated. The adjoint-based method is a popular approach for history matching (Wu et al., 1999; Li et al., 2003; Liu and Oliver, 2004) among researchers and practitioners as

the computation of gradients from the adjoint-based method is highly efficient. Any history matching method that uses the adjoint approach, however, requires access to the simulator source code and needs significant code development and maintenance on the user's side. A number of challenges like these, involved in using the gradient-based approach prohibit its application for history matching in reservoir engineering on a large scale.

In the past few years, a technique called as the ensemble Kalman filter (EnKF) has attracted significant attention in the research community as a viable data assimilation method. The EnKF technique has a great advantage over the gradient-based approach as it approximates the gradient through the ensemble-based sensitivity which can be obtained from the ensemble of forecasts obtained from the simulator. Thus, the EnKF does not suffer from the challenges including high computational cost or the requirement of the access to the simulator source code as those faced by the gradient-based methods. One only has to work with the inputs and outputs from the reservoir simulator for computing the sensitivity which makes the reservoir simulator a black box. The EnKF was developed by Evensen (1994b) for reducing the limitations on linearity and for allowing data assimilation with reservoir models with large numbers of gridblocks. The EnKF utilizes the dynamical equations to evolve the most probable state (mean) and the error covariance matrix forward in time and the model parameters of each ensemble member are adjusted using the sensitivity computed from the ensemble of state vectors. The first application of the EnKF to the reservoir engineering field was presented by Lorentzen et al. (2001a) where the authors used the EnKF in a dynamic two-phase fluid flow model for updating model parameters. Since its introduction to the petroleum engineering field, the EnKF technique has gained increasing attention for history matching and continuous reservoir model updating (Nævdal et al., 2005; Evensen et al., 2007; Skjervheim et al., 2007). The application of the EnKF is not only limited to the synthetic data assimilation problems but the

method has successfully been demonstrated on a number of pseudo-field cases (Gu and Oliver, 2005; Chen and Oliver, 2010) and real-field data assimilation problems (Bianco et al., 2007; Haugen et al., 2008; Zhang and Oliver, 2010b).

The standard implementation of the EnKF performs well when the model dynamics is not highly nonlinear, the state variables can be approximated by a Gaussian distribution, and the ensemble size is large. In case of highly nonlinear problems, the standard implementation of the EnKF often encounters difficulties in obtaining a reasonable match to the observations. One solution to this issue is to use an iterative approach as discussed by a number of researchers (Zafari et al., 2006; Li and Reynolds, 2009a; Krymskaya et al., 2009). An iterative EnKF method, called as ensemble randomized maximum likelihood (EnRML), was proposed by Gu and Oliver (2007) which uses a Gauss-Newton formulation to iteratively adjust model parameters. The dynamic state variables such as gridblock pressures and phase saturations are obtained from the reservoir simulator. A major issue that occurs in the presence of highly nonlinear model dynamics, non-Gaussian state variables, and small ensemble size is that the standard EnKF results in large adjustments to state variables leading to non-physical updates during the data assimilation step. As the analysis step of the EnKF preserves only the first two moments, violation of the Gaussianity assumption, as in the case with water saturation during a water flood operation, often results in constraint violations where the updated water saturations take non-physical values. A number of solutions have been proposed in the past to avoid the issue of non-physical updates, however, to our knowledge, there are no well-established methods in the literature for incorporating the inequality constraints into the EnKF framework.

In this dissertation, we describe two different methods, the CEnKF and the IP-CEnKF, for incorporating the inequality constraints on the plausible values of state variables during the analysis step of the EnKF methodology. These methods are

derived based on two distinct principles. In the constrained EnKF method, an unconstrained solution is obtained using the standard implementation of the EnKF. The unconstrained solution is subsequently analyzed for any inequality constraint violations for each ensemble member. The inequality constraints (at the location of constraint violations) are converted into a small number of equality constraints before using them as pseudo-observations during the CEnKF data assimilation step. Although the CEnKF is an iterative method and additional constraint violations are identified during each iteration, it is observed during this study that a small number of iterations are often sufficient for obtaining a solution that either lies in the feasible region completely or gets sufficiently close to being in the feasible region. Thus, the CEnKF method utilizes the inequality constraints in the form of equality constraints for projecting the unconstrained solution on to the feasible region. On the other hand, the IPCEnKF method formulates the data assimilation problem in such a manner that the resulting solution stays within the boundaries of the feasible region at all times. The IPCEnKF method is also an iterative technique and shows a close resemblance to the EnRML. The fundamental difference between the CEnKF and the IPCEnKF methods is that the CEnKF method gradually moves the unconstrained solution towards the feasible region whereas the solution obtained from the IPCEnKF method always lies in the interior of the feasible region. Although both methods are iterative in nature, the iterative process does not require running the reservoir simulator.

This dissertation contains six chapters. Chapter 2 presents a brief summary of the EnKF technique. A number of issues involving the application of the EnKF method such as model error, sampling error, localization, assimilation of large amounts of observations, and handling of equality and inequality constraints on the state variables are discussed. The evolution of the EnKF technique in terms of its application

to models with increasing dimensions and estimation of different types of model parameters is also mentioned. Chapter 3 illustrates the application of the standard implementation of the EnKF to a synthetic 2D compositional model and points out the resulting issue with the injection of fluids into reservoir. A detailed analysis of one of the ensemble members is presented for demonstrating the absence of a constraint-enforcing mechanism during the analysis step of the EnKF. Chapter 4 introduces the constrained EnKF method (CEnKF) and presents two different approaches for its implementation. The effectiveness of the CEnKF is demonstrated through its application to a synthetic 1D linear example, to a synthetic 2D compositional model, and to a highly heterogeneous three-phase flow reservoir model. Chapter 5 presents the interior-point method for incorporating the inequality constraints into the EnKF methodology (IPCEnKF). The newly proposed IPCEnKF method is tested first on a 1D linear example to demonstrate its effectiveness when the nonlinearity is not an issue, then on a more realistic 3D, three-phase reservoir flow assimilation problem based on the modified SPE9 model. Chapter 6 summarizes the main findings and conclusions drawn from this study.

# CHAPTER II

# THE ENSEMBLE KALMAN FILTER AND RELATED ISSUES

## 2.1 Introduction

The EnKF is an ensemble-based extension of the Kalman filter (Kalman, 1960). The Kalman filter was originally developed to continuously update the states of linear systems to account for data assimilation in linear problem with Gaussian priors. The Kalman filter seeks a solution that minimizes the objective function involving the sum of two quantities: (1) mismatch between the actual observations and the simulated observations predicted from the solution and (2) mismatch between the prior knowledge about the state of the system and the solution obtained from the Kalman filter algorithm. Under this framework, the classical Kalman filter can be thought of as an optimization algorithm where the solution maximizes the posterior probability density of the state of the system given the observations and the prior knowledge of the state.

Due to the underlying structure, however, the Kalman filter becomes impractical for state vectors (a vector consisting of static model parameters and dynamic state variables) with large dimensions as the computational efforts required for updating the covariance matrix are prohibitive. The Kalman filter was extended for dealing with the nonlinearity in the models through the extended Kalman filter (Jazwinski, 1970) which uses linearizations of the model and observation equations around the estimated mean of the state. Even though the extended Kalman filter shows promising results for a wide variety of problems, it breaks down for highly nonlinear models. In

addition, the extended Kalman filter requires updating the covariance matrix which makes this method, like the standard Kalman filter, infeasible for models with large numbers of gridblocks. In 1994, Evensen (1994b) introduced the technique of the ensemble Kalman filter (EnKF) to reduce the limitations on linearity and to allow assimilation with large state vectors. The EnKF uses an ensemble of model realizations for approximating the probability distribution of the state vector and relies on the Monte Carlo approach to forecast the error statistics. The mean state and the associated error covariance matrix are propagated forward in time using a set of dynamical equations. Evensen (1994a) mentioned about the successful application of EnKF in resolving the closure problems reported from the applications of the extended Kalman filter. Unlike the extended Kalman filter, EnKF avoids the linearization of the nonlinear model dynamics and uses a reduced order approximation of the covariance matrix which makes it suitable for data assimilation problems with large numbers of state variables.

A number of improvements to the implementation of the EnKF have been suggested since the method was originally introduced by Evensen (1994b). Due to the ease of implementation and the robustness, the EnKF in its standard form and its variants have gained increasing attention for data assimilation in a wide spectrum of fields ranging from meteorology (Houtekamer and Mitchell, 1998; Hamill et al., 2000; Anderson, 2001; Houtekamer et al., 2005), to oceanography (Evensen and van Leeuwen, 1996; Keppenne and Rienecker, 2002; Bertino et al., 2003; Gerber and Joos, 2010), and to groundwater (Reichle et al., 2002; Moradkhani et al., 2005; Chen and Zhang, 2006; Zhang et al., 2007). Evensen (2003) reviewed some of the important results from a number of studies and documented a chronological sequence of the developments in the application of the EnKF. Although the EnKF is being used in a number of different fields, Oliver et al. (2010) summarized some of the key points of the data assimilation for multiphase flow problems in reservoir engineering that make

the problem distinctly different from data assimilation problems in other areas such as meteorology and oceanography. Aanonsen et al. (2009) presented a comprehensive review of a number of applications where the EnKF has been used for problems specific to the petroleum industry.

The EnKF updates reservoir models by assimilating observations in a sequential manner. In the process, it uses the cross-covariances between observations and model parameters estimated from the ensemble. Liu and Oliver (2005b) cited a number of reasons for which large, nonlinear models show difficulty in terms of sequential data assimilation. In spite of the challenges, the method has gained widespread acceptance in the petroleum industry simply because it uses a reduced order approximation of the covariance matrix and the computation of the entire cross-covariance is not required which makes the method more suitable for large-scale fluid flow problems in reservoir engineering. Also, the correlation between the model parameters and theoretical data is only a function of the information contained in the ensemble, thus making the EnKF method independent of the reservoir simulator. The historical production data are sequentially assimilated which makes EnKF suitable for reservoir monitoring and performance prediction (Nævdal et al., 2002, 2005). Since its introduction to petroleum engineering field, the EnKF technique has gained increasing attention for history matching and continuous reservoir model updating (Lorentzen et al., 2005; Liu and Oliver, 2005b; Bianco et al., 2007; Evensen et al., 2007; Haugen et al., 2008; Agbalaka and Oliver, 2009; Chen and Oliver, 2010; Zhang and Oliver, 2010b).

There has been a remarkable progress in terms of the application of the EnKF technique for history matching with reservoir models that are increasingly complex in nature and size. In the first application of the EnKF to the petroleum engineering problem, Lorentzen et al. (2001b) used the EnKF technique for tracking observations in case of a underbalanced drilling operation. In this study, model parameters for a dynamic two-phase model of fluid flow in a vertical well were estimated by assimilating

measured data such as the pump pressure, the bottomhole pressure, and the liquid and gas flow rates at the outlet. The first application of the EnKF to reservoir monitoring was presented by Nævdal et al. (2002) where the EnKF was used on a 2D (vertical cross-section) model for forecasting production of a near-well reservoir and for tracking the production of the two phases in different inflow zones. In 2005, the reservoir engineering field saw the first application of the EnKF technique on field-like synthetic cases (Nævdal et al., 2005) where a simplified 2D field model was constructed by taking out a horizontal layer from a model of a North Sea field. Nævdal et al. (2005) applied the EnKF for assimilating historical production data from 14 production wells and for estimating permeability field. In addition, Nævdal et al. (2005) used the EnKF technique on a 2D synthetic example with model size of 2500 gridblocks where the permeability field was tuned by assimilating well pressure, oil production rate, producing gas-oil ratio, and water cut data at two producers and a single injector.

Gu and Oliver (2005) initialized a series of studies using the PUNQ-S3 model where both porosity and permeability fields were estimated. The dimensions of the PUNQ-S3 reservoir model were $19\times28\times5$ with 1,761 active gridblocks. There were 6 producing wells with no injectors in the field. The oil production rate was identical for all producers, however, as Gu and Oliver (2005) reported, the actual oil production rate values varied as wells in some simulation models were unable to make the target. Gu and Oliver (2005) assimilated four kinds of production data including bottomhole pressure, producing gas-oil ratio, water cut, and oil production rate. In another study, Gao et al. (2006) applied the EnKF on the PUNQ-S3 problem for uncertainty quantification, however, as all six producing wells were constrained to a fixed production rate, Gao et al. (2006) did not include oil production rate data in the state vector, but instead assimilated only water cut, wellbore pressure, and producing gas-oil ratio during the data assimilation procedure. The atmospheric pressure was set as the

minimum bottomhole pressure in their study. Skjervheim et al. (2007) presented the first application of the EnKF technique on a field case where the authors used the EnKF for assimilating 4D seismic data and production data including water cut and producing gas-oil ratio for a North Sea field. The reservoir properties such as porosity and permeability were considered to be uncertain parameters for the reservoir model with 14 layers and 29,580 active gridblocks.

Although the EnKF was used to estimate spatially correlated rock properties such as porosity and permeability during the initial applications of EnKF, since then, the types of variables that could be estimated using the EnKF have been expanded dramatically to include, for example, discontinuous variables such as reservoir rock facies (Liu and Oliver, 2005b,a; Agbalaka and Oliver, 2008, 2009). Evensen et al. (2007) and Seiler et al. (2009) used EnKF as a tool for assisted history matching on a complex North Sea field. The reservoir simulation model consisted of 40 layers with 60,000 active gridblocks. A number of model parameters including permeability, porosity, relative permeability, fault transmissibility multipliers, and vertical transmissibility multipliers were estimated by assimilating oil production rate, water cut, and producing gas-oil ratio data. A number of researchers have also attempted to estimate the uncertainty in the fluid contacts such as water oil contact and gas oil contact (Evensen et al., 2007; Thulin et al., 2007). Chen and Oliver (2010) also estimated the uncertainty in the relative permeability curves along with other model parameters such as porosity, permeability (in three directions), and net-to-gross ratio during the application of an iterative form of EnKF to the Brugge benchmark study. In an application of the EnKF to a deepwater reservoir field with 5 layers and over 95,000 active gridblocks, Zhang and Oliver (2010b) used a hierarchical stochastic model and included porosity and permeability along with the uncertain trend coefficients into the state vector. The known application of EnKF for history matching and uncertainty quantification on the largest reservoir simulation model was presented by Cominelli

et al. (2009). The simulation model of the deepwater reservoir field used by Cominelli et al. (2009) consisted of 44 layers and over 410,000 active gridblocks where porosity and permeability fields were tuned by assimilating oil production rate, water production rate, and static pressure data. An attempt has also been made to estimate the geomechanical properties such as stress, strain, and displacement fields (Chang et al., 2010). Recently, Irving and Robert (2010) presented a workflow for quantifying the uncertainty in the structural parameters such as fault permeability and thickness. Irving and Robert (2010) used the Barnhill field to compare the performance of the EnKF-based workflow with another method named factorial experimental design and pointed out different advantages and disadvantages of these two methods.

In the following section, we briefly discuss the methodology for the standard implementation of the EnKF.

## 2.2 The ensemble Kalman filter

The EnKF is a sequential data assimilation method. It is a Monte Carlo technique in which an ensemble of states is propagated forward in time to obtain the time evolution of the mean and the covariance of the ensemble of model realizations. The mean of the ensemble forecast represents the best estimate of the population mean whereas the ensemble members quantify the corresponding uncertainty in the estimate. Most of the Monte Carlo methods require a large sample size, while the EnKF utilizes much less number of samples due to the conditioning of state variables to data along the sequential data assimilation process. In other words, the samples obtained during the analysis step of the EnKF move to the region of higher probability as data are continually assimilated. In petroleum engineering field, the process of evolving reservoir models forward in time is often accomplished by running a reservoir simulator which can be expensive for reservoir models with large numbers of gridblocks. A small ensemble size can be advantageous while working with such models where the

computational time required for running the simulator can be the main concern.

### 2.2.1 Parameterization

In order to characterize a reservoir and quantify the uncertainty in its future reservoir performance, one needs to quantify the uncertainty within a number of gridblock-based parameters as well as global properties. Therefore, the first step in the history matching process is to identify the important parameters which influence the performance of the simulation model and therefore need to be estimated. Traditional history matching methods, e.g. manual history match, are not capable of estimating large numbers of uncertain variables and often require to include a low number of model parameters in the workflow. Currently, one of the most popular history matching methods used in the petroleum industry is the Design of Experiments which uses a sensitivity study to identify the most significant parameters. Since it is necessary to evaluate the performance of the reservoir model for different combinations of the unknown model parameters during the sensitivity study, the number of unknown parameters that can be estimated using such technique is severely limited and only global parameters such as fault transmissibility multiplier or water oil contact can be determined. The EnKF, however, does not suffer from these disadvantages and the numbers of unknown parameters that can be estimated are not limited. The fundamental reason is that the EnKF uses an ensemble of realizations for representing the probability distribution of state variables. With this ensemble approximation, the EnKF effectively reduces the dimension of the problem from the number of uncertain variables in the state vector to the number of ensemble members. As a direct consequence of this, the solution obtained from the most basic form of the EnKF algorithm lies in the subspace spanned by the ensemble members.

Due to the improvements to the EnKF methodology, a wide range of uncertain model parameters can be estimated using the EnKF including porosity, horizontal

permeability, vertical permeability, depth of fluid contacts, horizontal and vertical barriers to fluid flow, net-to-gross ratio, relative permeability curves, fault transmissibility multipliers, vertical transmissibility multipliers, and trend coefficients, etc. The most important gridblock-based parameters included into the state vector, however, are often porosity and permeability. Since the logarithm of permeability (referred to as log-permeability in this dissertation) is sometimes approximately Gaussian, the log-permeability instead of permeability is often included in the data assimilation process. The reservoir gridblock properties such as porosity and log-permeability do not change with time. Hence, they are termed as the static model parameters. The uncertainty associated with some of the global reservoir parameters such as fluid contacts, fault transmissibility, and net-to-gross ratio can also be quantified using the EnKF. However, in all the reservoir flow problems discussed in this dissertation, we use the standard parameterization for the EnKF and include the two most important static model parameters, gridblock porosity and log-permeability, into the state vector.

### 2.2.2 Generation of the initial ensemble

An initial ensemble of porosity and log-permeability realizations can be generated using a number of statistical tools such as sequential Gaussian simulation (SGSim), Cholesky decomposition, and the method of moving averages, etc. If all the realizations from the initial ensemble are required to honor the porosity and log-permeability data at well locations, one can use the conditional simulation for generating the initial ensemble. The prior mean and the prior covariance for generating the initial ensemble of porosity and log-permeability can be determined from the geological reservoir model (if available). In industrial practice, the geologists often take into account the information available from sources such as cores, well-logs, and seismic data for estimating the statistical parameters (e.g. mean and covariance) of model parameters.

As mentioned earlier, the dynamic state variables are also included into the state vector. For reservoir engineering applications, pressures and phase saturations (water, gas, and oil) at each gridblock represent the dynamic state of the reservoir at any given time. These properties change with time and therefore, they are termed as the dynamic state variables. The reservoir field is assumed to be in a static equilibrium condition at the initial time prior to any production from the field. As a result, the spatial distribution of the dynamic state variables at the initial condition is often assumed to be known without any uncertainty. The ensemble of reservoir models is initialized with these static model parameters and dynamic state variables.

### 2.2.3 Forecast and analysis

The EnKF methodology is a recursive two step process. The first step is the forecast step in which the system is evaluated forward in time using a forward model. The second step, often called the updating or the analysis step, is based on the Bayesian framework for merging the prior (or background) information about the current state based on the previous data up to the current time, and current data (or observations). Each ensemble member is evaluated from the current time step to the next data assimilation time step at which the observations are available. The discrepancy encountered due to the lack of the knowledge and the expertise to model the actual system dynamics is usually the source of the model error. To this date, a reliable method for quantifying model error in petroleum reservoir flow problems has not been developed. An inappropriate use of the model error, however, can introduce bias in the estimates. For all the examples discussed in this dissertation, we consider a "perfect model" with no model error.

Let us assume that the static model parameters are denoted by $\mathbf{m}$ and the dynamic state variables, $\mathbf{u}^k$, at time $t_k$ are a nonlinear function of the state variables at time

$t_{k-1}$ and the model parameters, $\mathbf{m}$,

$$\mathbf{u}^k = f(\mathbf{u}^{k-1}, \mathbf{m}).$$

In reservoir engineering, a reservoir fluid flow simulator often represents the forward model, $f(\cdot)$, and the model error is typically neglected. The predicted data, $\mathbf{d}^k$, at time $t_k$ are functions of the state and model variables at $t_k$.

$$\mathbf{d}^k = g(\mathbf{u}^k, \mathbf{m}).$$

It is assumed that the observed data are functions of the true state and model variables with zero-mean Gaussian noise added.

$$\mathbf{d}_{\text{obs}} = g(\mathbf{u}^{\text{true}}, \mathbf{m}^{\text{true}}) + \nu$$

where, the covariance of the data is $\mathbf{C}_D$. The measurements are generally independent of each other, and therefore, $\mathbf{C}_D$ is often a diagonal matrix where the diagonal elements represent the measurement error variances. In reservoir engineering applications, the measurements used in the analysis step often include the production data (e.g. oil production rate, gas-oil ratio, water cut at the well, etc.) recorded during the production history of the reservoir. To reduce the number of subscripts and superscripts, we will omit the notation for the time index in the following equations.

For most reservoir engineering applications, a joint vector consisting of the static model parameters ($\mathbf{m}$), the dynamic state variables ($\mathbf{u}$), and the predicted (simulated) data ($\mathbf{d}$) is often used. In that case, the augmented state vector $\mathbf{y}_j$ for the $j$th ensemble member can be represented as

$$\mathbf{y}_j = \begin{bmatrix} \mathbf{m} \\ \mathbf{u} \\ \mathbf{d} \end{bmatrix}_j \tag{2.1}$$

and the collection of all ensemble state vectors is denoted by $\mathbf{Y}$ as

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_{N_e}]. \tag{2.2}$$

The matrix $\mathbf{Y}$ is of dimensions $N_y \times N_e$ where $N_e$ is the ensemble size. In Eq. 2.1, $\mathbf{m} \in \mathbb{R}^{N_m}$ and $\mathbf{u} \in \mathbb{R}^{N_s}$ where $N_m$ and $N_s$ denote the dimensions of the static model parameters and the dynamic state variables, respectively. As a result of the use of the augmented state vector, the measurement operator $\mathbf{H}$, is a matrix consisting of 0s and 1s that selects the simulated data from the state vector. The analysis step in the standard EnKF computes variables that approximately minimize

$$J_j(\mathbf{y}) = \frac{1}{2}(\mathbf{y} - \mathbf{y}_j^f)^T(\mathbf{C}_Y^f)^{-1}(\mathbf{y} - \mathbf{y}_j^f) + \frac{1}{2}(\mathbf{H}\mathbf{y} - \mathbf{d}_{\text{obs},j})^T\mathbf{C}_D^{-1}(\mathbf{H}\mathbf{y} - \mathbf{d}_{\text{obs},j}). \quad (2.3)$$

As a result, each state vector of the ensemble is updated at a data assimilation step using the ensemble approximation to the Kalman gain in the following manner:

$$\mathbf{y}_j^u = \mathbf{y}_j^f + \mathbf{K}(\mathbf{d}_{\text{obs},j} - \mathbf{H}\mathbf{y}_j^f), \qquad \text{for} \quad j = 1, 2, \ldots, N_e \quad (2.4)$$

where, the superscripts $f$ and $u$ stand for forecast and update, respectively. In this study, we have used the perturbed observation approach (Burgers et al., 1998; Houtekamer and Mitchell, 1998), for which random observation perturbations with the same covariance as the true observation error are added to observations in each ensemble member,

$$\mathbf{d}_{\text{obs},j} = \mathbf{d}_{\text{obs}} + \nu_j.$$

Although the perturbed observation approach is widely used by the data assimilation community, the perturbations to the observations can serve as a source of additional sampling error in which case Tippett et al. (2003) suggest using methods that do not require perturbations. In Eq. 2.4, $\mathbf{K}$ is the ensemble approximation to the Kalman gain and is computed as

$$\mathbf{K} = \mathbf{C}_Y^f\mathbf{H}^T(\mathbf{H}\mathbf{C}_Y^f\mathbf{H}^T + \mathbf{C}_D)^{-1}. \quad (2.5)$$

In Eq. 2.5, $\mathbf{C}_Y^f$ denotes the prior covariance matrix for the state variables. An alternate expression for the Kalman gain $\mathbf{K}$ in Eq. 2.4 can be given as

$$\mathbf{K} = \mathbf{C}_{yd}^f(\mathbf{C}_{dd}^f + \mathbf{C}_D)^{-1}, \quad (2.6)$$

where, $\mathbf{C}_{yd}^f$ is the sample covariance between state variables and simulated data and $\mathbf{C}_{dd}^f$ is the sample covariance between different simulated data. In EnKF, the covariances $\mathbf{C}_{yd}^f$ and $\mathbf{C}_{dd}^f$ are estimated from an ensemble of realizations.

Substituting Eq. 2.5 into Eq. 2.4, the expression for updating each state vector of the ensemble can be expressed as

$$\mathbf{y}_j^u = \mathbf{y}_j^f + \mathbf{C}_Y^f \mathbf{H}^T (\mathbf{H}\mathbf{C}_Y^f \mathbf{H}^T + \mathbf{C}_D)^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{H}\mathbf{y}_j^f), \quad \text{for} \quad j = 1, 2, \ldots, N_e. \tag{2.7}$$

The mean of the updated ensemble can be computed from Eq. 2.7 as

$$\overline{\mathbf{y}^u} = \overline{\mathbf{y}^f} + \mathbf{C}_Y^f \mathbf{H}^T (\mathbf{H}\mathbf{C}_Y^f \mathbf{H}^T + \mathbf{C}_D)^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{H}\overline{\mathbf{y}^f}), \tag{2.8}$$

where, $\mathbf{d}_{\text{obs}}$ denotes the vector of true observations. The prior covariance matrix for the state variables, $\mathbf{C}_Y^f$ can be estimated from the ensemble using the following expression:

$$\mathbf{C}_Y^f = \frac{1}{N_e - 1} (\mathbf{Y}^f - \overline{\mathbf{Y}^f})(\mathbf{Y}^f - \overline{\mathbf{Y}^f})^T \tag{2.9}$$

where, $\overline{\mathbf{Y}^f}$ is the mean of state variables calculated across the ensemble members and is a vector of dimension $N_y$. The state covariance matrix can be very large, but it is never necessary to compute the entire matrix which makes the implementation of the traditional EnKF so efficient. Instead, we compute products of the form $(\mathbf{Y}^f - \overline{\mathbf{Y}^f})^T \mathbf{H}^T$ which are typically much smaller in dimension than the state covariance matrix.

## 2.3 Application of EnKF to reservoir engineering problems

The performance of the EnKF for reservoir engineering related applications with highly nonlinear model dynamics and non-Gaussian state variables is well documented. Gu and Oliver (2005) reported the use of the EnKF for automatic history matching and quantifying the uncertainty in the future reservoir performance of the

PUNQ-S3 synthetic model. Gu and Oliver (2005) showed that the EnKF gave satisfactory history-matching results and required significantly less computational work compared to traditional methods. The results of the cumulative oil production after 16.5 years obtained from the EnKF (Gu and Oliver, 2005) were compared with those summarized by Floris et al. (2001) which showed that the EnKF outperformed the traditional methods including multiple maximum likelihood or maximum *a posteriori* solution using different initial models. Zhang and Oliver (2010b) reported success with a hierarchical EnKF for matching the historical production data in case of a Gulf of Mexico deepwater reservoir field. A significant increase in the ability to match the water cut and other production data using this method was demonstrated by Zhang and Oliver (2010b). In a recent article, Peters et al. (2010) summarized the results of the Brugge benchmark study which was designed to test the combined use of flooding-optimization and history-matching methods. A total of nine participants took part in this study. The history-matching methods used by the participants varied from ensemble-based approach (EnKF and EnRML) to the gradient-based approach to the streamline-based generalized travel time inversion method. The participants who used EnKF or the iterative form of the EnKF (EnRML) approach for the history-matching part obtained the highest net present value (NPV) of the reservoir which underlines the power and the ability of the EnKF for history-matching large-scale reservoir models with complex geology.

Although the EnKF is being regarded as one of the promising data assimilation techniques in reservoir engineering related applications, there are a number of issues involving the practical application of the EnKF that are being studied for making the method more robust and increasing its applicability. Some of these key elements are described in the following section.

### 2.3.1 Model error

For systems of interest in the reservoir engineering field, one is often interested in estimating model parameters (e.g. porosity and permeability) using reservoir response (e.g. different types of production data) rather than determining exactly the initial reservoir conditions. The production data including oil production rate, water production rate, producing gas-oil ratio, well bottom hole pressure, etc. are typically measured in the field during various stages of production. During the history matching process using the EnKF, an accurate forecast is critical; only a perfect model guarantees such an accurate forecast. For real-field applications, a reservoir simulator is often used to simulate reservoir conditions and production scenarios. Although there has been a great progress in the field of reservoir simulators, the state-of-the-art simulator does not completely honor the physics of the multiphase fluid flow through porous media and is an 'imperfect model' as compared to the true reservoir. Using such an imperfect model, it is highly unlikely to produce correct probabilistic forecasts capable of bounding the true state of the system.

Hansen (2002) proposed to sample from both state uncertainty space and model uncertainty space while dealing with the imperfect model. Hansen (2002) suggested that in order to handle the model error, or rather to ignore the influence of model error, one should use a probabilistic framework for ensemble forecasting and should switch to a deterministic framework for data assimilation. Mitchell and Houtekamer (2000) developed a method of accounting for model error in the EnKF context by adding an ensemble of realizations of model error to the ensemble of model predictions. Mitchell and Houtekamer (2000) parameterized the model error in terms of a small number of parameters which were estimated at each analysis time. A number of researchers (Pham, 2001; Anderson, 2001; Whitaker and Hamill, 2002) used the approach where the ensemble-based covariances were scaled using a tunable factor to account for both inbreeding effects in the data assimilation process and model error. To account for

model error, Mitchell et al. (2002) used an additive representation which the authors claimed, is more general than multiplication by a tunable factor and presented a method of generating balanced model perturbations which were used to generate the initial ensemble and to simulate model error. Evensen (2006) provides a thorough discussion of model errors in connection with the EnKF and suggests a method for estimating model error as a part of the data assimilation process. Lødøen and Omre (2008) have developed a method for dealing with model bias caused by grid coarseness.

Most of the research on handling of the model error has been done for weather systems which are inherently different from the petroleum reservoir flow problems. In addition, a number of techniques discussed above require the use of a large size of ensemble which prohibits their implementation on a real-scale data assimilation problem with large numbers of state variables. To this date, a reliable method for quantifying model error in reservoir engineering related applications has not been developed. An inappropriate use of the model error, however, can lead to undesired results.

### 2.3.2 Sampling error and localization

Sampling error is another important aspect that needs to be looked at carefully during the implementation of the EnKF. There are two main sources of sampling error: (1) stochastic perturbations added to the observations and (2) small ensemble size. Although the original formulation of the EnKF presented by Evensen (1994b) did not include stochastic perturbations to the observations, the importance of including random observations to each ensemble member was pointed out independently by Burgers et al. (1998) and Houtekamer and Mitchell (1998). Deterministic methods for data assimilation were also developed to avoid sampling issues associated with the use of perturbed observations in the stochastic analysis ensemble update methods. These deterministic methods include the ensemble transform Kalman filter (Bishop

et al., 2001), the ensemble adjustment filter (Anderson, 2001), and the ensemble square-root filter (Whitaker and Hamill, 2002; Tippett et al., 2003; Evensen, 2004). In the implementation of the square-root filter and all its variants, the ensemble mean is updated first followed by the update of the ensemble perturbations.

A small ensemble size is another element which introduces sampling error in the data assimilation process. During the forecast step of the EnKF, all model realizations of the reservoir are run forward in time using a reservoir simulator. Hence, for reservoir models with few hundreds of thousands of gridblocks, the forecast step often becomes highly expensive. Therefore, for all practical purposes, one wishes to use as small an ensemble as possible. One problem that arises as a result of a small ensemble size, however, is that the sample covariance computed from a small ensemble may suffer from unrealistic spurious correlations. These spurious correlations, when used for computing the Kalman gain, cause changes to model parameters and state variables in regions of no real influence. Lorenc (2003) showed that the unrealistic updates can ultimately result in the loss of the ensemble variability and final breakdown of the EnKF, and concluded that the "harm done by spurious covariances with distant grid points is greater than the local benefit".

In the EnKF, the covariance is approximated from an ensemble that is often orders of magnitude smaller than the number of variables in the state vector. The error in the estimate of the covariance decreases at a rate which is inversely proportional to the square-root of the ensemble size. Therefore, using a large ensemble can result in an accurate estimate, while a small ensemble can lead to an estimate with large errors. An effective approach for reducing the harmful effect of spurious correlations on the update is to use distance-dependent covariance localization. The first application of localization in the EnKF was presented by Houtekamer and Mitchell (1998) where a distance-based cutoff was applied to the Kalman gain. Following on their previous study, Houtekamer and Mitchell (2001) described the use of a tapering function for

localizing covariances of the background error. A fifth-order compactly supported correlation function of Gaspari and Cohn (1999) appears to be a widely accepted choice for covariance localization. A tapering form of the localization function, similar to Houtekamer and Mitchell (2001), has been used by a number of researchers (Hamill et al., 2001; Houtekamer et al., 2005; Furrer and Bengtsson, 2007). The distance-based localization is often applied to the prior estimate of the covariance matrix where an element-by-element product is computed between the covariance matrix and a compactly supported positive-definite matrix. Refer to Ehrendorfer (2007) and Aanonsen et al. (2009) for additional discussion on covariance localization. The distance-based localization can also be applied to the Kalman gain and has been used by a number of researchers leading to improved results (Anderson, 2007; Agbalaka and Oliver, 2008; Zhang and Oliver, 2010b; Chen and Oliver, 2010).

Although the distance-dependent localization is an effective method for eliminating the spurious correlations, this approach suffers from a number of challenges including determining optimal correlation length and determining appropriate type of localization for a given datum. In a comprehensive study, Chen and Oliver (2009) presented a number of critical points about the application of distance-based covariance localization for data assimilation in reservoir engineering related applications. Some of their conclusions include: (1) the sensitivity and the prior covariance should be considered together for determining the region of non-zero cross-covariance, (2) the same type of datum might require different localization at different times, and (3) different model parameters and state variables should be localized differently. Due to the difficulty in applying distance-based localization, more general methods have been proposed (Anderson, 2007; Hacker et al., 2007) which assume that the prior estimate of cross-covariance may not be zero beyond a certain distance as the case is with the distance-dependent localization. More recently, Zhang and Oliver (2010c) developed

a bootstrap-based screening method for assessing the confidence level of each element from the Kalman gain matrix. In a follow-up study, Zhang and Oliver (2010a) evaluated the performance of bootstrap-based screening on the Kalman gain and on the covariances. They showed that the screening approach (either Kalman gain or covariance) results in better estimation and the screening Kalman gain outperforms the distance-based localization methods in terms of generality for application.

### 2.3.3 Assimilation of large numbers of observations

As mentioned earlier, the number of ensemble members are usually kept to a reasonably low value during all practical applications of the EnKF. The analysis step of the EnKF carries out a linear update where each updated ensemble member is a linear combination of the initial ensemble. In case of a field application with large numbers of wells, it is possible to have large numbers of production data, which are greater than the number of ensemble members, for assimilation during each data assimilation step. All the production data, however, may not be independent of each other. Localization, as discussed in Section 2.3.2, can also help in increasing effective rank of the ensemble and thus may help in resolving the issue with the degrees of freedom required for assimilating large numbers of production data. The application of the EnKF for assimilating production data along with 4D seismic data was demonstrated by Skjervheim et al. (2007). The 4D seismic data usually involves the difference between seismic attributes recorded at the time of the first and second seismic surveys and such type of observations are typically available for each gridblock in the reservoir simulation model. Therefore, assimilating 4D seismic data in case of a reservoir model with few hundreds of thousands of gridblocks with relatively small ensemble size (typically a couple of hundreds) results in insufficient degrees of freedom as the number of independent data is greater than the number of ensemble members. As Aanonsen et al. (2009) pointed out, assimilating large amounts of

observations, such as seismic data, also results in a computational problem where it becomes infeasible to compute the Kalman gain matrix as it requires inverting the matrix $(\mathbf{H}\mathbf{C}_Y^f\mathbf{H}^T + \mathbf{C}_D)$. This issue can be resolved by computing the singular value decomposition of $\mathbf{H}\Delta\mathbf{Y}$. The data mismatches and simulated measurement errors can then be projected onto the subspace spanned by the principal left singular vectors of $\mathbf{H}\Delta\mathbf{Y}$ (Evensen, 2004; Evensen et al., 2007; Skjervheim et al., 2007). This effectively transforms the matrix $(\mathbf{H}\mathbf{C}_Y^f\mathbf{H}^T + \mathbf{C}_D)$ from its original dimensions of $N_d \times N_d$ to the new dimensions of $N_e \times N_e$. With this transformation, it becomes feasible to invert the matrix $(\mathbf{H}\mathbf{C}_Y^f\mathbf{H}^T + \mathbf{C}_D)$ needed in the computation of the Kalman gain.

### 2.3.4 Nonlinear model dynamics and non-Gaussian priors

A successful application of the standard EnKF is based on two fundamental assumptions: (1) the prior distribution of model parameters can be approximated by a Gaussian distribution and (2) the relationship between observations and model parameters is not strongly nonlinear. In addition, use of a large ensemble size can also aid in obtaining good results during the application of EnKF. As discussed in Section 2.2.3, the EnKF consists of forecast and analysis steps. In reservoir engineering related applications, the forecast step is often carried out by running a numerical reservoir simulator which can be highly expensive for reservoir models containing hundreds of thousands of gridblocks. Therefore, a practical application of the EnKF on a realistic reservoir model requires the use of as small ensemble size as possible. The governing principles behind a reservoir simulator used in the forecast step of the EnKF are the highly nonlinear dynamical equations for multiphase flow in porous media which are solved in a discretized manner. In case of such nonlinear problems, the dynamic state variables (e.g. pressure, phase saturation, etc.) updated at the analysis step of the EnKF are likely to be inconsistent with those obtained from rerunning the reservoir model from initial time using the updated static model parameters (e.g. porosity,

permeability, etc.).

For a nonlinear problem, the augmented state vector for the $j$th ensemble member (Eq. 2.1) can be represented as

$$\mathbf{y}_j = \begin{bmatrix} \mathbf{m} \\ \mathbf{u} \\ \mathbf{d} \end{bmatrix}_j .$$

In reservoir engineering related applications, $\mathbf{u}$ and $\mathbf{d}$ are highly nonlinear functions of model parameters, $\mathbf{m}$. If $\mathbf{u}$ and $\mathbf{d}$ are linear functions of $\mathbf{m}$ then let us assume that

$$\mathbf{u} = \mathbf{F}\mathbf{m} \qquad \text{and} \qquad \mathbf{d} = \mathbf{G}\mathbf{m}, \tag{2.10}$$

where, $\mathbf{F}$ and $\mathbf{G}$ are matrices of dimensions $N_s \times N_m$ and $N_d \times N_m$, respectively. Here, $N_d$ is the number of observations. With these notations and without the loss of generality, a state vector, $\mathbf{y}$ for a linear problem can be represented as

$$\mathbf{y} = \begin{bmatrix} \mathbf{m} \\ \mathbf{F}\mathbf{m} \\ \mathbf{G}\mathbf{m} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{F} \\ \mathbf{G} \end{bmatrix} \mathbf{m}. \tag{2.11}$$

If $\mathbf{C}_M$ represents the covariance of the model parameters $\mathbf{m}$, then the covariance of the state vector $\mathbf{y}$ can be represented as

$$\mathbf{C}_Y = \begin{bmatrix} \mathbf{I} \\ \mathbf{F} \\ \mathbf{G} \end{bmatrix} \mathbf{C}_M \begin{bmatrix} \mathbf{I}^T & \mathbf{F}^T & \mathbf{G}^T \end{bmatrix}. \tag{2.12}$$

The right hand side of Eq. 2.12 can be expanded as

$$\mathbf{C}_Y = \begin{bmatrix} \mathbf{C}_M & \mathbf{C}_M\mathbf{F}^T & \mathbf{C}_M\mathbf{G}^T \\ \mathbf{F}\mathbf{C}_M & \mathbf{F}\mathbf{C}_M\mathbf{F}^T & \mathbf{F}\mathbf{C}_M\mathbf{G}^T \\ \mathbf{G}\mathbf{C}_M & \mathbf{G}\mathbf{C}_M\mathbf{F}^T & \mathbf{G}\mathbf{C}_M\mathbf{G}^T \end{bmatrix}. \tag{2.13}$$

Using Eq. 2.8, the EnKF analysis step for a linear problem can be written as (Chen, 2008)

$$\begin{aligned}\overline{\mathbf{y}^u} &= \begin{bmatrix} \overline{\mathbf{m}^u} \\ \mathbf{F}\ \overline{\mathbf{m}^u} \\ \mathbf{G}\ \overline{\mathbf{m}^u} \end{bmatrix} \\ &= \begin{bmatrix} \overline{\mathbf{m}^f} \\ \mathbf{F}\ \overline{\mathbf{m}^f} \\ \mathbf{G}\ \overline{\mathbf{m}^f} \end{bmatrix} + \begin{bmatrix} \mathbf{C}_M\mathbf{G}^T \\ \mathbf{F}\mathbf{C}_M\mathbf{G}^T \\ \mathbf{G}\mathbf{C}_M\mathbf{G}^T \end{bmatrix} (\mathbf{G}\mathbf{C}_M\mathbf{G}^T + \mathbf{C}_D)^{-1}(\mathbf{d}_{\text{obs}} - \mathbf{G}\overline{\mathbf{m}^f}). \end{aligned} \quad (2.14)$$

Eq. 2.14 can be reduced to

$$\overline{\mathbf{m}^u} = \overline{\mathbf{m}^f} + \mathbf{C}_M\mathbf{G}^T(\mathbf{G}\mathbf{C}_M\mathbf{G}^T + \mathbf{C}_D)^{-1}(\mathbf{d}_{\text{obs}} - \mathbf{G}\overline{\mathbf{m}^f}). \quad (2.15)$$

For a linear problem represented by Eq. 2.10, Eq. 2.14 states that, the dynamic state variables ($\mathbf{u}$) and the simulated observations ($\mathbf{d}$) obtained from the EnKF analysis step will be equivalent to those that could be obtained by rerunning the simulation model from the initial time using the updated static model parameters ($\mathbf{m}$). For nonlinear problems like those encountered in the reservoir engineering field, however, this relationship does not hold true and the updated dynamic state variables (e.g. pressures, phase saturations, etc.) are often found to be inconsistent with those could be obtained by rerunning the simulation model using the updated static model parameters (e.g. porosity, permeability, etc.). The most common evidence of this inconsistency is the violation of the inequality constraints on the plausible values of state variables and the occurrence of non-physical state variables upon data assimilation. These non-physical state variables can cause unwanted results during the forecast step of the EnKF. When the model is not strongly nonlinear, the inconsistency is not strong and heuristic methods like truncation can be used to handle the constraint violations. When the nonlinearity is strong, however, heuristic method is not an appropriate solution.

The EnKF relies only on the first two moments, the mean and the covariance, of the state vector. The Gaussian probability distribution is the only distribution which can be completely characterized using the first two moments, the mean and the variance or the covariance. Hence, the EnKF works well and shows a good match to the observations when the prior probability distributions of the static model parameters and the dynamic state variables are approximately Gaussian. The violation of the Gaussianity assumption may result in large adjustments to variables in the state vector and can introduce severe problems when using the EnKF for data assimilation in multiphase flow reservoir problems. Chen (2008) pointed out two reasons for the occurrence of the non-Gaussian probability distributions of state variables: (1) the prior probability distributions of the state variables are non-Gaussian and (2) nonlinearity present in the dynamic model (e.g. reservoir simulator) which can result in non-Gaussian distributions even when the prior probability distributions of the state variables are perfectly Gaussian.

A familiar example of the non-Gaussian behavior of the state variables is often seen in case of a two-phase displacement process where the saturation of the displacing phase (e.g. water) is high behind the saturation front and is low ahead of the front. Chen et al. (2009) showed that, in a water flood problem, the water saturation near the flood front is nearly discontinuous and its probability distribution in the front area is bimodal which cannot be accurately approximated by a Gaussian distribution. During the forecast step of the EnKF for a water flood problem, the reservoir simulator is able to honor and propagate the non-Gaussian probability distribution of the water saturation. The analysis step of the EnKF which is based on the first two moments, however, is not capable of handling the bimodal probability distribution and often results in saturations that violate the inequality constraints on their plausible values leading to non-physical saturations. It has been demonstrated that, because of the nonlinear/non-Gaussian effects, the analysis step of the EnKF

can make large adjustments to the water saturations and may result in non-physical water saturation values when the EnKF is applied for history matching of facies distribution and production data (Agbalaka and Oliver, 2008). A reservoir with fluvial structure consists of high-permeability channels in a background of low-permeability region and the location of these high-permeability channels is often uncertain. During the water flooding operation, the injected water, as expected, rapidly flows through the highly permeable channels with a small fraction of water advancing into the low-perm region. Such uneven advancement of water flood front through the high- and low-perm regions introduces the non-Gaussian effect. As Zhao et al. (2008) point out, the analysis step of the standard implementation of the EnKF can result in smeared saturations where water is not confined to the channels.

### 2.3.5 Constraint violations

Most optimization problems, including the problem of estimating model and state variables using the EnKF, come with a set of feasibility constraints on the elements of the optimal solution. In the EnKF framework, the static model parameters and the dynamic state variables are often subject to physical laws or constraints defined by the maximum and/or the minimum value they can take. Even though the EnKF has been highly successful for data assimilation, the underlying mathematical framework does not take into account the physical constraints on state variables during the data assimilation step. Highly nonlinear model dynamics, non-Gaussian state variables, and small ensemble size are some of the key elements which can result in non-physical updates during the data assimilation step of the standard EnKF. The computational cost grows with increase in the ensemble size. Therefore, for all practical purposes, it is required to use as small an ensemble size as possible. The standard implementation of the EnKF, however, suffers from a poor approximation to the covariance matrix when the ensemble size is small. The poor approximation of the covariance matrix

can result in non-physical updates to the state variables.

The widely studied example of the EnKF resulting in non-physical updates in case of reservoir data assimilation problem is that the updated saturations, when included directly into the state vector, may take values which are negative or are greater than unity. Gu and Oliver (2007) showed examples in which this problem occurred. The authors showed that the EnKF in its standard form failed to handle a highly nonlinear model (1D water flood) involving non-Gaussian state variable (water saturation) resulting in non-physical water saturations during data assimilation. A similar problem was demonstrated by Chen et al. (2009) where the authors showed that the standard implementation of the EnKF failed to handle the bimodal probability distribution of water saturation in the front area in case of a water flooding example. Mandel et al. (2004) also reported a problem of updated state variables taking non-physical values when the EnKF was applied to a simple wildfire model. The authors observed that, application of the EnKF resulted in non-physical changes to the temperatures in some locations, which caused a complete breakdown of the simulations in subsequent advancements.

In case of history matching compositional models using the EnKF, a similar problem occurs when the molar densities of components, instead of phase saturations, are included in the state vector. In Chapter 3, we illustrate an application of the standard implementation of the EnKF to a compositional model in which the updated molar density of $CO_2$ in some regions is observed to take negative values while molar densities of the remaining components are increased. Standard truncation schemes avoid negative values of molar densities, but do not address the problem of increased molar densities of other components. The results can include a spurious increase in reservoir pressure with a subsequent inability to maintain injection. Chapter 3 includes the reservoir description of a compositional model and discusses the results obtained from the application of the standard EnKF to it.

## 2.4 Summary

The ensemble Kalman filter (EnKF) was developed to allow data assimilation with nonlinear models consisting of large numbers of state variables. The EnKF uses an ensemble of model realizations for propagating forward in time the mean state and the associated error covariance matrix. The EnKF, however, is only sub-optimal for nonlinear and non-Gaussian problems due to the linear update at the analysis step. In reservoir engineering related applications, highly nonlinear model dynamics and non-Gaussian state variables are often encountered which present a significant challenge for data assimilation when the standard implementation of the EnKF is used.

A brief review of the EnKF technique and its application to petroleum engineering related applications is presented. The review consists of the progress in terms of the application of the EnKF method to reservoir engineering history matching problems with increasing dimensions and estimation of different types of model parameters. Some of the important studies including the first application of the EnKF to petroleum engineering history matching problem and to reservoir monitoring, and the application of the EnKF to different pseudo-field and real-field cases are discussed in the review. The methodology for the implementation of the standard EnKF technique which include parameterization, generation of the initial ensemble, and the forecast and the analysis steps are discussed. Some of the important issues that arise during the application of the EnKF to reservoir engineering related history matching problems are briefly mentioned. These issues include model error, sampling error, localization, assimilation of large numbers of observations, nonlinear model dynamics and non-Gaussian state variables, and constraint violations on the plausible values of the state variables.

# CHAPTER III

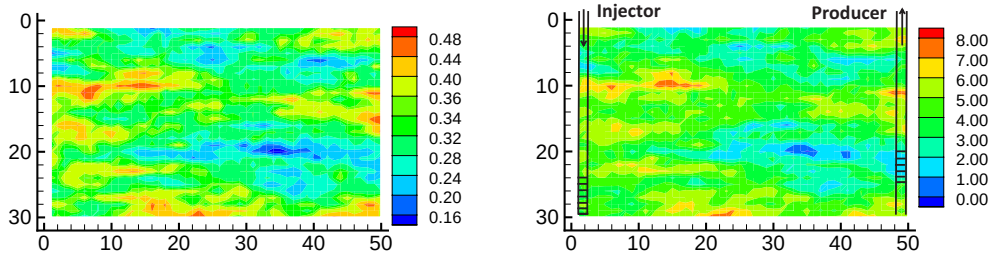
# HISTORY MATCHING OF A COMPOSITIONAL MODEL USING THE STANDARD ENKF

In this chapter, we describe the application of the standard implementation of the EnKF to a data assimilation problem of $CO_2$ injection into a compositional reservoir model with 12 components, 11 in the hydrocarbon phase. A detailed discussion about the reservoir model is given below.

## 3.1 Reservoir model description

The synthetic 2-dimensional ($x$-$z$ cross-section) compositional reservoir model has immobile water saturation with no initial gas phase. The 11 components in the oil phase range from $C_1$ to $C_{7+}$ along with $CO_2$ and $N_2$. There is one producer and one injector in the model. The producer is located on the east boundary of the model and is operated with a bottom hole pressure (BHP) constraint of 1600 psia. The $CO_2$ injector is located on the west boundary of the model with a BHP constraint of 2200 psia. The reservoir model is divided into a uniform grid of dimensions 50 $\times$ 1 $\times$ 30. Grid block dimensions are 50 ft $\times$ 30 ft $\times$ 20 ft. The porosity and log-permeability are realizations of a correlated anisotropic Gaussian random field. The mean porosity is 0.32 with standard deviation of 0.06. The mean log-permeability is 4.0 with standard deviation of 1.4. The log-permeability field is correlated to the porosity field with a correlation coefficient of 0.90. Figs. 3.1(a) and 3.1(b) show the porosity and log-permeability fields for the reference case, respectively. Completion

locations for the injector and the producer are also shown in Fig. 3.1(b).




(a) Reference: porosity (b) Reference: log-permeability

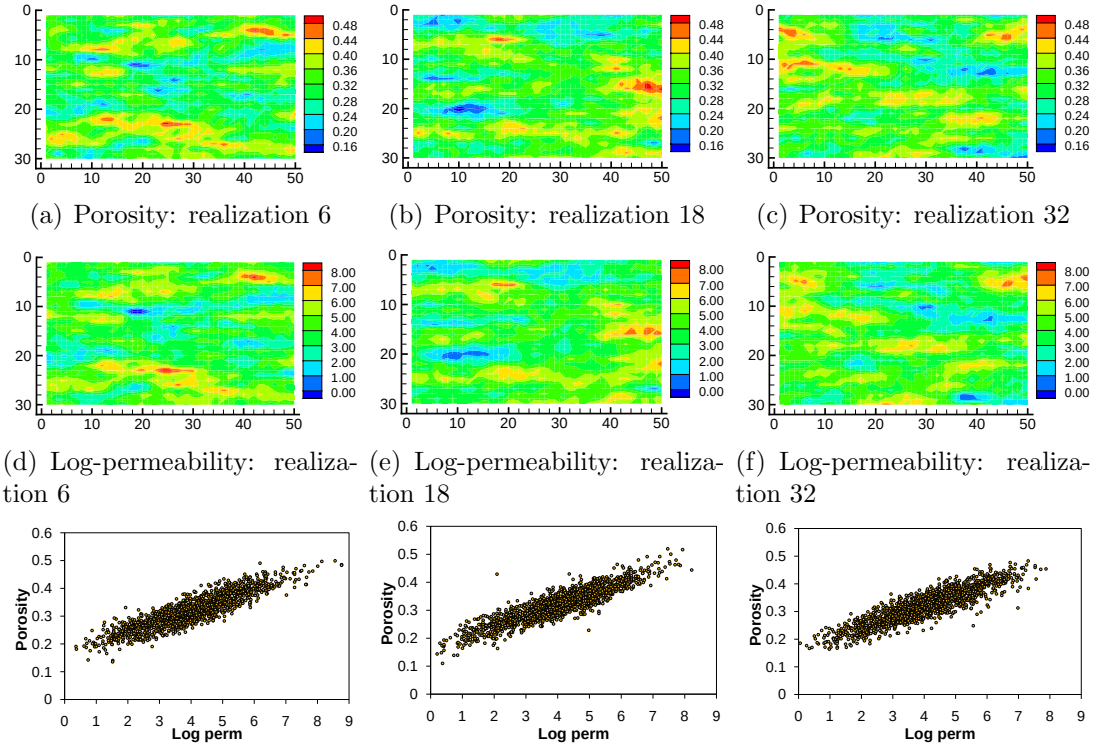
**Figure 3.1:** The reference porosity and log-permeability fields for the compositional model.

## 3.2 Parameters for the standard EnKF

### 3.2.1 Initial ensemble

Forty initial realizations of porosity and log-permeability are generated using the same geostatistical parameters as the reference model. None of the realizations are conditioned to the static well data. Fig. 3.2 shows the initial realizations of porosity and log-permeability for different ensemble members and the cross-plots between porosity and log-permeability for each of these realizations. The variability among the log-permeability fields of different ensemble members is sufficiently high which is important for obtaining different flow behavior (e.g. early or late $CO_2$ breakthrough at the producer) during the history matching period. A commercial compositional reservoir simulator (Schlumberger, 2006) is used for the forward modeling of flow and transport. Measurements used in the data assimilation include oil production rate, producing gas-oil ratio, $CO_2$ injection rate, gas-phase $CO_2$ mole fraction at the producer, and liquid-phase $CO_2$ mole fraction at the producer. The state vector for model updating includes the molar densities of 12 components from the reservoir fluid system and gridblock-based variables including pressure, porosity, and log-permeability. Thus, there are a total of 15 variables in each gridblock.

Fig. 3.3 shows the production data obtained from the initial ensemble together




(a) Porosity: realization 6 (b) Porosity: realization 18 (c) Porosity: realization 32

(d) Log-permeability: realization 6 (e) Log-permeability: realization 18 (f) Log-permeability: realization 32

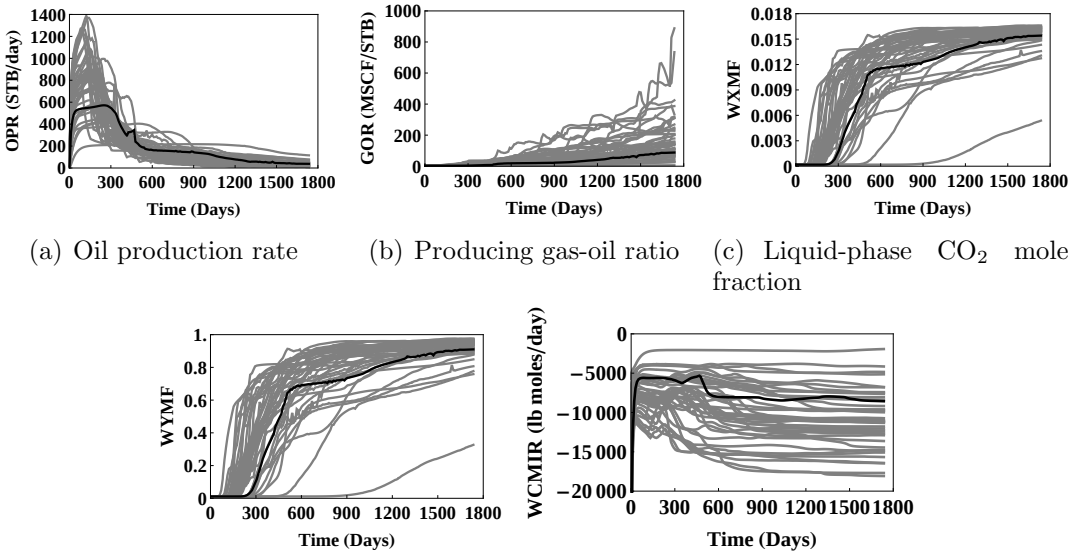
(g) Cross-plot: realization 6 (h) Cross-plot: realization 18 (i) Cross-plot: realization 32

**Figure 3.2:** The model parameters for different realizations from the initial ensemble.

**Table 3.1:** Compositional model - standard deviation values for different types of production data

| **Type of observation** | **Standard deviation** |
|---|---|
| Oil production rate | 10.0 STB/day |
| $CO_2$ injection rate | 20.0 lb moles/day |
| Producing gas-oil ratio | 0.5 MSCF/STB |
| Producing gas-phase $CO_2$ mole fraction | 0.0005 |
| producing liquid-phase $CO_2$ mole fraction | 0.00005 |

with the reference production data. The gray curves indicate the ensemble forecasts and the black curve indicates the production data from the reference model. A significant fraction of the initial ensemble members have high oil production rates and early $CO_2$ breakthrough compared to the reference model. The total simulated production history time is 1106 days and production data are assimilated approximately every 100 days. The production data are assimilated 11 times during the history matching period. The noise associated with the production data is sampled from a normal distribution with a mean of 0 and standard deviation which varies by type of observation. The same initial realizations and the same noisy observations are used in performing the standard EnKF and the CEnKF (CEnKF is discussed in Chapter 4). The standard deviation values used for different types of production data are given in Table 3.1.




(a) Oil production rate

(b) Producing gas-oil ratio

(c) Liquid-phase $CO_2$ mole fraction

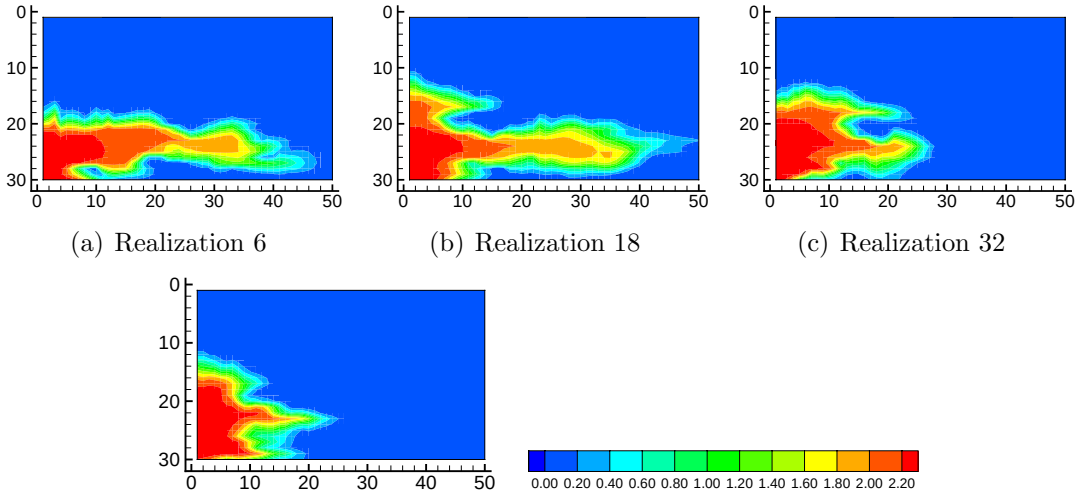
(d) Gas-phase $CO_2$ mole fraction

(e) $CO_2$ injection rate

**Figure 3.3:** The production forecasts from the initial ensemble (gray curves) and the reference production data (black curve).

### 3.2.2 Additional observation of $CO_2$ molar density

An additional observation of $CO_2$ molar density at gridblock (16,22), located in the layer 22 of the 2D reservoir model and at the 16th gridblock to the right from the west boundary of the model, is assimilated at the first data assimilation time. Although the additional $CO_2$ molar density measurement in the reservoir interior is unlikely to occur in practice, it is selected because it significantly increases the likelihood of state constraint violations in the interior. The initial ensemble of porosity and log-permeability are run from the initial time to the first data assimilation time (day 102) using the compositional module of the reservoir simulator. Fig. 3.4 shows the distribution of $CO_2$ throughout the reservoir at day 102 before data assimilation for different ensemble members and the reference model. Because of the relatively large degree of heterogeneity in permeability, the molar densities are highly variable. The movement of $CO_2$ flood front inside the reservoir for the reference model is shown in Fig. 3.4(d) which indicates that the $CO_2$ front has not moved a long way into the reservoir at this time.



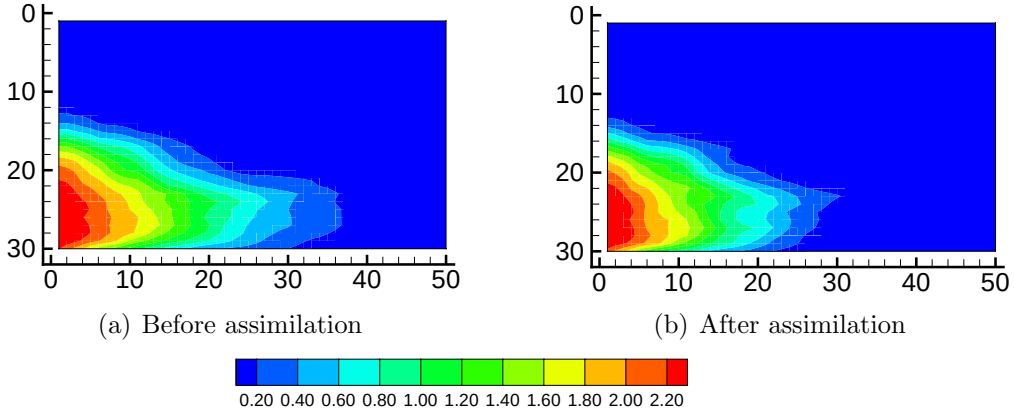

(a) Realization 6 (b) Realization 18 (c) Realization 32

(d) Reference (e) Scale

**Figure 3.4:** Forecast $CO_2$ molar density distributions at day 102 for different ensemble members.

The ensemble mean and the standard deviation of $CO_2$ molar density at day

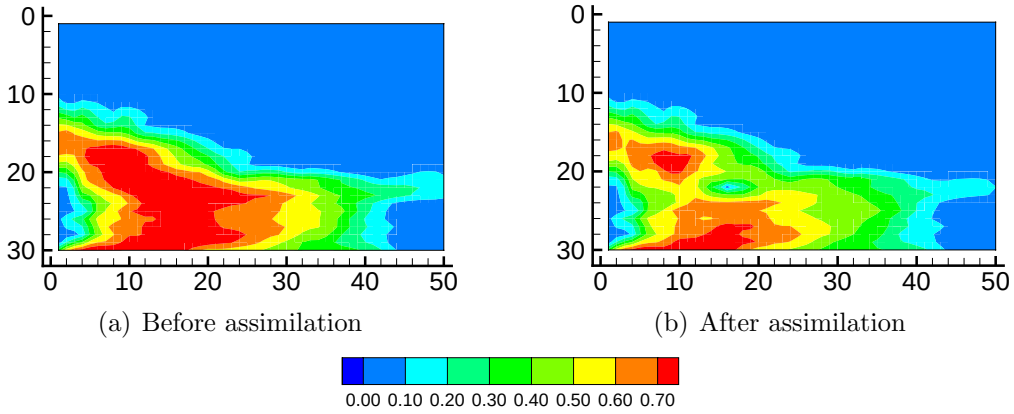
102 prior to data assimilation are shown in Figs. 3.5(a) and 3.6(a), respectively. The forecast ensemble mean and the standard deviation around the measurement location, gridblock (16,22), are close to 1.5 and 0.70, respectively. Thus, the region around the measurement location has fairly large variability in terms of $CO_2$ molar density. The standard EnKF method is applied for assimilating production data and the additional observation of $CO_2$ molar density. The ensemble mean and the standard deviation of $CO_2$ molar density at day 102 after data assimilation are shown in Figs. 3.5(b) and 3.6(b), respectively. After the analysis step, the variability in terms of $CO_2$ molar density at the gridblock (16,22) is reduced approximately to 0 as seen in Fig. 3.6(b).




(a) Before assimilation (b) After assimilation

(c) Scale

**Figure 3.5:** The ensemble mean of $CO_2$ molar density at day 102.



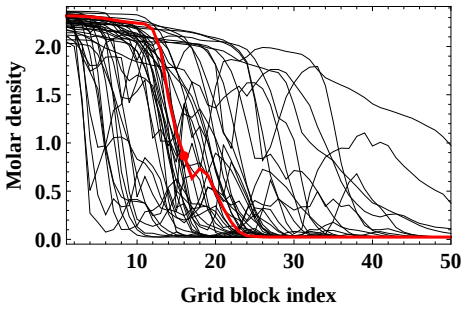

(a) Before assimilation (b) After assimilation

(c) Scale

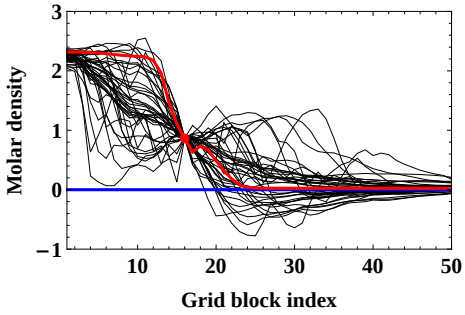
**Figure 3.6:** The ensemble standard deviation of $CO_2$ molar density at day 102.

## 3.3 Results from the standard EnKF

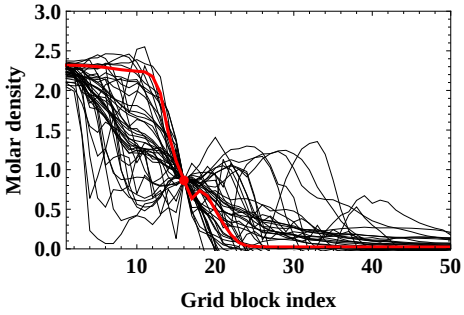
At the first data assimilation time (day 102), in addition to the observations of production data, we assume that the $CO_2$ molar density is measured at one location in the reservoir at that time. Fig. 3.7 shows the $CO_2$ molar densities for layer 22 of the reservoir model before and after data assimilation. The black curves represent the ensemble members, the thick red curve is the reference model, and the red dot denotes $CO_2$ molar density measurement. For most of the ensemble members, the updated $CO_2$ molar density from the analysis step in the EnKF (Fig. 3.7(b)) is negative at some locations.




(a) Forecast $CO_2$ molar density




(b) Updated $CO_2$ molar density (without truncation)




(c) Updated $CO_2$ molar density (with truncation)

**Figure 3.7:** The prior and posterior ensemble of $CO_2$ molar density for layer 22 at the first data assimilation time (day 102). The black curves are the ensemble members, the thick red curve is the reference, and the red dot represents $CO_2$ molar density measurement.

The most common method for dealing with negative saturations or molar densities in the EnKF is to truncate the state variables that violate constraints to the value

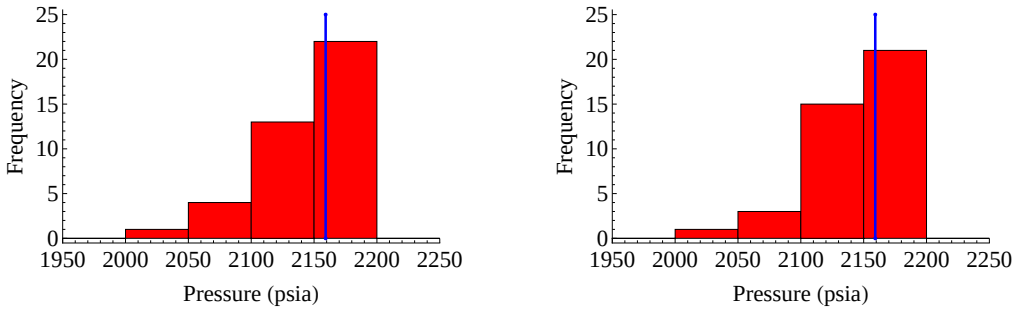
at the boundary of the feasible region. Molar density cannot be negative, so all $CO_2$ molar densities in the model that are less than 0 would be set equal to 0 at the end of the analysis step. When that approach is taken in this case, the result is an inability to inject fluid into the reservoir.

The injector and the producer, both are controlled by the bottom hole pressure (BHP) as the primary constraint. The injector, which is completed in the gridblocks (1,24) through (1,29), is operated at a BHP of 2200 psia. Upon data assimilation, a substantial increase in the gridblock pressure around an injector, exceeding the maximum allowable BHP (2200 psia), can cause an injector to shut in during the subsequent forecast step. The distribution of pressure data at day 102, before and after data assimilation, for several gridblocks immediately adjacent to an injector are analyzed and are shown in Fig. 3.8. The blue straight lines on the histograms indicate the corresponding gridblock pressure values for the reference model. For all the ensemble members, the pressure (before and after data assimilation) at gridblocks adjacent to an injector lies within the limit of the maximum allowable BHP for an injector. This analysis indicates that the updated pressure in the adjacent gridblocks to an injector should not result in an inability to maintain fluid injection into the reservoir.

The effect of the truncation approach during data assimilation on the reservoir behavior can easily be understood through the investigation of one particular realization. We will examine the effect of the use of the standard EnKF update with truncation on the reservoir behavior of realization 31.

### 3.3.1 Analysis of realization 31

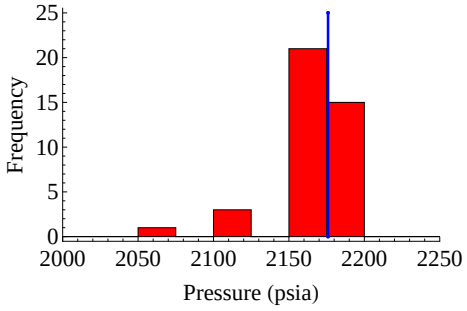
At the end of the first data assimilation step, the values of all state variables are updated, truncated if necessary, and returned to the reservoir simulator. Fig. 3.9



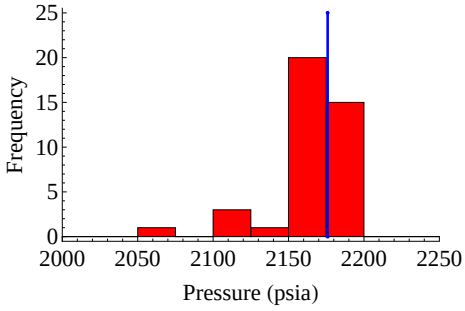

(a) Forecast: gridblock (2,24)



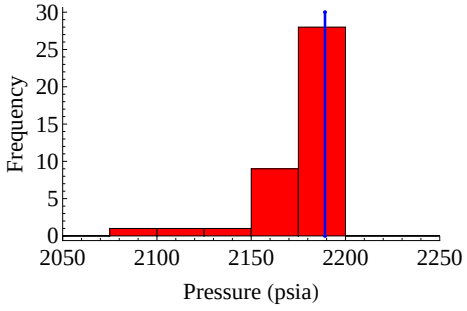

(b) Updated: gridblock (2,24)



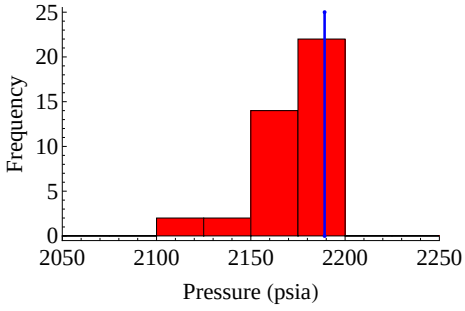

(c) Forecast: gridblock (2,26)



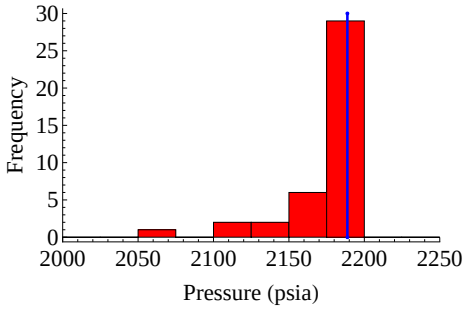

(d) Updated: gridblock (2,26)



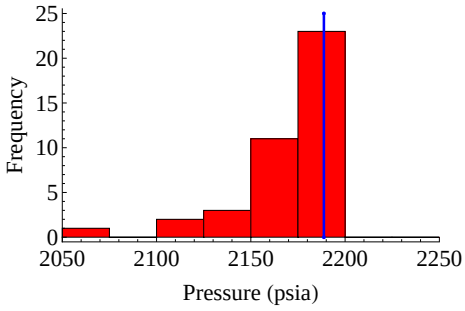

(e) Forecast: gridblock (2,28)




(f) Updated: gridblock (2,28)



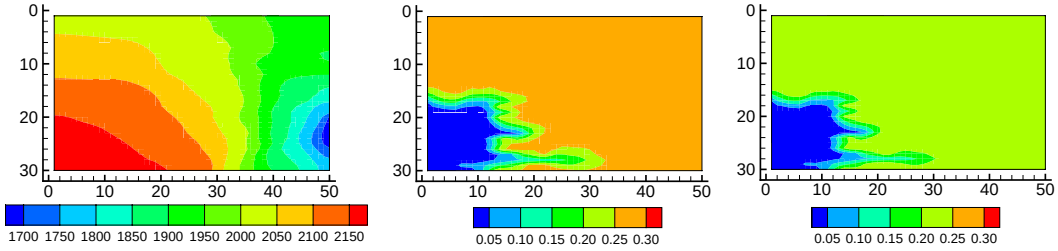

(g) Forecast: gridblock (2,29)




(h) Updated: gridblock (2,29)

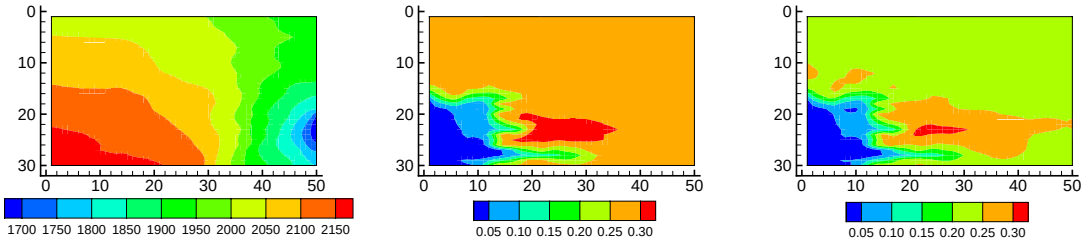
**Figure 3.8:** The distribution (from the ensemble) of forecast and updated pressure data at day 102 for different gridblocks adjacent to an injector. The blue line indicates the gridblock pressure for the reference model.

shows the forecast pressure and molar densities of two of the components. The updated pressure and molar densities of these two components obtained at the first data assimilation time step are shown in Fig. 3.10. The pressure after updating is quite reasonable with high pressure on the left and low pressure on the right near the producing completions. The molar densities of $NC_4$ and $C_7$ prior to data assimilation at the first data assimilation step (Figs. 3.9(b) and 3.9(c)) are also reasonable. The updated molar densities of $NC_4$ (Fig. 3.10(b)) and $C_7$ (Fig. 3.10(c)) are somewhat odd, however, as they show anomalously high densities to the right of the measurement location. These densities are higher than at any other location in the reservoir.




(a) Pressure (b) $NC_4$ molar density (c) $C_7$ molar density

**Figure 3.9:** Realization 31 - pressure and molar density distributions in the reservoir before the first data assimilation step (day 102).

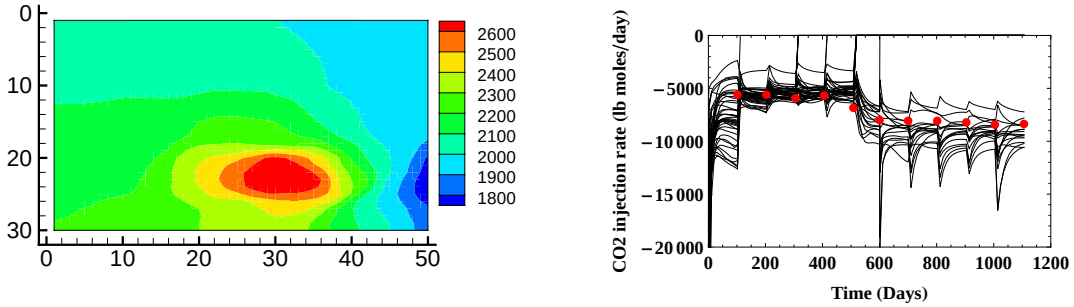



(a) Pressure (b) $NC_4$ molar density (c) $C_7$ molar density

**Figure 3.10:** Realization 31 - pressure and molar density distributions in the reservoir after the first data assimilation step (day 102).

The model is restarted using all the static and dynamic variables updated at the first data assimilation time step (day 102). At day 102.5, a region of high pressure can be seen between the injector and the producer (Fig. 3.11(a)). The magnitude of the maximum pressure in this region is about 2700 psia which is substantially higher

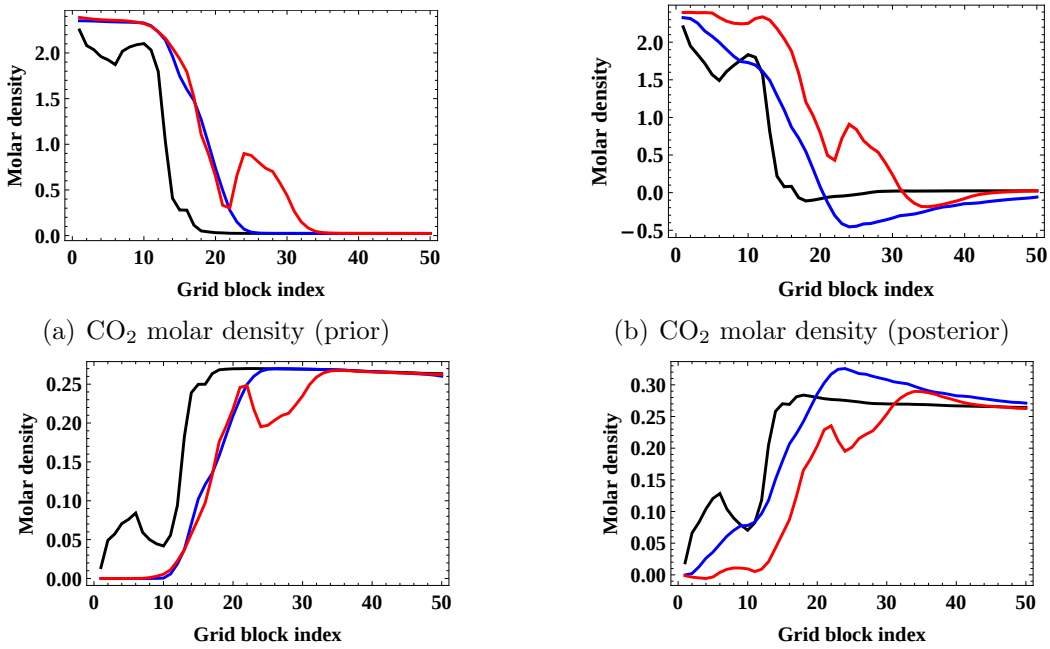
than the maximum allowable BHP for the injector (2200 psia). The anomalously high pressure due to the excess mass is significant enough that it results in an inability to inject fluid into the reservoir (Fig. 3.11(b)). Nearly 60% of the ensemble members exhibited periods during which an injector became shut-in due to excess pressure when truncation was used following the standard update. Realization 31 is selected because it is one of the more extreme examples. The regions with higher molar densities of components $NC_4$ and $C_7$ in Figs. 3.10(b) and 3.10(c) coincide closely with the region of pressure build-up shown in Fig. 3.11(a). Similar behavior of increased molar density occurs for all other components from the oil phase with the exception of $CO_2$. Fig. 3.12 shows the prior and the posterior molar densities of two of the components, $CO_2$ and $NC_4$, for different layers of the simulation model at the first data assimilation time step (day 102). The updated $CO_2$ molar density (Fig. 3.12(b)) at some locations in the reservoir takes non-physical value. The updated molar density of $NC_4$ (Fig. 3.12(d)) at these locations appears to be increased. The results of the production forecasts during the course of data assimilation and the predictions using the updated estimates of porosity and permeability from the standard EnKF are discussed in Chapter 4 and are compared with the results obtained from the CEnKF.




(a) Pressure at day 102.53 (immediately after restart)

(b) $CO_2$ injection rate

**Figure 3.11:** Pressure in realization 31 at the end of the first forecast time step after data assimilation (left) and the ensemble of $CO_2$ injection rates during data assimilation for the standard EnKF (right).




(a) $CO_2$ molar density (prior)

(b) $CO_2$ molar density (posterior)

(c) $NC_4$ molar density (prior)

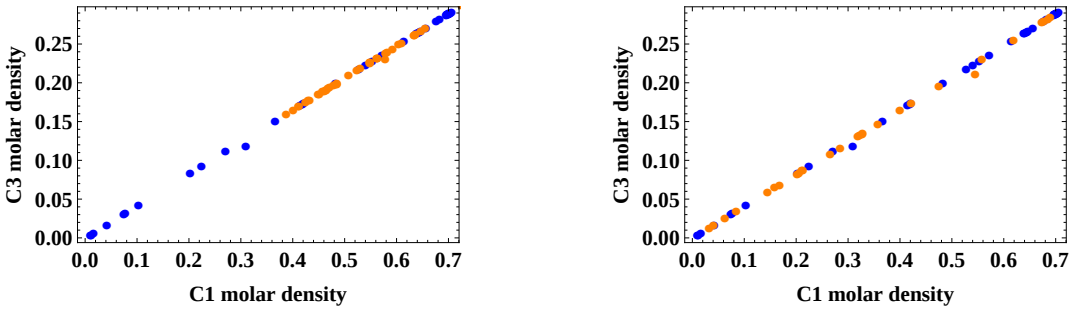
(d) $NC_4$ molar density (posterior)

**Figure 3.12:** Realization 31 - the prior and posterior molar densities of $CO_2$ and $NC_4$ at the first data assimilation time for different layers of the simulation model: layer 18 (black curve), layer 22 (blue curve), and layer 27 (red curve).

### 3.3.2 Transformation of variables

An alternate approach that avoids the need for truncation or for applying inequality constraints is to transform the variables such that the back-transformation always results in values within the feasible region. Gu and Oliver (2006) applied a "normal score transform" to saturation for this purpose, with marginal results. While the updated saturations correctly honored the inequality constraints, the spatial distributions were sometimes highly oscillatory. Simon and Bertino (2009) also reported the normal score transformation of state variables to be an effective solution in an oceanographic application. In the case of compositional simulation, however, there is a more compelling reason not to use transformed variables: the updated fluid compositions may be significantly changed as a result of updating in the transformed system. Here we illustrate using the same example that has been discussed earlier, but using the logarithm of molar densities in the state vector. We first note that

the molar densities of the various components are highly correlated in the forecast ensemble as shown by the blue dots in Figs. 3.13 and 3.14. These molar densities correspond to a location several gridblocks away from the location at which the molar density of $CO_2$ is observed. When the standard variables (molar densities) are used in the state vector, the correlation between molar densities is well-preserved after the update (orange dots in Figs. 3.13(a) and 3.14(a)). Although the compositions of components $C_3$ and $C_1$ (both present in the hydrocarbon phase at the initial time) appear to show a good correlation after updating using the transformed state variables (Fig. 3.13(b)), the consistency of the fluid composition is completely destroyed after updating as shown by the orange dots in Fig. 3.14(b). While there are many examples where transformation of the variables might be the appropriate method for preventing non-physical updates, it does not appear to be an appropriate approach for compositional state variables.

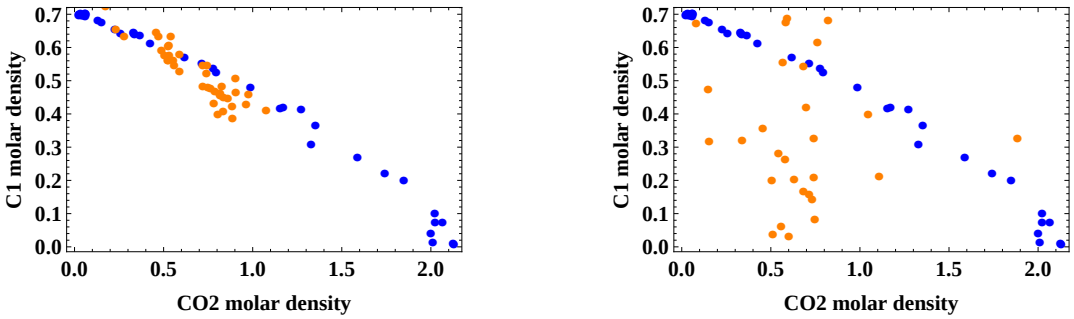



(a) Without transformation

(b) With transformation

**Figure 3.13:** The cross-plot of $C_3$ and $C_1$ molar densities at location (18,22) obtained from the standard EnKF without transformation and with transformation at the first data assimilation time step. The forecast and the updated molar densities are shown in the blue and the orange colors, respectively.

Although the transformation of state variables approach used with the compositional model ensures that the updated state variables honor the inequality constraints, it results in a similar issue of the increased molar densities of components, which has been discussed earlier in the case of the application of the standard EnKF without transformation. Fig. 3.15 shows the updated pressure and molar densities of two




(a) Without transformation

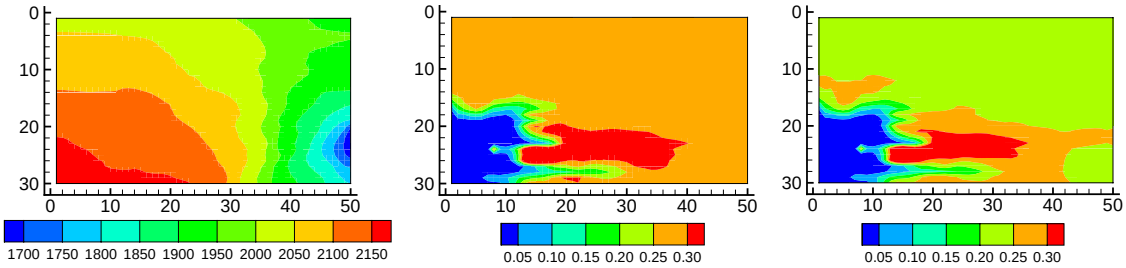
(b) With transformation

**Figure 3.14:** The cross-plot of $C_1$ and $CO_2$ molar densities at location (18,22) obtained from the standard EnKF without transformation and with transformation at the first data assimilation time step. The forecast and the updated molar densities are shown in the blue and the orange colors, respectively.

of the components obtained from the application of the standard EnKF with transformation. The updated molar densities (Figs. 3.15(b) and 3.15(c)) show similar behavior with those obtained during the application of the standard EnKF without transformation (Figs. 3.10(b) and 3.10(c)). The issue of the increased molar densities, however, appears to be more severe in the case of using the log-transformed molar densities where the region with the increased molar densities is greater than in the case where molar densities alone are included in the state vector. Although it has not been shown here, the excess mass, in the case of the standard EnKF with transformation, will result in the anomalously high pressure in the reservoir causing an injector to shut in. It is clear that, the transformation of state variables method prevents the non-physical updates to state variables, but it is not able to eliminate the occurrence of the excess mass after data assimilation.

## 3.4 Summary

The application of the standard implementation of the EnKF to a data assimilation problem of $CO_2$ injection into a compositional reservoir model is described. Along with the production data, an additional observation of $CO_2$ molar density measured at a particular location in the reservoir was used during the first data assimilation time




(a) Pressure (b) $NC_4$ molar density (c) $C_7$ molar density

**Figure 3.15:** Realization 31 - pressure and molar density distributions in the reservoir after the first data assimilation step (standard EnKF with transformation).

step. We have shown that, upon data assimilation, the standard implementation of the EnKF resulted in non-physical (or negative) $CO_2$ molar densities and the molar densities of the rest of the components were increased at some locations in the reservoir. When a truncation scheme was used to truncate the state variables that violate constraints, to the value at the boundary of the feasible region, the negative molar densities of $CO_2$ were set equal to 0 at the end of the analysis step. This truncation approach, however, did not affect the increased molar densities of the rest of the components, resulting in an inability to maintain fluid injection into the reservoir. Transformation of state variables is an alternate approach which can be used to replace truncation or the need for applying inequality constraints. In the case of the compositional model, however, we have demonstrated that, the transformation of variables approach not only destroyed the consistency of the fluid composition but it did not prevent the molar densities of the other components (except $CO_2$) from increasing after data assimilation. The standard EnKF coupled with either truncation or transformation of state variables was not successful in maintaining a continuous fluid injection into the reservoir.

# CHAPTER IV

# CONSTRAINED ENSEMBLE KALMAN FILTER

## 4.1 Introduction

One of the attractive features of the EnKF is its ability to update both the static model parameters and the dynamic state variables given the observations. These model parameters and state variables are usually subject to physical laws or constraints defined by the maximum and/or the minimum value they can take. Most optimization problems, including the problem of estimating model and state variables using the EnKF, come with a set of feasibility constraints on the elements of the optimal solution. In the EnKF framework as well as in any other data assimilation method, the state variables of a realization conditioned to the measurements should honor the constraints on their plausible values. Finding a solution that minimizes (or maximizes) the given objective function without honoring constraints (unconstrained optimization) may not be difficult, but finding a solution that solves the optimization problem and simultaneously satisfies constraints on each of its elements (constrained optimization) is often a challenging task. Although the EnKF has been highly successful for data assimilation, the underlying mathematical framework does not take into account the physical constraints on state variables during the data assimilation step. The physical constraints are often ignored during updating, and are handled post-analysis in a rather heuristic manner. Thus, the standard implementation of the EnKF can be viewed as an unconstrained optimization problem. Often times, highly nonlinear and non-Gaussian processes are encountered in dynamic estimation problems. Under these conditions and in the absence of any constraint-enforcing

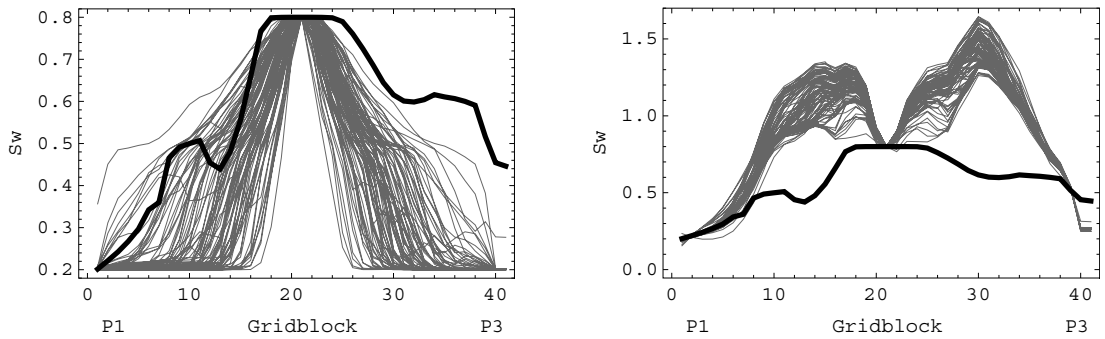
[1]Much of the material in this chapter has been accepted for publication in the SPE Journal.

mechanism, the solution (or part of it) obtained from the standard implementation of the EnKF often violates the constraints on its elements and may result in non-physical values of model parameters and/or state variables.

A number of solutions have been proposed in the past including transformation of state variables, reparameterization, and truncation to avoid the issue of non-physical updates. Gu and Oliver (2006) proposed three different methods for preventing non-physical water saturations: (1) normal score transformation of the saturation variable, (2) use of the location of water shock front instead of saturations as a state variable, and (3) use of an iterative filter whenever the updated saturations are found to exceed the physical bounds. A normal score transformation of state variables which transforms the constrained variables to unconstrained variables was reported to be an effective solution in an oceanographic application by Simon and Bertino (2009). Gu and Oliver (2006) also reported partial success with using normal score transform of saturation for updating states in the water flood problem. This approach resulted in updated saturation values within plausible limits, however, it did not prevent non-physical spatial oscillations in saturation. Although the transformation of state variables approach ensures the updated state variables to honor the inequality constraints, as discussed in Chapter 3, it might not be an attractive solution in the case of a compositional simulation. On the other hand, Gu and Oliver (2006) did report good success with the use of the location of the saturation front as a state variable, but the method's effectiveness seemed to be partly a result of the 1D geometry of the flow. Gu and Oliver (2006) concluded that the most generally successful methods are iterative forms of the EnKF in which the Kalman gain is used only to update model variables and the updated state variables, including saturations, are obtained by rerunning the simulator from the initial conditions (Gu and Oliver, 2007; Li and Reynolds, 2009b). As the state variables are obtained using the simulator, they naturally satisfy the inequality constraints. The iterative forms of the EnKF

present an effective way of mitigating the issue of constraint violations. However, the effectiveness comes at a price of substantial computational efforts required for rerunning the simulator.

The reparameterization technique has also been investigated in order to avoid non-physical updates to state variables. When the water cut data at producers were assimilated in a 2D water injection problem, Chen et al. (2009) showed that it resulted in constraint violations on the plausible values of water saturation as shown in Fig. 4.1. In order to resolve this issue, Chen et al. (2009) included the time of saturation arrival, instead of saturations, as a state variable in the state vector. The Gaussian assumption worked well with the time of saturation arrival and resulted in improving the quality of the estimates compared to the standard implementation of the EnKF. However, it required substantial additional computations beyond that required by the EnKF. Even though transformation and reparameterization techniques have been used, truncation of state variables where the non-physical estimates are truncated to their limiting values remains the widely used method for dealing with the issue. The truncation approach works fairly well in most of the situations, however, as we showed in Chapter 3, it can fail at times resulting in significant problems during the forecast step of the EnKF.




(a) Forecast $S_w$ at second assimilation

(b) Updated $S_w$ at second assimilation

**Figure 4.1:** The saturation profiles on the diagonal of producers P1 and P3 from the standard EnKF. The thin curves are the ensemble solutions and the thick black curve is the reference (after Chen et al. (2009)).

A number of attempts have been made in the past for incorporating the constraints on the state variables into the data assimilation framework. The equality constraints have been addressed by a number of authors in the context of the Kalman filter. Qureshi et al. (1989) considered the constraints on signal amplitude and signal local variance in the image restoration process. The deterministic signal constraints were transformed into constraints on the Kalman gain. Qureshi et al. (1989) solved the resulting constrained nonlinear optimization of the Kalman gain using a penalty function approach. De Geeter et al. (1997) presented smoothly constrained Kalman filter for incorporating nonlinear constraints. The nonlinear constraints were used as nearly perfect observations and incorporated in an iterative manner. Ko and Bitmead (2007) presented a constrained Kalman filter for dealing with the state estimation problem for linear systems with linear state equality constraints. Simon and Chia (2002) used the projection of the unconstrained Kalman filter solution onto the state constraint surface in order to incorporate state equality constraints in the Kalman filter framework. Pan and Wood (2006) proposed a two-step procedure for incorporating equality constraints in the data assimilation process using the EnKF. Even though the constrained data assimilation approach can successfully be used for an equality constrained problem, it does not address the inequality constraints which are generally more important and inherently more complicated than the equality constraints in reservoir engineering applications.

The problem of applying inequality constraints to estimation of state variables has been addressed in the context of the Kalman filter. Massicotte et al. (1995) modified the standard Kalman filter algorithm where the state variables were transformed in accordance with the inequality constraints on them, prior to assimilating data. Massicotte et al. (1995) presented the application of the modified Kalman filter algorithm to a spectrometric data correction example for preventing the negative values of the solution. Rao et al. (2003) used the moving horizon estimation (MHE) approach for

incorporating the inequality constraints on the state variables and the noise associated with the state variables and the observations. For linear systems with quadratic objective functions, the optimal solution presented by Rao et al. (2003) reflects the standard Kalman estimate for the state of the system. Gupta and Hauser (2007) presented two different methods for incorporating the inequality constraints into the Kalman filter framework. In the first method, Gupta and Hauser (2007) projected the unconstrained solution to lie in the feasible space where the active set of inequality constraints was obtained using Lagrange multipliers. In the second method, Gupta and Hauser (2007) proposed to restrict the optimal Kalman gain such that the updated state estimate lies in the constrained space. Recently, Rotea and Lana (2008) considered a minimax estimation problem where the hard constraints on the state variables and input were replaced by the probability constraints that restrict the choices of the covariance matrices for the state and input. Rotea and Lana (2008) showed that the solution to the minimax problem is a conventional Kalman filter for suitably chosen covariance matrices for the state and input. Ungarala et al. (2007) presented a formulation for incorporating inequality constraints on state variables in an extended Kalman filter technique where active constraints were identified by an exhaustive search of all combinations of state variables. While their approach was successfully applied to a nonlinear batch reactor with small number of state variables and limited number of constraints, direct application to reservoir problems with few hundred thousands of state variables and constraints is not feasible. Often times, a relatively easy approach is considered for enforcing the inequality constraints where the inequality constraints, upon violation, are converted into equality constraints and are used as perfect observations. Porrill (1988) and Hayward (1998) treated state constraints as perfect measurements. On the other hand, Thacker (2007) discussed two different approaches, Lagrange multipliers and augmented data, for incorporating inequality constraints into the Kalman filter framework. He concluded that when the

number of state variables is relatively large, it may not always be important to determine the active set of constraints precisely. Lauvernet et al. (2009) solved a similar problem by modifying the Kalman gain to explicitly account for the assumption of a truncated Gaussian probability distribution for the state variables.

Recently, Wang et al. (2010) presented accept/reject and projection methods for incorporating inequality constraints into the EnKF framework. Depending on the nature of the problem, Wang et al. (2010) proposed to sample the model error and/or observation error iteratively until the forecast and the updated states honor all inequality constraints. Although the method is easy to implement and has been successfully demonstrated, its extension to the reservoir engineering applications does not seem feasible. Even though a number of attempts have been made to incorporate the inequality constraints into the EnKF framework, most of the techniques depend on the treatment of the unconstrained solution obtained from the standard implementation of the EnKF. None of the previous studies shows the direct involvement of inequality constraints in the mathematical formulation of the EnKF.

In this chapter, we propose a method for the constrained EnKF (CEnKF) for incorporating the physical constraints on the plausible values of state variables during the data assimilation such that the resulting solution is as close as possible to the unconstrained solution obtained from the standard EnKF, and at the same time, it lies in the feasible region. The proposed method is based upon the earlier work by Thacker (2007) where he demonstrated the effectiveness of incorporating additional equality constraints during data assimilation process for a Kalman filter. The CEnKF method can be implemented in two different approaches, both of which convert inequality constraints to a small number of equality constraints. The first approach uses Lagrange multipliers to enforce the active constraints. In the second approach, the active constraints are used as virtual observations for calibrating the model parameters within plausible ranges. Applying the CEnKF technique in an iterative manner ensures that

the resulting solution is within the limits given by the constraints. The application of the proposed CEnKF method shows that the technique is able to enforce the non-negativity constraints on molar densities in the case of a 2D compositional model and the bound constraints on phase saturations (all phase saturations must be between a lower and an upper bound) in the case of a highly heterogeneous three-phase black-oil flow reservoir model. The CEnKF method is able to achieve a better estimate of reservoir properties than is obtained using only truncation with the standard EnKF.

## 4.2 Methodology

While constraints on reservoir variables can be linear or nonlinear, the kinds of constraints that are of particular interest in this paper are linear inequality constraints of the form

$$\mathbf{A}\mathbf{y} \geq \mathbf{b}$$

which are of the type satisfied by phase saturations, concentrations, pressures, and molar densities. Note that this form includes variables such as saturations with upper and lower bounds. It is not easy to determine which of the inequality constraints should be converted to equality constraints and which might be ignored when the problem has many variables. The approach we have taken is to first assimilate the data without any constraints (data assimilation using the standard implementation of the EnKF) and then check for constraint violations. In this study, we apply an iterative procedure for approximately identifying the constraints that are active in a particular data assimilation step. Once the active set is identified, equality constraints are applied to the problem in the update step. If constraints are still violated, the procedure is repeated. When it is determined that a state variable is on the boundary of the feasible region, then the inequality constraint can be replaced by equality constraint. The system of active constraints for the $j$th ensemble member can be

written in the following form,

$$\mathbf{P}_j\mathbf{y} = \mathbf{c}_j \tag{4.1}$$

where, the vector $\mathbf{c}_j$ contains the constraint values.

Thacker (2007) discussed two different approaches for incorporating equality constraints into a Kalman filter framework. The first approach uses the undetermined Lagrange multipliers and it is shown that, the resulting formulation has a form very close to the Kalman filter updating equation. The second approach incorporates the constraints along with data in an expanded data-vector and subsequently, the entire system of equations is expanded to reflect the constraints. Following is a brief description of the two approaches.

### 4.2.1 Lagrange multipliers

This approach uses the undetermined Lagrange multipliers (Thacker and Long, 1988) for incorporating equality constraints into the data assimilation formalism. While the analysis step in the standard EnKF computes variables that approximately minimize

$$J_j(\mathbf{y}) = \frac{1}{2}(\mathbf{y} - \mathbf{y}_j^f)^T(\mathbf{C}_Y^f)^{-1}(\mathbf{y} - \mathbf{y}_j^f) + \frac{1}{2}(\mathbf{H}\mathbf{y} - \mathbf{d}_{\text{obs},j})^T\mathbf{C}_D^{-1}(\mathbf{H}\mathbf{y} - \mathbf{d}_{\text{obs},j}), \tag{4.2}$$

the CEnKF will compute variables that minimize

$$L_j(\mathbf{x}, \mathbf{y}) = J_j(\mathbf{y}) - (\mathbf{P}_j\mathbf{y} - \mathbf{c}_j)^T\mathbf{x}. \tag{4.3}$$

The vector $\mathbf{x}$ is a vector of the Lagrange multipliers that enforce the constraints. The values of the Lagrange multipliers must be determined as a part of the solution.

The constrained update (which is a solution of Eq. 4.3) can be written as a correction to the unconstrained solution $\mathbf{y}^u$ of Eq. 4.2

$$\mathbf{y}_j^{u,\text{con}} = \mathbf{y}_j^u + \mathbf{C}_Y^u\mathbf{P}_j^T(\mathbf{P}_j\mathbf{C}_Y^u\mathbf{P}_j^T)^{-1}(\mathbf{c}_j - \mathbf{P}_j\mathbf{y}_j^u), \tag{4.4}$$

where, $\mathbf{y}_j^{u,\text{con}}$ and $\mathbf{y}_j^u$ are the constrained and the unconstrained updated solutions for the $j$th ensemble member, respectively; $\mathbf{C}_Y^u$ is the posterior covariance matrix; $\mathbf{P}_j$ is

the coefficient matrix for the constraint equations for the $j$th ensemble member; and $\mathbf{c}_j$ is the vector of constraint values for $j$th ensemble member.

The constrained solution (Eq. 4.4) is clearly similar in form to the unconstrained solution from the EnKF (Eq. 2.4). The largest difference is that, while in the standard implementation of the EnKF, the same Kalman gain is used to update each ensemble member, here, each ensemble member may have a different set of active constraints, so terms such as $\mathbf{P}_j$, $\mathbf{C}_Y^u\mathbf{P}_j^T$, $\mathbf{P}_j\mathbf{C}_Y^u\mathbf{P}_j^T$, $\mathbf{c}_j$, and $\mathbf{P}_j\mathbf{y}_j^u$ must be computed for each ensemble member at each data assimilation time step in order to compute the constrained solution.

### 4.2.2 Augmented data

This approach also starts by computing the unconstrained solution. Once the unconstrained solution is obtained, the next step includes checking for any constraint violations that need to be enforced. The constraint data $\mathbf{c}_j$ are then appended to the data-vector $\mathbf{d}_{\text{obs},j}$. The data-error covariance matrix $\mathbf{C}_D$ should be enlarged with zero values on the diagonal corresponding to the constraint errors, and $\mathbf{H}$ should be enlarged to reflect the constraints. With these modifications, an augmented Kalman gain that includes columns corresponding to data and active constraints can be constructed for computing the constrained updates to the original forecast $\mathbf{y}_j^f$. The modified form of Eq. 2.4 for implementing the CEnKF method can be given as

$$\mathbf{y}_j^{u,\text{con}} = \mathbf{y}_j^f + \mathbf{C}_Y^f\tilde{\mathbf{H}}^T(\tilde{\mathbf{H}}\mathbf{C}_Y^f\tilde{\mathbf{H}}^T + \tilde{\mathbf{C}}_D)^{-1}(\tilde{\mathbf{d}}_{\text{obs},j} - \tilde{\mathbf{H}}\mathbf{y}_j^f) \tag{4.5}$$

where, $\tilde{\mathbf{H}}$ is the modified form of the measurement operator which extracts the state variables that are subject to constraint enforcement along with the simulated data; $\tilde{\mathbf{C}}_D$ is the modified form of the data-error covariance matrix which also reflects the constraints; and $\tilde{\mathbf{d}}_{\text{obs},j}$ is the modified vector of the observations appended with the constraint values.

As we can notice from Eqs. 4.4 and 4.5, the primary difference between the two

approaches is in the sequence of incorporating the equality constraints. In the Lagrangian approach, the constraints are applied after assimilating data, while in the augmented data approach, the constraints are assimilated simultaneously with data. For linear observation operators and Gaussian probability densities, the two approaches are equivalent. In the examples that follow, we use the augmented data approach.

## 4.3 Algorithm description for the constrained EnKF

Following is a brief generalized algorithm for implementing the augmented data approach of the CEnKF method. This algorithm describes various steps which need to be carried out in order to enforce the constraints on any type of state variable.

1. For assimilation times $t_k$, $(k = 1, \ldots, k_{\max})$, compute updates of the ensemble members using the standard EnKF analysis formula (Eq. 2.4).

   (a) For each ensemble member $(j = 1, \ldots, N_e)$, evaluate the constraint equations $\mathbf{A}\mathbf{y}_j - \mathbf{b}$ on the updated state vector. If any values are less than $-\epsilon$ where $\epsilon$ is some small predetermined value (e.g. 0.0001 for molar densities) then proceed to the next step. Otherwise, increase $j$ by one and repeat this step.

   (b) For ensemble member $j$, identify the constraint equation $\mathbf{A}\mathbf{y}_j - \mathbf{b}$ with the largest constraint violation. That inequality constraint will be converted to an equality constraint. (Note that constraint equations may need to be normalized to make comparison of magnitudes meaningful for state variables in different units.)

   (c) Append the new constraint to the vector of actual observations (and previously appended constraints, if any) and modify the observation operator

matrix to include the equality constraint relation. Also expand the data-error covariance matrix, $\mathbf{C}_D$ for the newly added constraint.

(d) With the newly expanded matrices $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{C}}_D$, compute the modified Kalman gain for ensemble member $j$.

(e) Update the state vector for ensemble member $j$ using Eq. 4.5.

(f) If the number of iterations is less than the limit, return to 1a, otherwise apply truncation to any remaining constraint violations and return to 1 for next assimilation time.

Because our emphasis is on efficiency, we set a fairly low limit (10) on the maximum number of iterations for identification of active constraints.

We will consider four synthetic examples to demonstrate the performance of the CEnKF. The augmented data approach of the CEnKF is implemented and the results obtained from the CEnKF are compared with the results obtained from the standard EnKF. In the first example, we consider a synthetic 1D linear model to demonstrate the validity of the CEnKF technique. The second test problem shows the benefits of the constrained solution in a compositional model for which truncation of negative molar densities, as discussed in Chapter 3, results in significant problems with the mass of fluid in the model. The third example is a three-phase black-oil example for which the state variables have both lower and upper bounds. The final example is a simple 1D Buckley-Leverett flow example that investigates the effects of constraints on mass conservation during the update step.

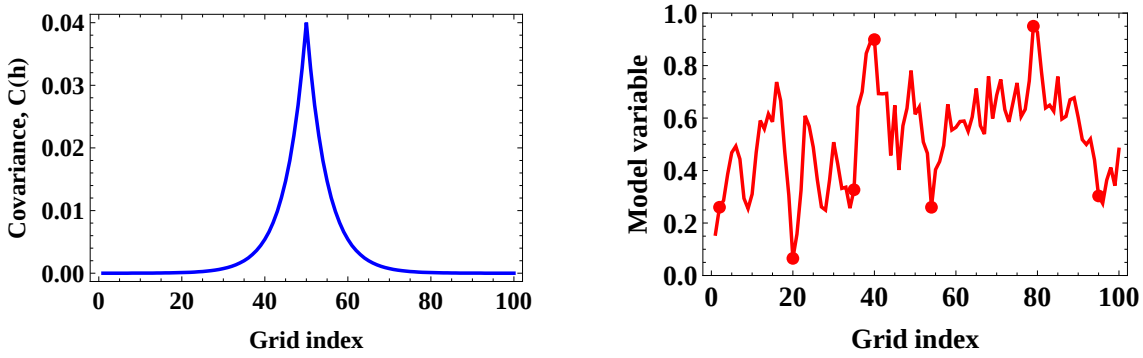
## 4.4 1D linear example

### 4.4.1 Model description

A simple 1D model is considered to demonstrate the validity of the CEnKF method. The 1D model ($x$-direction) is divided into 100 gridblocks. The present model is not

a reservoir flow model and therefore, the dimensions of a gridblock are not important. Instead, let us assume that the model represents a random field such that the random variable at each gridblock is subject to two inequality constraints defined by a lower and an upper bound. The feasible region is bounded by 0 on the lower side and by 1 on the upper side. We assume that the random variable is measured at a number of locations throughout the field. The standard EnKF and the CEnKF methods are implemented on this synthetic example in order to match the observations.

In the current investigation, the same exponential covariance model with standard deviation of 0.20 and a range of 15 is used for generating the reference (true) model and the initial ensemble of model realizations. The true covariance for the model variable at gridblock 50 with respect to model variables at other gridblocks in the model is shown in Fig. 4.2(a).




(a) True covariance

(b) Reference model and observations

**Figure 4.2:** (a) The true covariance for the model variable at gridblock 50, and (b) the reference model (red curve) together with the true observations (red dots) for the synthetic 1D example.

### 4.4.2 Reference model and initial ensemble

The random field is assumed to have a mean value of 0.50. The reference model and the initial ensemble of realizations of the random field are generated by sampling the prior Gaussian probability density function. Thus, the $i$th realization of the ensemble,

$\mathbf{y}_i$, can be generated as

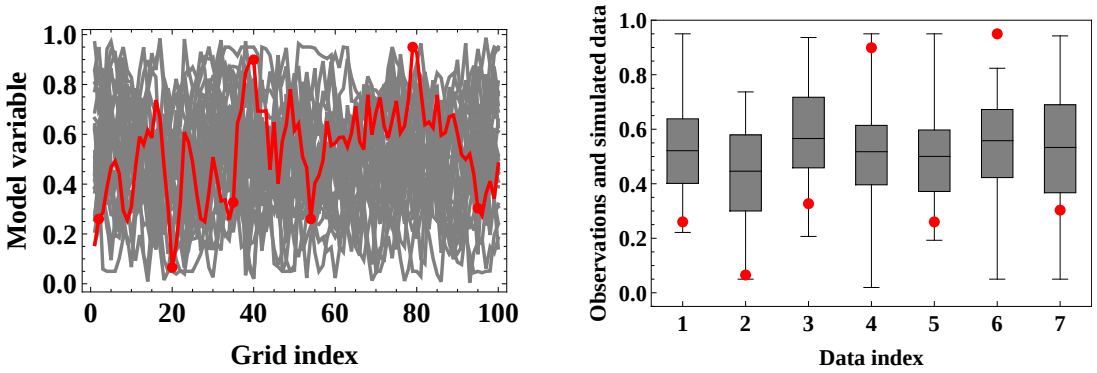
$$\mathbf{y}_i = \text{mean} + \mathbf{L}\mathbf{Z}_i \tag{4.6}$$

where, $\mathbf{L}$ is a lower triangular matrix obtained from the Cholesky decomposition of the (true) covariance matrix, and $\mathbf{Z}_i$ is the $i$th vector of independent normal deviates. The reference model generated in this manner is shown in Fig. 4.2(b).

Each realization of the initial ensemble is generated using Eq. 4.6. The random variable at each gridblock is subject to two inequality constraints. To ensure that each model variable of the individual ensemble member (prior) honors the inequality constraints, it is truncated to the value at the boundary of the feasible region whenever it is observed to take value outside the feasible region. Thus, all the model variables from the initial ensemble of realizations lie in the feasible region. The initial ensemble of realizations of the random variable is shown by gray curves in Fig. 4.3(a) and the box-plot (Fig. 4.3(b)) shows the statistical distribution of the simulated data obtained from the initial ensemble. The box contains 50% probability, the horizontal line in the box indicates the median, and the whiskers indicate the maximum and the minimum values. The predictions from the initial ensemble are generally very different than the observations from the reference model and show very high uncertainty. Note that, in Fig. 4.3(a) and in all the subsequent figures, unless stated otherwise, the red curve represents the reference model and the gray curves denote the initial ensemble of realizations of the random variable. Similarly, in Fig. 4.3(b) and all the subsequent figures, the red dots represent the observations from the reference model and the gray boxes denote the statistical distribution of the simulated data from the initial ensemble of realizations.

### 4.4.3 Observations for data assimilation

We consider a linear observation model where the observations used for data assimilation are the measurements of the random variable at a number of locations in the




(a) Model variables (b) Simulated data

**Figure 4.3:** (a) The initial ensemble of model variables (gray curves) and the reference model (red curve), and (b) the simulated data from the initial ensemble (gray boxes) together with the true observations (red dots) from the reference model.

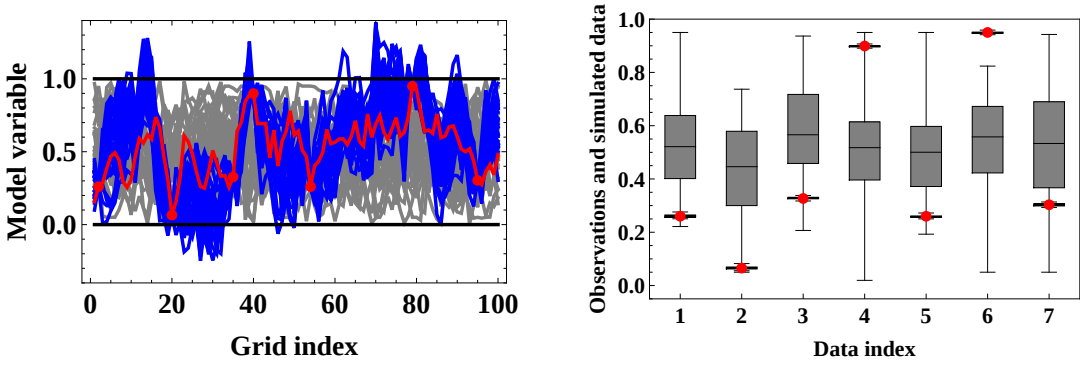
field. Let us assume that the random field has been measured at seven different locations. These measurements are assumed to be made at gridblock locations 2, 20, 35, 40, 54, 79, and 95. The measurements and their corresponding locations are given in Table 4.1. The measurement error in the observations has a standard deviation of 0.005. An ensemble of 30 ensemble members is considered for implementing the standard EnKF and the CEnKF. Thus, we have 100 model variables, 30 ensemble members, and 7 observations for the data assimilation process.

**Table 4.1:** The observations and the measurement locations for 1D linear example

| Location | Observation |
|---|---|
| 2 | 0.2602 |
| 20 | 0.0653 |
| 35 | 0.3268 |
| 40 | 0.8991 |
| 54 | 0.2600 |
| 79 | 0.9500 |
| 95 | 0.3032 |

### 4.4.4 Standard EnKF

The standard EnKF is implemented using the parameters discussed earlier. During the implementation of the standard EnKF and the CEnKF methods, perturbed observations, instead of true observations, are used for data assimilation. We refer to the solution obtained from the standard EnKF as the unconstrained solution. Fig. 4.4(a) shows the updated ensemble of model variables obtained from the standard EnKF. The blue curves represent the updated ensemble members and the horizontal black lines indicate the boundaries of the feasible region on the lower and the upper sides. The reference model is contained within the spread of the updated ensemble obtained from the standard EnKF. The data-match obtained from the updated ensemble is shown in Fig. 4.4(b) where the box-plots (black) represent the simulated data from the updated ensemble obtained from the standard EnKF.



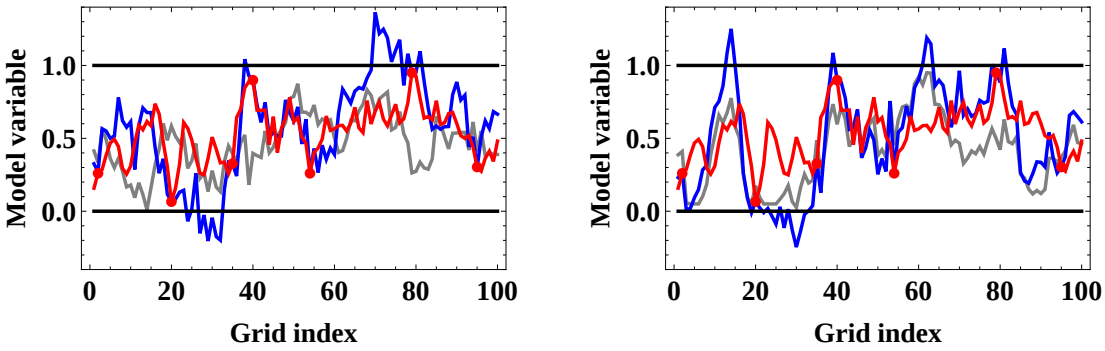

(a) Estimate of model variables

(b) Data-match

**Figure 4.4:** (a) The updated ensemble of model variables (blue curves) and (b) the simulated data (black boxes) obtained from the standard EnKF together with the true observations (red dots) from the reference model.

The results obtained from the standard EnKF show a good match to the observations. A number of updated realizations, however, consist of model variables that violate the inequality constraints and take values which lie outside the feasible region (either less than 0 or greater than 1). Fig. 4.5 shows two such ensemble members, realizations 1 and 15. The gray and the blue curves represent the prior and the

unconstrained solutions for realizations 1 and 15 obtained from the standard EnKF, respectively, the red curve is the reference model, and the red dots denote the true observations.




(a) Realization 1 (b) Realization 15

**Figure 4.5:** The prior and posterior solutions for realizations 1 and 15 (gray - prior realization; blue - unconstrained solution from the standard EnKF; red - reference model; red dots - true observations)

We define the following quantitative measure of the data mismatch for the standard EnKF and the CEnKF methods.

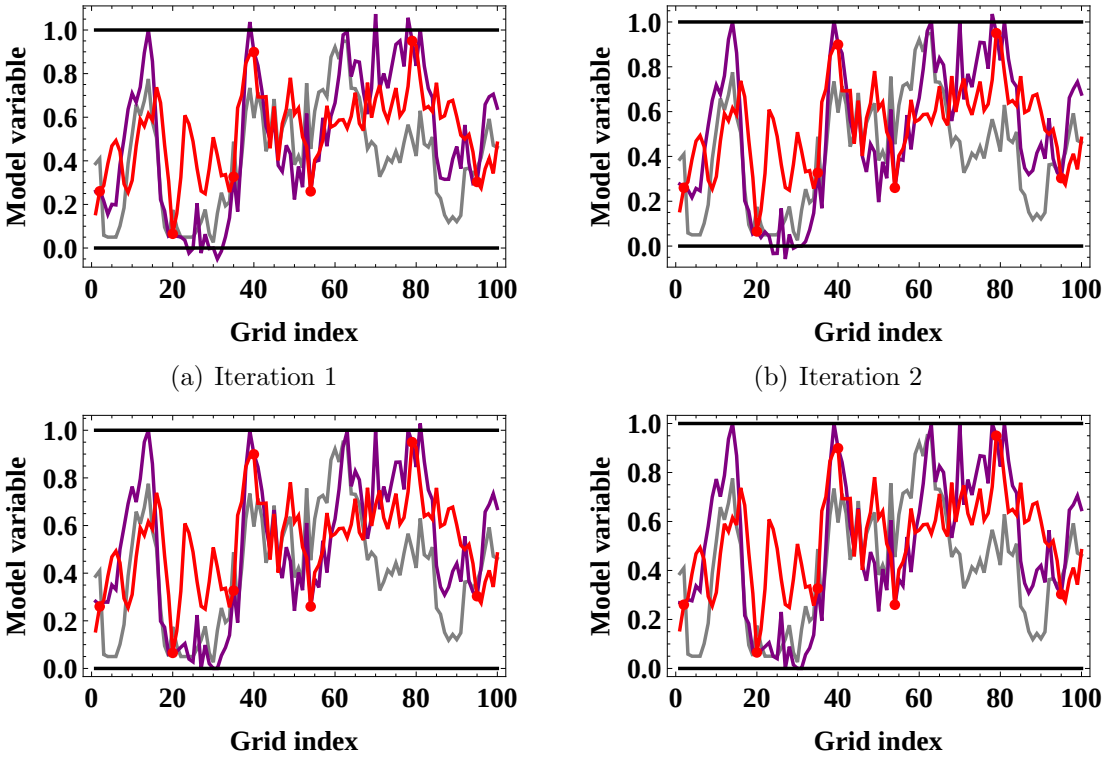
$$\text{Obj}_\text{D} = \frac{1}{2N_e}\sum_{i=1}^{N_e}[(\mathbf{d}_i - \mathbf{d}_{\text{obs},i})^T\mathbf{C}_D^{-1}(\mathbf{d}_i - \mathbf{d}_{\text{obs},i})] \tag{4.7}$$

where, $\mathbf{d}_i$ is the vector of the simulated data obtained from the updated solution for the $i$th ensemble member, $\mathbf{d}_{\text{obs},i}$ is the vector of the perturbed observations for the $i$th ensemble member used during the data assimilation process, and $\mathbf{C}_D$ is the observation-error covariance matrix. The data mismatch objective function computed for the standard EnKF has a value of 0.053.

### 4.4.5 Constrained EnKF

The results obtained from the application of the CEnKF to the 1D synthetic linear example are discussed in this subsection. We first examine the effectiveness of the CEnKF using realization 15. The unconstrained solution for realization 15 obtained from the standard EnKF is shown by the blue curve in Fig. 4.5(b). The unconstrained

solution for realization 15 showed values of 1.25 and -0.25 for the updated model variables at gridblocks 14 and 30, respectively, as the largest violations of constraints on the upper and the lower bounds, respectively. The vector of actual observations is expanded to include the new constraints of 1 and 0 (during iteration 1 of the CEnKF method) on the model variables at gridblocks 14 and 30, respectively. The data-error covariance matrix, $\mathbf{C}_D$, and the measurement operator, $\mathbf{H}$, are also expanded to reflect the newly added constraints. The constrained solution for realization 15 is computed using Eq. 4.5. Fig. 4.6 shows the constrained solutions for realization 15 at different iterations of the CEnKF. The degree of the largest constraint violation on both sides of the feasible region decreases with each iteration of the CEnKF method and the constrained solution moves towards the interior of the feasible region.




(a) Iteration 1

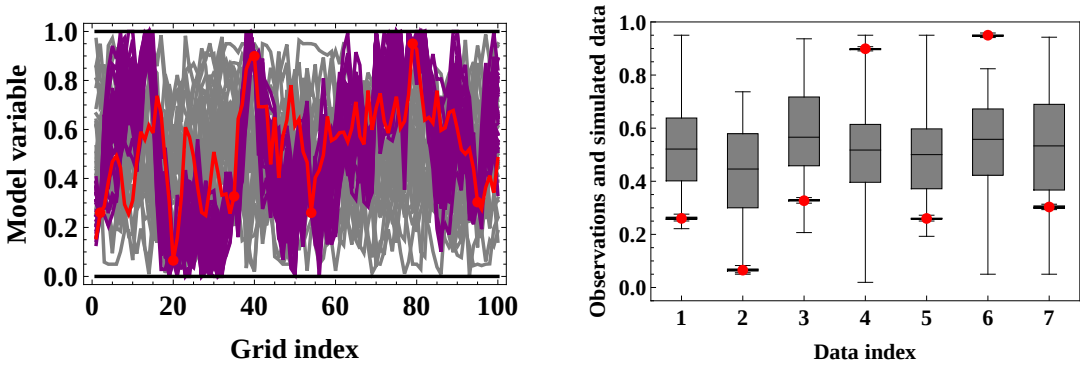
(b) Iteration 2

(c) Iteration 3

(d) Iteration 10

**Figure 4.6:** Realization 15 - constrained solutions at different iterations of the CEnKF method (gray - prior realization; purple - constrained solution from the CEnKF; red - reference model; red dots - true observations).

Fig. 4.7(a) shows the updated ensemble of model variables obtained from the CEnKF. The purple curves represent the updated realizations of model variables and the horizontal black lines indicate the boundaries of the feasible region on the lower and the upper sides. As Fig. 4.7(a) indicates, the reference model (red curve) is contained within the spread of the updated ensemble obtained from the CEnKF and model variables from each updated ensemble member honor the inequality constraints on the lower and the upper side of the feasible region. The match between the true observations (red dots) and the simulated data (black box-plots) from the updated ensemble obtained from the CEnKF is shown in Fig. 4.7(b).




(a) Estimate of model variables

(b) Data match

**Figure 4.7:** (a) The updated ensemble of model variables (purple curves), and (b) the simulated data (black boxes) obtained from the updated ensemble from the CEnKF method. The true observations are denoted by the red dots.

The results obtained from the CEnKF method show a good match to the observations and the updated model variables from all ensemble members honor the inequality constraints on the lower and the upper side of the feasible region. The data mismatch objective function (Eq. 4.7) computed for the CEnKF method has a value of 0.104. The results obtained from the application of the CEnKF to a 1D linear model show that the CEnKF is successful in enforcing the given constraints while simultaneously assimilating the observations. The CEnKF is also successful in restoring a part of the model space which otherwise tends to violate the inequality constraints and thus, results in a solution that lies in the feasible region.

## 4.5 Compositional model

In this section, we describe in detail the application of the CEnKF method to the data assimilation problem of $CO_2$ injection into a compositional reservoir model studied in Chapter 3. The reference reservoir model description and the initial ensemble of porosity and log-permeability have already been discussed in Section 3.1. The parameters used for data assimilation with the standard EnKF are mentioned in Section 3.2. The only additional parameters needed for the implementation of the CEnKF method are the set of the active constraints which are identified during each iteration.

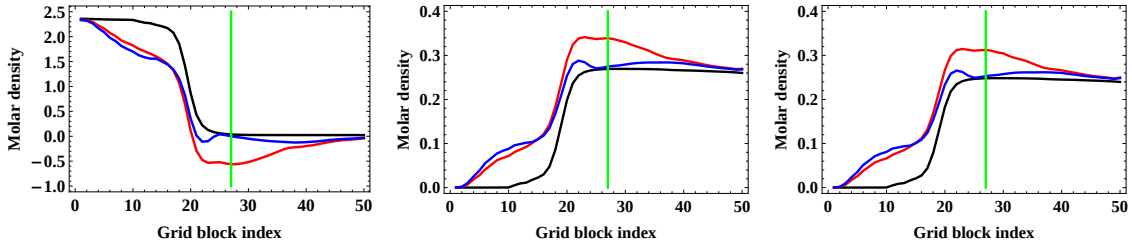
### 4.5.1 Results and discussion

The CEnKF method uses a different Kalman gain matrix and a different observation vector for each ensemble member. Following are the results obtained from the application of the CEnKF to the problem of data assimilation for the same ensemble member, realization 31, studied in Chapter 3.

#### *4.5.1.1 Analysis of realization 31*

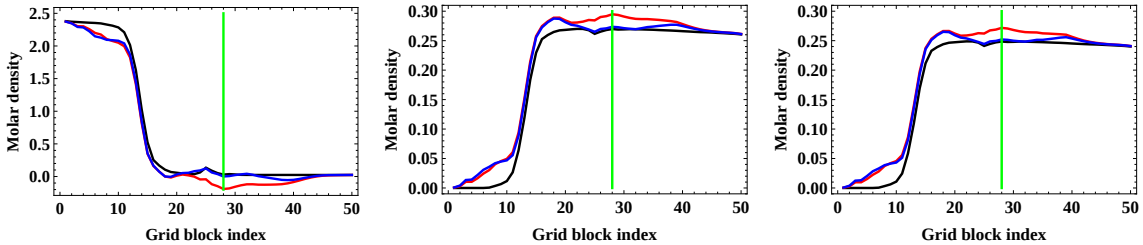
For the first data assimilation time step, the unconstrained solution for realization 31 showed a value of -0.56 for the updated $CO_2$ molar density at the gridblock (27,23) as the largest violation of the positivity constraint on $CO_2$ molar density value. Fig. 4.8 shows the results for the forecast, the unconstrained, and the constrained solutions for the molar densities of components $CO_2$, $NC_4$, and $C_7$ obtained from iteration 1 for layer 23 in the model. The green vertical line indicates the location of the gridblock 27 in layer 23 at which the constraint is enforced. At subsequent iterations, the largest constraint violation occurs in other layers, but there is still a small effect from the application of constraints on $CO_2$ molar density in other layers on the density in layer 23. It should be noted that, during the subsequent iterations, the constrained solution obtained from the previous iteration of the CEnKF is treated as the unconstrained

solution. After iteration 4, the largest violation of the positivity constraint on $CO_2$ molar density for realization 31 is observed at the gridblock 28 in layer 25, where the updated $CO_2$ molar density showed a value of -0.19. Figs. 4.9 and 4.10 show the results for the forecast, the unconstrained (constrained solution at iteration 4), and the constrained solutions for the molar densities of components $CO_2$, $NC_4$, and $C_7$ obtained from iteration 5 for layers 25 and 23, respectively, in the model. As Fig. 4.10(a) indicates, the application of constraints on $CO_2$ molar density in layer 25 during iteration 5 of the CEnKF results in a small change in $CO_2$ molar density in layer 23.



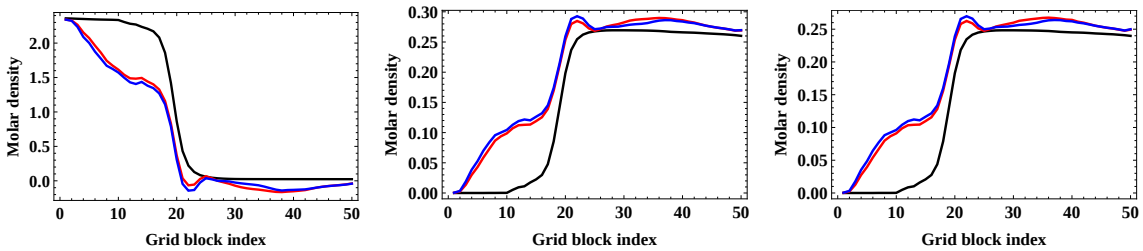

(a) $CO_2$ molar density (b) $NC_4$ molar density (c) $C_7$ molar density

**Figure 4.8:** Layer 23 - the forecast, the unconstrained (standard EnKF), and the constrained (CEnKF) solutions obtained at the first data assimilation time step using iteration 1 for the ensemble member 31. The black curve stands for the forecast, red curve denotes the unconstrained EnKF update, and the blue curve represents the constrained EnKF update.




(a) $CO_2$ molar density (b) $NC_4$ molar density (c) $C_7$ molar density

**Figure 4.9:** Layer 25 - the forecast, the unconstrained (CEnKF, iteration 4), and the constrained (CEnKF, iteration 5) solutions obtained at the first data assimilation time step using iteration 5 for the ensemble member 31. The black curve stands for the forecast solution, red curve denotes the unconstrained EnKF update, and the blue curve represents the constrained EnKF update.

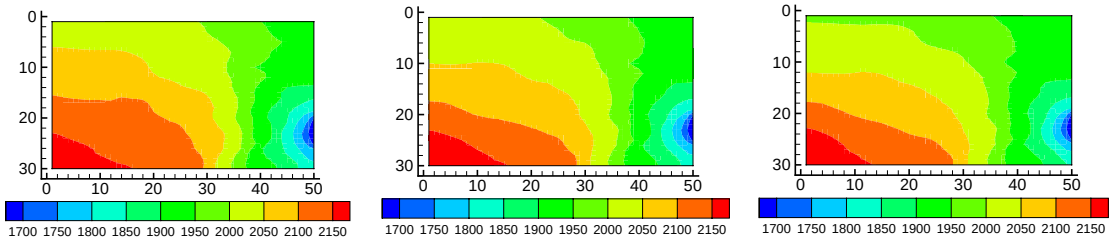



(a) $CO_2$ molar density (b) $NC_4$ molar density (c) $C_7$ molar density

**Figure 4.10:** Layer 23 - the forecast, the unconstrained (CEnKF, iteration 4), and the constrained (CEnKF, iteration 5) solutions obtained at the first data assimilation time step using iteration 5 for the ensemble member 31. The black curve stands for the forecast solution, red curve denotes the unconstrained EnKF update, and the blue curve represents the constrained EnKF update.

In the regions where the $CO_2$ molar density takes negative values, the unconstrained solution from the standard implementation of the EnKF shows anomalously high molar densities for all other components in the hydrocarbon phase. Figs. 4.8 and 4.9 show that the consequence of applying the constraints on the molar density of $CO_2$ is to reduce the increase in molar densities of components $NC_4$ and $C_7$. The CEnKF technique helps to ensure that the magnitudes of the increases and decreases in molar densities of different components are applied in a consistent manner. This consistency is missing when either transformation or truncation is used during the implementation of the standard EnKF. The results showed that the CEnKF is able to solve the problem of increased molar density (and thereby, additional mass) of components in a particular region of the reservoir field which was responsible for the high pressure build-up in that region, ultimately shutting the injector off.

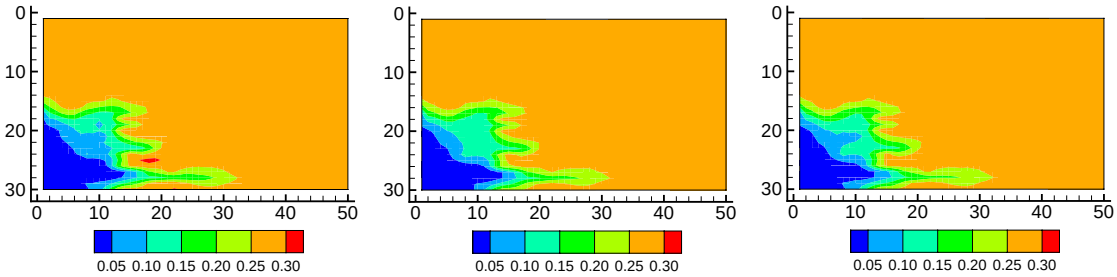
Figs. 4.11 through 4.13 show pressure and distributions of $NC_4$ and $C_7$ molar densities, respectively, for realization 31 in the reservoir at different iterations of the CEnKF method at the first data assimilation time step. The change in pressure and distribution of molar densities of two of the components between iteration 1 and iteration 7 is significant which is indicative of the large effect of applying constraints on $CO_2$ molar density during iteration 1 and iteration 7. The magnitude of the




(a) Iteration 1 (b) Iteration 7 (c) Iteration 10

**Figure 4.11:** Realization 31 - pressure in the reservoir at different iterations of the CEnKF method at the first data assimilation step.

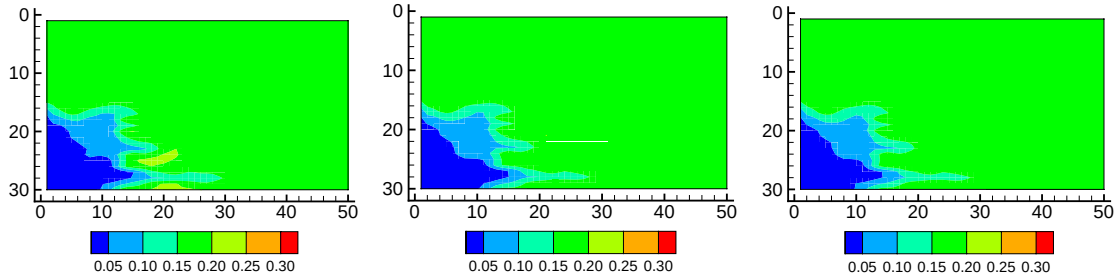
largest constraint violation for $CO_2$ molar density, however, reduces during subsequent iterations. As a result, the change in pressure and molar densities of $NC_4$ and $C_7$ between iteration 7 and iteration 10 is small.




(a) Iteration 1 (b) Iteration 7 (c) Iteration 10

**Figure 4.12:** Realization 31 - $NC_4$ molar density distribution in the reservoir at different iterations of the CEnKF method at the first data assimilation step.

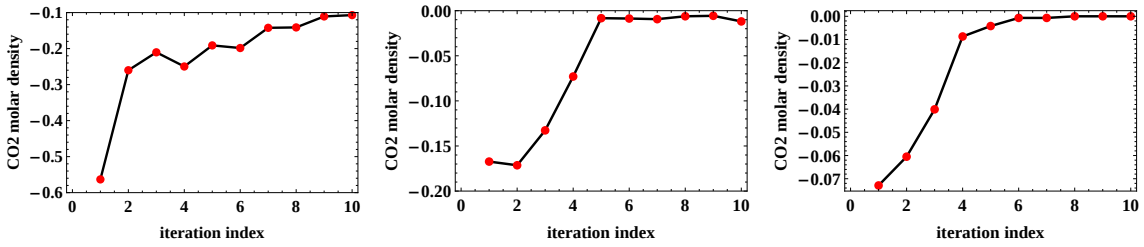
Fig. 4.14 shows the iterative reduction in magnitudes of the largest constraint violations in realization 31 for $CO_2$ molar density within the entire reservoir model for three data assimilation time steps. In this particular test problem, we observed fairly rapid reduction in the magnitude of the maximum constraint violation for the first 4 or 5 iterations, after which, the rate of reduction was decreased. The reduction is not necessarily monotonic, because the location of the largest constraint violation for $CO_2$ molar density changes with each iteration and spurious correlations can result in an increase in the magnitude of a violation at another location after updating. Because our emphasis was on efficiency, we arbitrarily set the maximum number of iterations at 10 with good results. Although the CEnKF method is an iterative




(a) Iteration 1 (b) Iteration 7 (c) Iteration 10

**Figure 4.13:** Realization 31 - $C_7$ molar density distribution in the reservoir at different iterations of the CEnKF method at the first data assimilation step.

process, the computational cost required during each iteration is small compared to the cost of running the reservoir simulator. The additional computational cost for each iteration arises from the identification of the active constraints and the computation of the modified Kalman gain for individual ensemble member. Although the Kalman gain is computed separately for each ensemble member, the data-storage requirement during each iteration is fairly small as the largest difference is the need to store both the forecast and the updated state vectors at each assimilation time. As a result, the implementation of the CEnKF method on a real-field reservoir problem with few hundred thousands of gridblocks should not impose any computational or data-storage challenge.

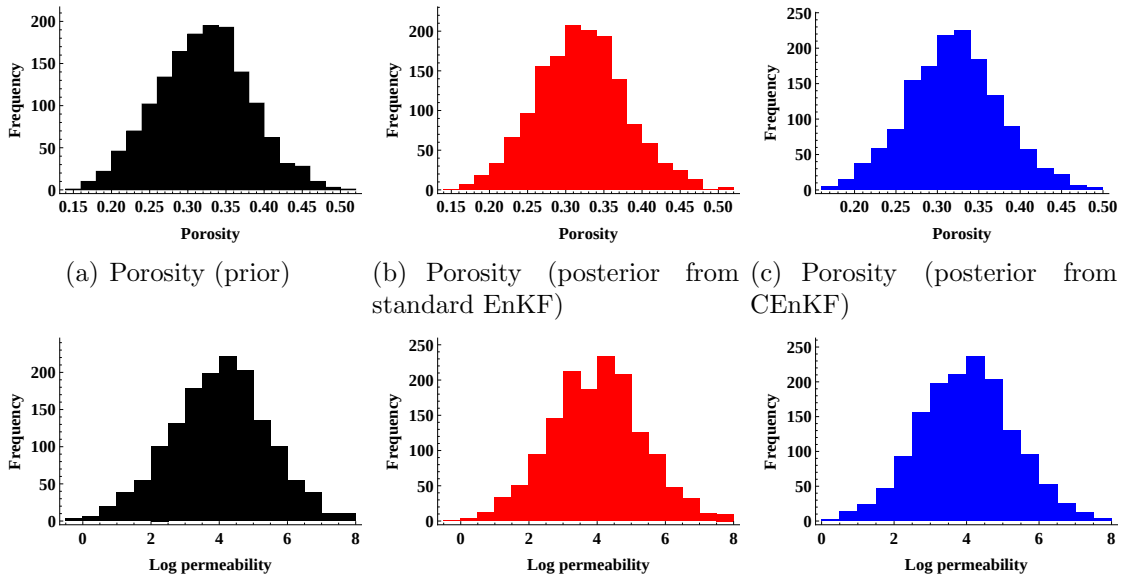



(a) Assimilation time step 1 (b) Assimilation time step 4 (c) Assimilation time step 7

**Figure 4.14:** Realization 31 - the minimum value of $CO_2$ molar density at different iterations of the CEnKF method for some of the data assimilation time steps.

In addition to the ease of implementation, the CEnKF method is robust and no extra assumptions about the distributions of the state variables are required for its

application. Figs. 4.15(a) and 4.15(d) show that the prior distributions of porosity and log-permeability fields for realization 31 are Gaussian. The posterior distributions of porosity and log-permeability for realization 31 from the standard EnKF and the CEnKF methods are shown in Fig. 4.15. During the implementation of the CEnKF method, no constraints are enforced on either porosity or log-permeability. The posterior porosity and log-permeability data obtained from the CEnKF method appear to follow the Gaussian distribution. There is no noticeable effect of the CEnKF on the prior distributions of porosity and log-permeability.




(a) Porosity (prior)

(b) Porosity (posterior from standard EnKF)

(c) Porosity (posterior from CEnKF)

(d) Log-permeability (prior)

(e) Log-permeability (posterior from standard EnKF)

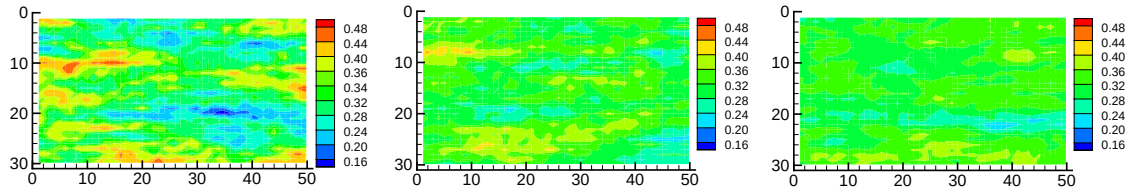
(f) Log-permeability (posterior from CEnKF)

**Figure 4.15:** The distributions of porosity and log-permeability for realization 31 at the first data assimilation time step and iteration 1 obtained from different methods.

#### *4.5.1.2 Model parameter estimates*

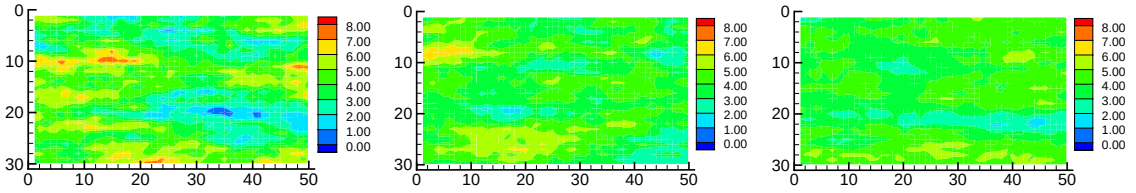
At the end of each data assimilation time, the standard EnKF and the CEnKF provide an ensemble of model realizations which are conditioned to all measurements up to that time. For a good performance of the EnKF, it is necessary that the resulting ensemble is a good representative of the underlying probability space throughout the entire history matching period. For synthetic examples where the reference model

is known, the mean of the final ensemble of estimates obtained from the EnKF can be used to evaluate the performance of the method. Fig. 4.16 shows the estimates of porosity fields obtained from the standard EnKF and the CEnKF, respectively, together with the reference porosity field. The final estimates of porosity provided by the standard EnKF and the CEnKF do not show a close resemblance to the reference porosity field. Both methods, however, are able to capture some of the features seen in the reference porosity field. The final estimate of the porosity from the CEnKF shows a region with low porosity near the producer, similar to the reference porosity field. Similar results were observed for the final estimates of the log-permeability field obtained from the standard EnKF and the CEnKF which are shown in Fig. 4.17.




(a) Reference (b) Standard EnKF (c) CEnKF

**Figure 4.16:** Final mean porosity fields from the standard EnKF and the CEnKF methods together with the reference porosity field.




(a) Reference (b) Standard EnKF (c) CEnKF

**Figure 4.17:** Final mean log-permeability fields from the standard EnKF and the CEnKF methods together with the reference log-permeability field.

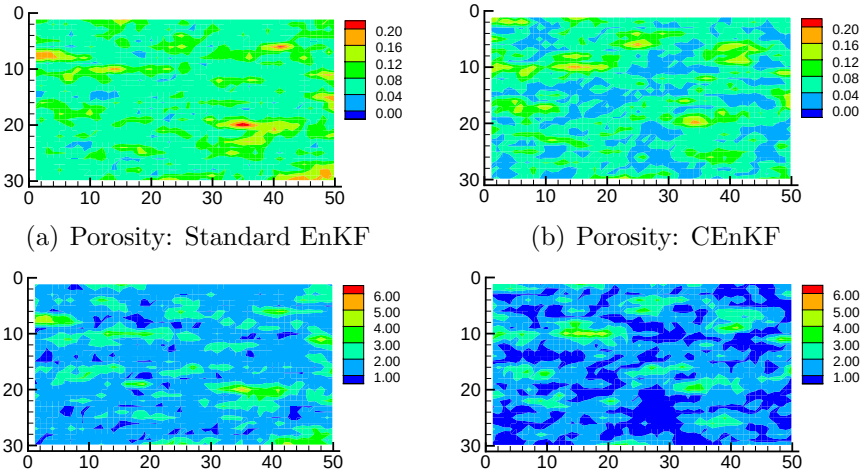
Another measure of the performance which can be used to compare different methods in case of synthetic examples is the gridblock-based root mean squared error (RMSE) of estimates. The RMSE of the estimate $x$ for the $i$th gridblock in the model

can be computed as

$$\mathbf{RMSE}_i = \sqrt{\frac{1}{N_e}\sum_{j=1}^{N_e}(x_{i,j} - x_i^{true})^2} \qquad (i = 1, 2, \ldots, N_x), \tag{4.8}$$

where, $x_i^{true}$ is the true value of variable of interest at the $i$th gridblock, $x_{i,j}$ is the estimate of the variable at the $i$th gridblock for the $j$th ensemble member, $N_e$ represents the ensemble size, and $N_x$ is the total number of gridblocks in the simulation model. The RMSE values of the final updated porosity and log-permeability estimates obtained from the standard EnKF and the CEnKF methods are computed and compared in Fig. 4.18. The RMSE values for the final estimates of porosity from the standard EnKF show large variability and are generally in the range of 0.04 to 0.12 for most part of the reservoir model with the exception of few locations where the RMSE values are close to 0.20 (Fig. 4.18(a)). The final estimates of porosity obtained from the CEnKF show RMSE values lower than 0.04 for most of the regions. We observed similar results for the final estimates of log-permeability field (Figs. 4.18(c) and 4.18(d)). The results show that the final estimates of model parameters obtained from the CEnKF are closer to the reference model compared to the estimates obtained from the standard EnKF.

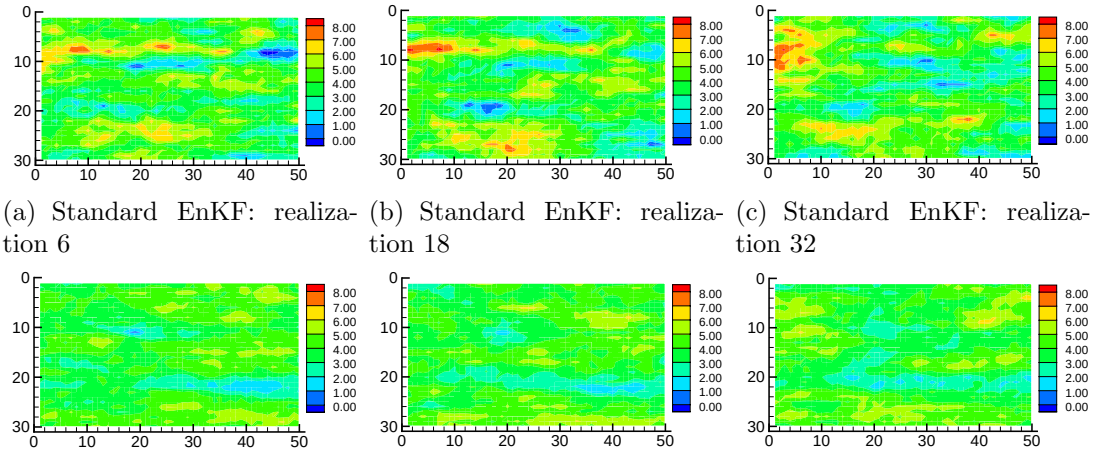
Fig. 4.19 compares three realizations of the final updated log-permeability fields obtained from the standard EnKF and the CEnKF. The updated realizations from the standard EnKF appear to have larger variability, however, they do not show the characteristics present in the reference log-permeability field. The updated realizations obtained from the CEnKF method show less variability compared to the standard EnKF. All three realizations of log-permeability from the CEnKF method shown in Fig. 4.19 appear to have the low-permeability channel, similar to the reference field, near the producer completions. Fig. 4.20 shows the standard deviation of log-permeability computed from the final updated ensembles obtained from the standard EnKF and the CEnKF. The results indicate that there is a larger variability




(a) Porosity: Standard EnKF (b) Porosity: CEnKF

(c) Log-permeability: Standard EnKF (d) Log-permeability: CEnKF

**Figure 4.18:** The root mean squared error (RMSE) for porosity and log-permeability computed from the final estimates obtained from the standard EnKF and the CEnKF methods.

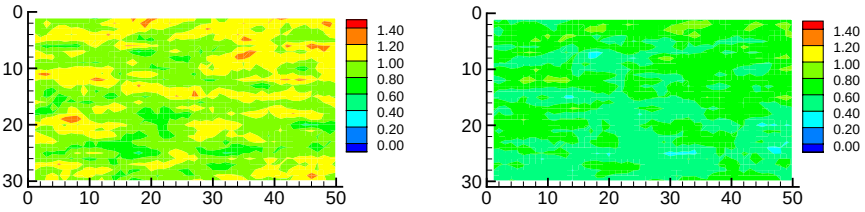



(a) Standard EnKF: realization 6 (b) Standard EnKF: realization 18 (c) Standard EnKF: realization 32

(d) CEnKF: realization 6 (e) CEnKF: realization 18 (f) CEnKF: realization 32

**Figure 4.19:** The final updated realizations of log-permeability from the standard EnKF and the CEnKF for different ensemble members.

among the final updated ensemble obtained from the standard EnKF compared to the CEnKF. The standard deviation of log-permeability from the CEnKF has values in the range of 0.4 to 0.6 in some part of the reservoir which is lower than the standard deviation used for generating the initial ensemble (1.4). In some parts of the reservoir, however, the standard deviation values have magnitude of around 1 which suggests that the final updated ensemble from the CEnKF has enough variability for assimilating additional production data in the future.



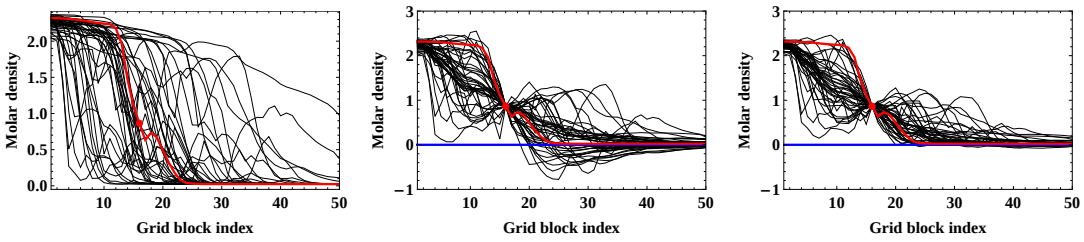

(a) Standard EnKF (b) CEnKF

**Figure 4.20:** The standard deviation of log-permeability from the standard EnKF and the CEnKF.

#### *4.5.1.3 Estimation of molar densities*

Fig. 4.21 compares the updated ensemble of $CO_2$ molar densities for layer 22 of the reservoir model from the standard EnKF and CEnKF methods. The black curves in Figs. 4.21(b) and 4.21(c) are the posterior $CO_2$ molar densities for ensemble members from the standard EnKF and the CEnKF, respectively. The thick red curve is the $CO_2$ molar density for layer 22 from the reference model at day 102, the red dot denotes the $CO_2$ molar density measurement, and the horizontal blue line represents the boundary of the feasible region for $CO_2$ molar density. The updated realizations of $CO_2$ molar density for layer 22 obtained from the standard EnKF and the CEnKF show a good agreement with the $CO_2$ molar density measurement (red dot in Figs. 4.21(b) and 4.21(c)). For most of the ensemble members, the updated $CO_2$ molar density from the analysis step in the standard EnKF is negative at some locations. The updated $CO_2$ molar densities from the CEnKF method honor the positivity constraint. Since the

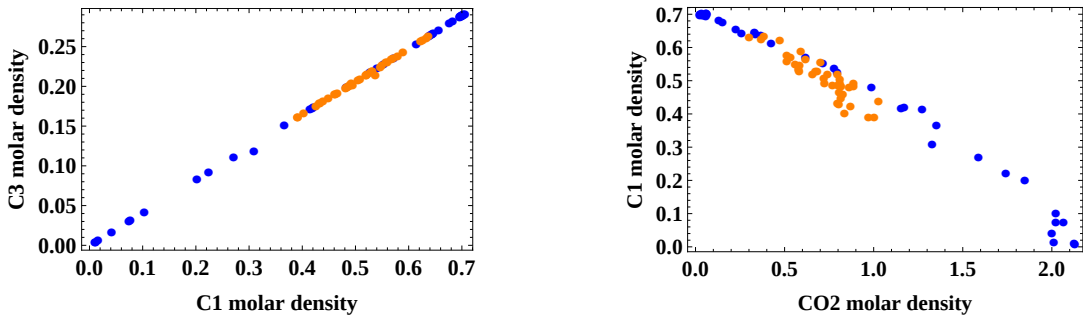
maximum number of iterations for the CEnKF method is set to 10, the updated $CO_2$ molar density (Fig. 4.21(c)) for some ensemble members shows a small violation of positivity constraint. Similar to the standard EnKF, the non-physical molar density of $CO_2$ from the CEnKF method is truncated to its value at the boundary of the feasible region before advancing the simulator to the next data assimilation time. Since the degree of constraint violation on $CO_2$ molar density in the CEnKF is significantly smaller than the standard EnKF, the truncation approach, unlike in the standard EnKF, does not lead to harmful results in the CEnKF method.




(a) Forecast (b) Standard EnKF (posterior) (c) CEnKF (posterior)

**Figure 4.21:** The posterior ensemble of $CO_2$ molar density for layer 22 from the standard EnKF and CEnKF (iteration 10) methods together with the forecast ensemble of $CO_2$ molar density at the first data assimilation step. The black curves are the ensemble members, the thick red curve is the reference, and the red dot represents $CO_2$ molar density measurement.

Fig. 4.22 shows the relationship between molar densities of different components at a location in the reservoir model before and after assimilating data at the first data assimilation step. The blue dots represent the forecast molar densities and the orange dots are the updated molar densities obtained from the CEnKF method. The forecast molar densities of components $C_3$ and $C_1$ are positively correlated (Fig. 4.22(a)). As $CO_2$ is being injected into the reservoir, the forecast molar densities of components $C_1$ and $CO_2$ are negatively correlated (Fig. 4.22(b)). The analysis step of the CEnKF is able to preserve the linear relationship between the molar densities of different components as indicated by the orange dots in Fig. 4.22.




(a) $C_3$ - $C_1$ (b) $C_1$ - $CO_2$

**Figure 4.22:** The cross-plots between molar densities of (a) $C_3$ - $C_1$ and (b) $C_1$ - $CO_2$ at location (18,22) obtained from the CEnKF at the first data assimilation step. The forecast and the updated molar densities are shown in the blue and the orange colors, respectively.

#### *4.5.1.4 Matching production data*

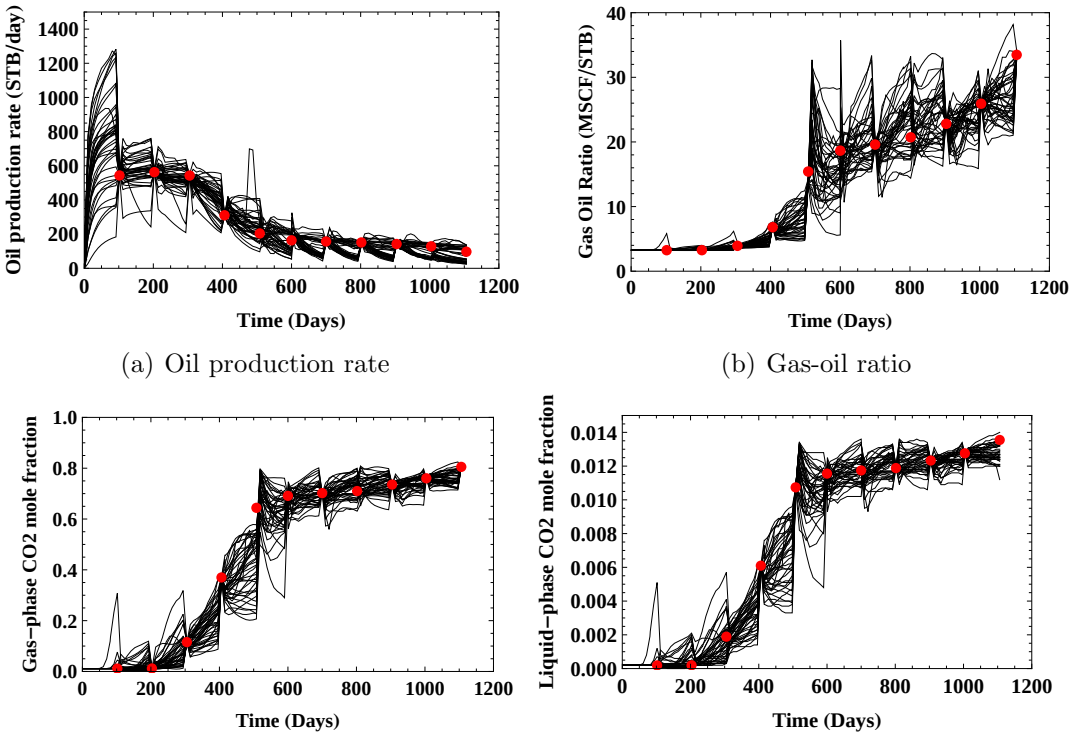
The static model parameters (porosity and permeability) and the dynamic state variables (pressure and molar densities) are updated sequentially in time using the information contained in the production data (e.g. oil production rate, $CO_2$ injection rate, etc.). Figs. 4.23 and 4.24 show forecast of different types of production data during the course of the data assimilation process for the standard EnKF and the CEnKF, respectively. The black curves indicate the forecast from the ensemble members and the red dots denote the actual measurements from the reference model. The static model parameters and dynamic state variables for each ensemble member updated at a data assimilation step are used for running the ensemble member from that data assimilation time to the next assimilation time. As the static model parameters and dynamic state variables are updated at each assimilation time, the ensemble forecasts of different types of production data show large changes after the data are assimilated. At the first data assimilation step, the observations are assimilated which provide a large amount of information to ensemble members about the reservoir flow dynamics (of the real field or the reference synthetic model). Therefore, large changes in the behavior of the forecast production data are seen after the first data assimilation

step (e.g. Figs. 4.23(a) and 4.24(a)). During the application of the standard EnKF to the compositional model, as discussed in Chapter 3, nearly 60% of the ensemble members exhibited periods during which an injector became shut-in after the first data assimilation time. The absence of $CO_2$ injection for some ensemble members causes incorrect molar density distributions in the reservoir during the subsequent forecast step. The result is incorrect prior covariances and incorrect updates to the static model parameters and dynamic state variables during each of the subsequent data assimilation steps of the standard EnKF. Therefore, forecasts of the production data from the standard EnKF do not show improvement and have large variability (e.g. Fig. 4.23(b)). Contrarily, forecasts of the production data from the CEnKF show continued improvement starting at the first assimilation time as indicated in Fig. 4.24. The estimates of the static model parameters and dynamic state variables updated at each data assimilation step from the CEnKF are better than those obtained from the standard EnKF which results in better production forecasts from the CEnKF compared to the standard EnKF.

The ensemble of $CO_2$ injection rate profiles obtained from the standard EnKF and the CEnKF methods during data assimilation are compared in Fig. 4.25. The application of the inequality constraint on $CO_2$ molar density in the CEnKF method ensures that the updated $CO_2$ molar densities are non-negative while the molar densities of other components are reduced in a consistent manner. Thus, the CEnKF significantly improves the estimates of molar density distributions in the reservoir compared to the estimates obtained from the standard EnKF. As a result, the variability in the injection rate seems more reasonable in the CEnKF (Fig. 4.25(b)).

#### *4.5.1.5 Predictability of the updated ensemble*

The ultimate goal of a history matching exercise is to obtain an ensemble of model parameters which honor all the previous production history, so that, it can be used



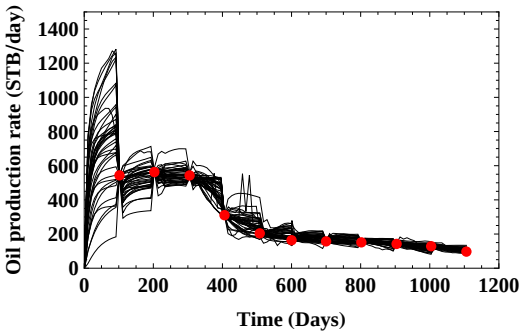

(a) Oil production rate

(b) Gas-oil ratio

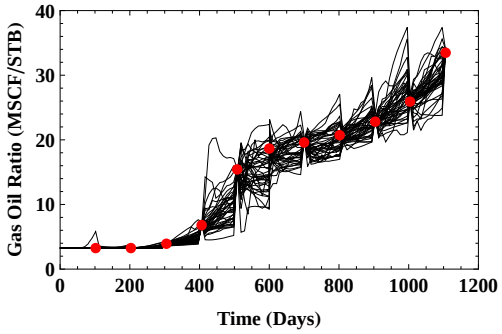
(c) Gas-phase $CO_2$ mole fraction

(d) Liquid-phase $CO_2$ mole fraction

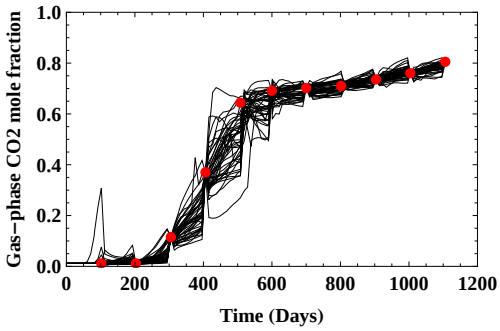
**Figure 4.23:** The production forecasts from the standard EnKF during the course of the data assimilation. The black curves denote the ensemble members and the red dots denote the true observations.



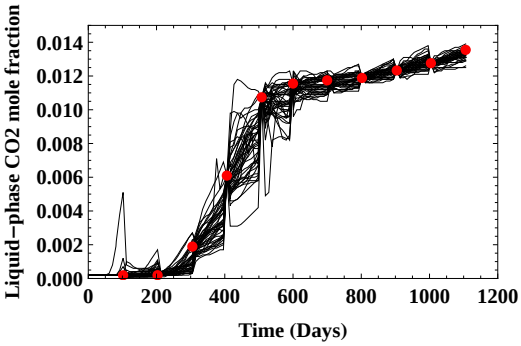

(a) Oil production rate




(b) Gas-oil ratio

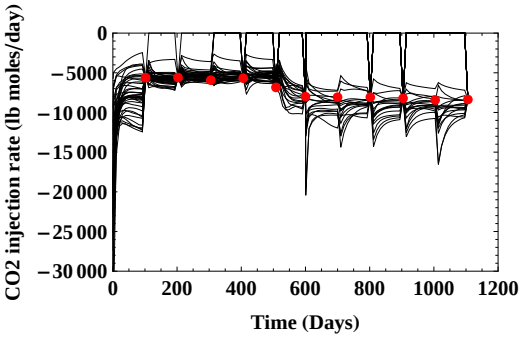



(c) Gas-phase $CO_2$ mole fraction

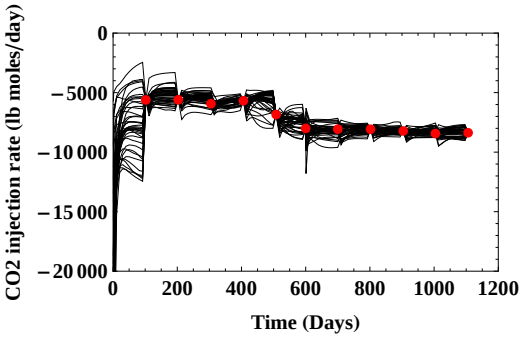



(d) Liquid-phase $CO_2$ mole fraction

**Figure 4.24:** The production forecasts from the CEnKF during the course of the data assimilation. The black curves denote the ensemble members and the red dots denote the true observations.




(a) Standard EnKF



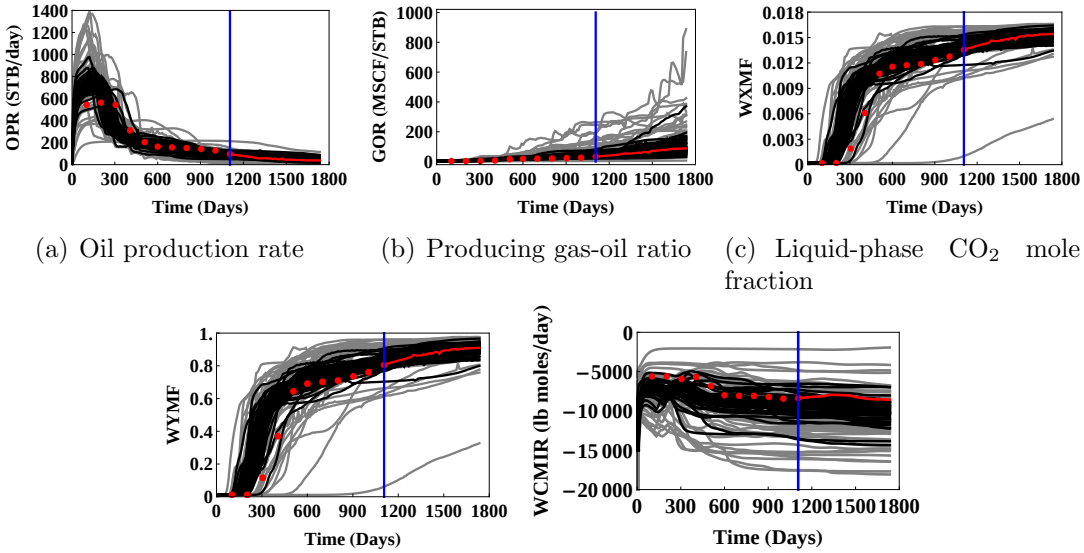

(b) CEnKF

**Figure 4.25:** The $CO_2$ injection rate profiles for all the ensemble members during data assimilation obtained from the standard EnKF and the CEnKF methods.

for predicting the future reservoir performance. The prediction using EnKF methods can start at the end of the data assimilation period, where the final updated static model parameters and dynamic state variables are directly used to advance the reservoir simulator into the future prediction period. An alternate way of measuring the performance of the EnKF methods is to use the updated ensemble of the static model parameters to rerun the entire history starting from the initial equilibrium reservoir conditions to predict the future performance.

We evaluated the performance of both EnKF methods by considering the quality of the production forecasts from the initial time (day 0). The final updated ensemble of porosity and permeability fields obtained from the standard EnKF and the CEnKF methods are run from the initial time (day 0) up to the end of the prediction period (day 1735). Figs. 4.26 and 4.27 show the prediction of the final updated ensemble from the standard EnKF and the CEnKF, respectively. The gray curves are the predictions from the initial ensemble, the black curves are the predictions from the final updated ensemble, and the red curve shows the production data obtained from the reference model in the prediction period (from day 1106 to day 1735). The different types of production data shown in Figs. 4.26 and 4.27 include oil production rate (OPR), producing gas-oil ratio (GOR), liquid-phase $CO_2$ mole fraction at the producer (WXMF), gas-phase $CO_2$ mole fraction at the producer (WYMF), and $CO_2$ injection rate (WCMIR). It is fairly clear that, although the final ensemble members for both methods match production data better than the initial ensemble members, the predictability of the ensemble updated by the CEnKF is better than that of the standard EnKF. In particular, the ensemble predictions of breakthrough time from the CEnKF (Figs. 4.27(c) and 4.27(d)) are better than the predictions from the standard EnKF (Figs. 4.26(c) and 4.26(d)). The predictions of oil production rate at early time from the CEnKF (Fig. 4.27(a)) are also better than the predictions from the standard EnKF (Fig. 4.26(a)). The RMSE for the CEnKF method was nearly

a factor of two better than for the standard EnKF for all types of production data (Table 4.2).




(a) Oil production rate

(b) Producing gas-oil ratio

(c) Liquid-phase $CO_2$ mole fraction

(d) Gas-phase $CO_2$ mole fraction

(e) $CO_2$ injection rate

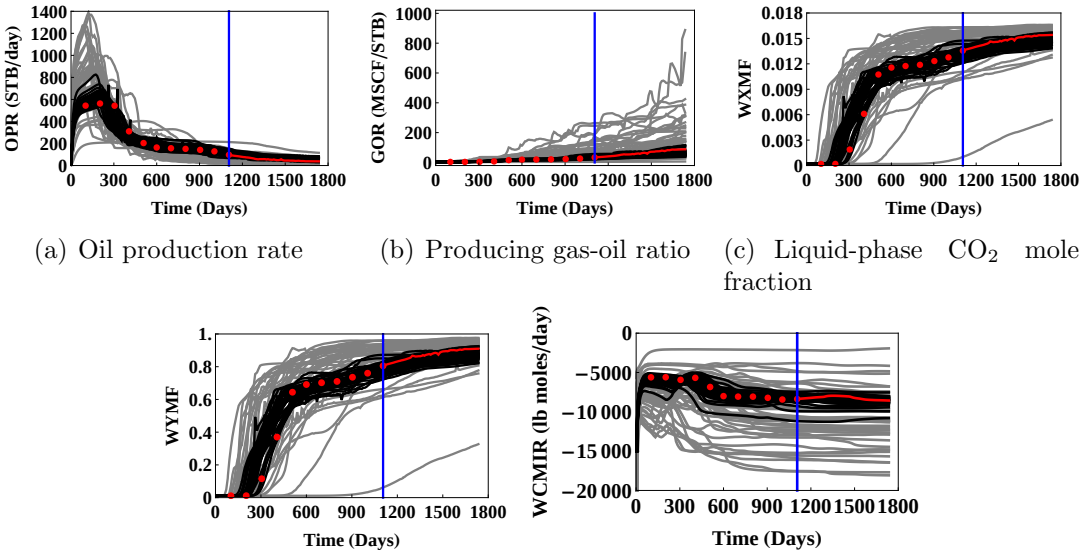
**Figure 4.26:** The production data predicted from the ensemble of porosity and permeability updated at day 1106 by the standard EnKF. The gray curves represent the initial ensemble, the black curves represent the updated ensemble, the red dots represent the observations from the reference model, the red curve denotes the production data from the reference model in the prediction period, and the vertical blue line indicates the end of the history matching period (day 1106).

## 4.6 Black-oil model

The implementation of the EnKF with truncation often works well for data assimilation problem of black-oil model involving only two phases (e.g. oil and water).

**Table 4.2:** Compositional model - RMSE values for different types of production data

| Observation | Standard EnKF | Constrained EnKF |
|---|---|---|
| Oil production rate (STB/day) | 79.3 | **49.7** |
| $CO_2$ injection rate (lb moles/day) | 1899. | **668.** |
| Producing GOR (MSCF/STB) | 8.41 | **3.27** |
| Producing gas-phase $CO_2$ mole fraction | 0.109 | **0.063** |
| Producing liquid-phase $CO_2$ mole fraction | 0.00185 | **0.00106** |




(a) Oil production rate

(b) Producing gas-oil ratio

(c) Liquid-phase $CO_2$ mole fraction

(d) Gas-phase $CO_2$ mole fraction

(e) $CO_2$ injection rate

**Figure 4.27:** The production data predicted from the ensemble of porosity and permeability updated at day 1106 by the CEnKF. The gray curves represent the initial ensemble, the black curves represent the updated ensemble, the red dots represent the observations from the reference model, the red curve denotes the production data from the reference model in the prediction period, and the vertical blue line indicates the end of the history matching period (day 1106).

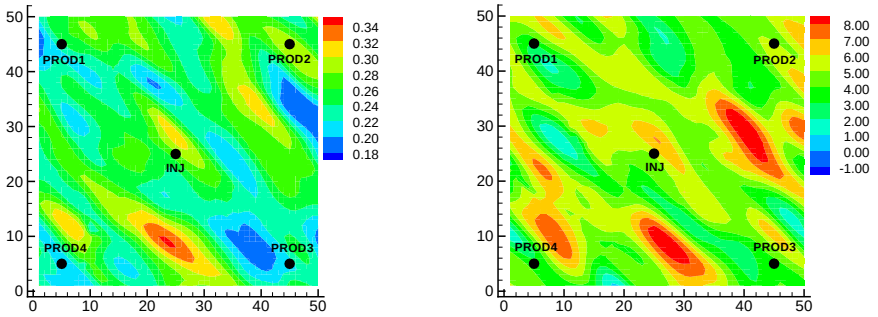
When it is important to consider all three phases (oil, gas, and water) present in the reservoir, the implementation of the EnKF with truncation, sometimes, fails to obtain good results. When the phase saturations of two phases are included in the state vector, often times, the updated phase saturations take non-physical values after data assimilation. Truncating the non-physical phase saturations of two phases directly effects the saturation of the remaining phase. The inconsistency in adjusting phase saturations using the truncation approach has more severe consequences on the reservoir performance of three-phase reservoir models compared to models where only two phases are present. In this section, the results from the implementation of the EnKF in its standard form with truncation and the implementation of the CEnKF method (augmented data approach) to a data assimilation problem of black-oil model involving three phases are discussed. The reservoir model description is given below.

### 4.6.1 Reservoir model description

The synthetic 2D ($x - y$ areal) three-phase black-oil reservoir model has connate water saturation of 0.15 with no initial gas saturation. The initial reservoir pressure is 3600 psia, which is also the bubble point pressure. The reservoir is divided into a uniform Cartesian grid of dimensions $50 \times 50 \times 1$. The gridblock dimensions are 30 ft $\times$ 30 ft $\times$ 20 ft. We assume that all four sides of the reservoir model are no-flow boundaries, and the phases present in the reservoir are oil, water, and gas. We neglect the capillary effects. There are four producers and one injector in the reservoir model arranged in a five-spot well pattern with the injector at the center and four producers at the corners. The injector, INJ, is located at gridblock (25,25), and the producers PROD1, PROD2, PROD3, and PROD4 are located at gridblocks (5,45), (45,45), (45,5), and (5,5), respectively. All four producers are operational from day 0. We also assume that the water injection is started from day 0 and is continued until the end of the production period (day 520). All four producers are operated with a

target oil production rate of 1500 STB/day and a minimum BHP of 1000 psia as a secondary constraint. The injector is operated with a target water injection rate of 8000 STB/day and a maximum BHP of 6000 psia as a secondary constraint.

The porosity field has a mean of 0.25 and a standard deviation of 0.03. The log-permeability field is generated with a mean and a standard deviation of 4.8 and 1.4, respectively, and is correlated to porosity with a correlation coefficient of 0.75. An anisotropic Gaussian variogram model with practical ranges of 30 and 5 gridblocks in the two principal directions is used to generate the static model parameter fields. Fig. 4.28 shows the porosity and log-permeability fields for the reference case.




(a) Porosity (b) Log-permeability

**Figure 4.28:** The porosity and log-permeability fields for the black-oil model (reference).

### 4.6.2 Data assimilation

The porosity and log-permeability fields for the reference model and the initial ensemble are generated using the same geostatistical model parameters as the reference model. An ensemble of 100 realizations is conditioned to values of permeability and porosity at well locations from the reference model. Fig. 4.29 shows the production data obtained from the initial ensemble together with the reference production data. The gray curves indicate the ensemble forecasts and the black curve shows the production data for the reference model. The initial ensemble shows good variability for different types of production data and the reference model (black curve in Fig. 4.29)

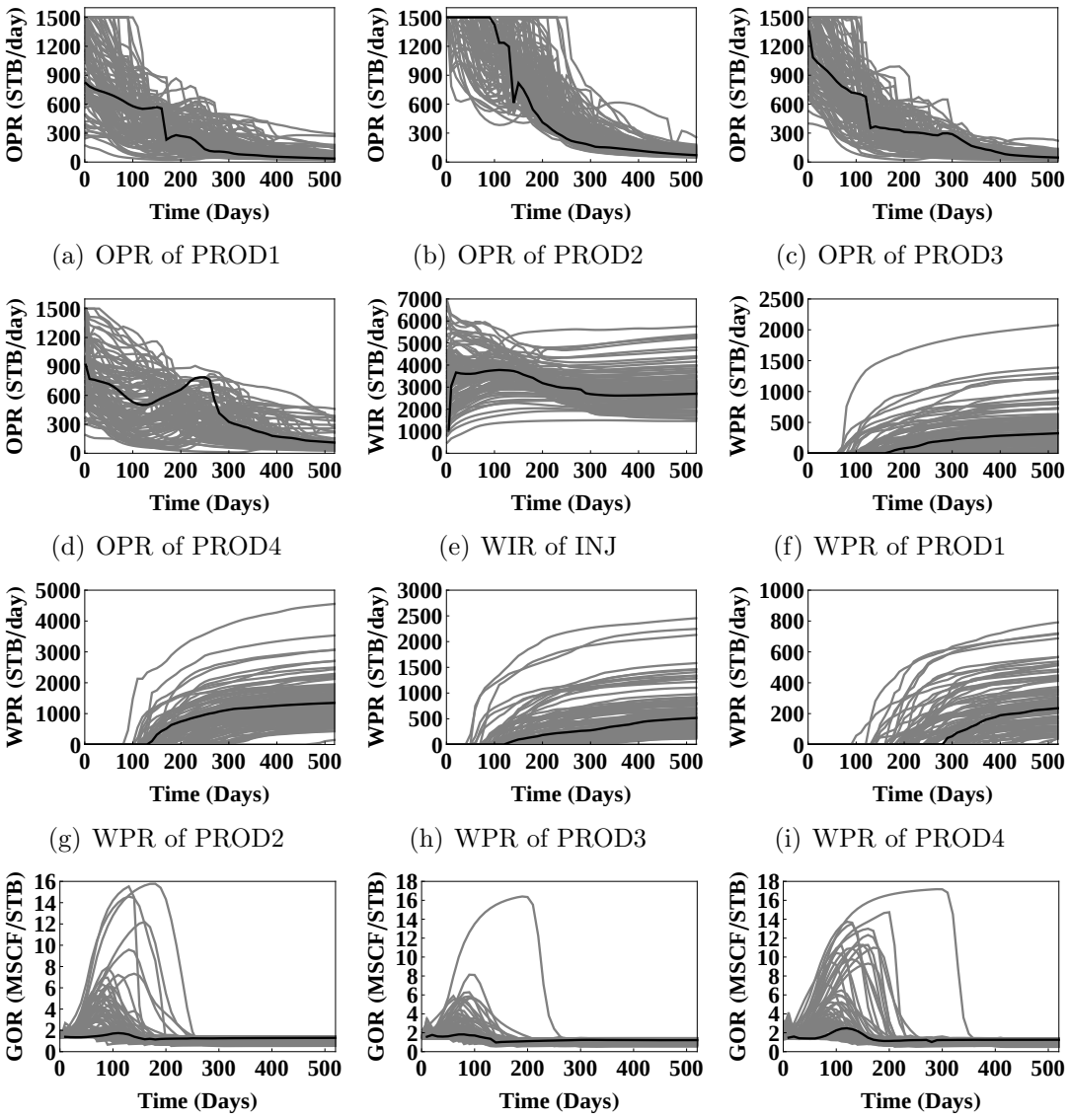
is contained within the spread of the initial ensemble.

The state vector includes static model parameters, such as porosity and log-permeability, along with dynamic state variables, such as pressure, water saturation, and gas saturation. The simulated production data, including oil production rate, producing gas-oil ratio, water production rate, and water injection rate are used in the data assimilation. The total simulated production history is 280 days. The production data are assimilated at days 10, 40, 80, 120, 160, 200, and 280. There are a total of 7 data assimilation time steps and 13 data at each assimilation step. The measurement noise associated with the production data is sampled from a normal distribution with a mean of 0 and a standard deviation which varies by type of observation. The standard deviation values used for different types of production data are given in Table 4.3. Note that, two different values of the standard deviation are used for the water production rate observation. When the true water production rate is less than 1.0 STB/day, the standard deviation is 0.0001 STB/day; otherwise it is 4.0 STB/day.

In the implementation of the CEnKF method, the updated water saturation is constrained to the connate water saturation value of 0.15 whenever the updated value is smaller than that. The gas saturation is constrained to a value of zero upon taking a negative value after data assimilation. The water and gas saturations from the updated state vector (unconstrained solution) are analyzed simultaneously for any constraint violations and constraints are enforced together.

**Table 4.3:** 2D black-oil model - standard deviation values for different types of production data

| Type of observation | Standard deviation |
|---|---|
| Oil production rate | 4.0 STB/day |
| Producing gas-oil ratio | 0.003 MSCF/STB |
| Water production rate | 0.0001 STB/day or 4.0 STB/day |
| Water injection rate | 5.0 STB/day |




(a) OPR of PROD1 (b) OPR of PROD2 (c) OPR of PROD3

(d) OPR of PROD4 (e) WIR of INJ (f) WPR of PROD1

(g) WPR of PROD2 (h) WPR of PROD3 (i) WPR of PROD4

(j) GOR of PROD1 (k) GOR of PROD3 (l) GOR of PROD4

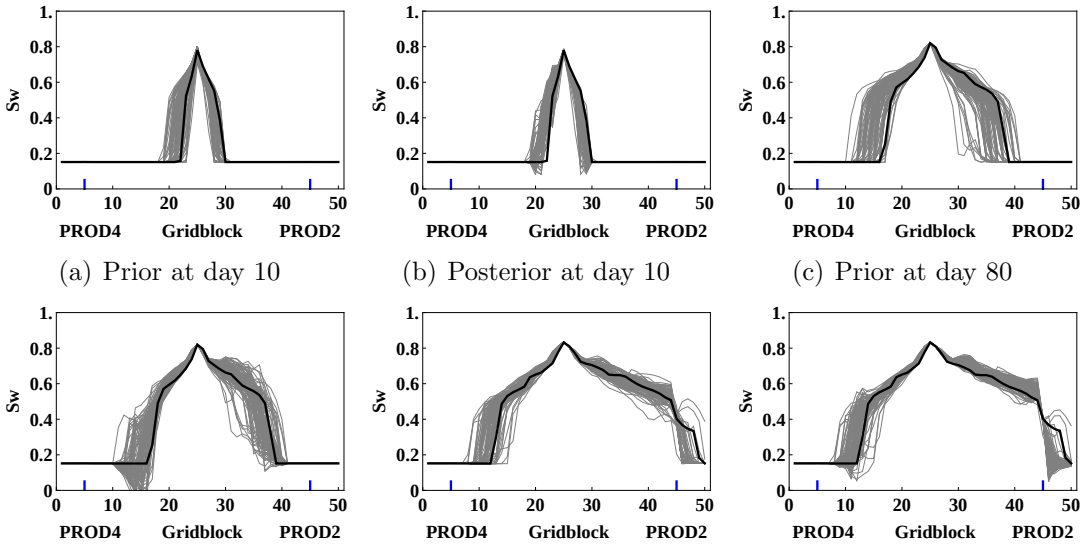
**Figure 4.29:** The production forecasts for the oil production rate (OPR), water injection rate (WIR), water production rate (WPR), and gas-oil ratio (GOR) from the initial ensemble and the reference production data. The gray curves denote the ensemble members and the black curve is the reference.

### 4.6.3 Results and discussion

#### *4.6.3.1 Estimation of saturation fields*

Figs. 4.30 and 4.31 show the prior and the posterior saturation profiles of water and gas, respectively, on the cross-section which passes through producers, PROD2 and PROD4, at different assimilation times for the standard EnKF. The gray curves are the ensemble members, the black curve shows the saturation profile for the reference model at the corresponding time, and the vertical blue lines indicate the locations of producers, PROD2 and PROD4. The injector, INJ, will be located at the center along this diagonal. The water and gas saturation profiles updated at the first assimilation time (day 10) are nearly similar to their corresponding forecast saturation profiles. The spread of the posterior ensembles, however, has been reduced as a result of the data assimilation. As we can see in Fig. 4.30(b), there are some ensemble members for which the updated water saturations are out of the physical bounds (less than 0.15) for gridblocks around the injector. A similar observation can be made in Fig. 4.31(b) where the updated gas saturations for some ensemble members are less than zero in the regions between the producers and the injector. The magnitude of constraint violations during the initial assimilation time, however, is small. The water saturation profiles updated at day 80 (Fig. 4.30(d)) show strong constraint violations for gridblocks located between the producer, PROD4, and the injector. The updated water saturations for some ensemble members are significantly lower than the minimum plausible value. The water saturations updated at day 160 (Fig. 4.30(f)) do not resemble the water saturation profile from the reference model for gridblocks near the producer PROD2, and show strong constraint violations. The gas saturation profiles updated at day 160 (Fig. 4.31(f)) also show strong constraint violations for gridblocks located between the injector and the producer PROD2.

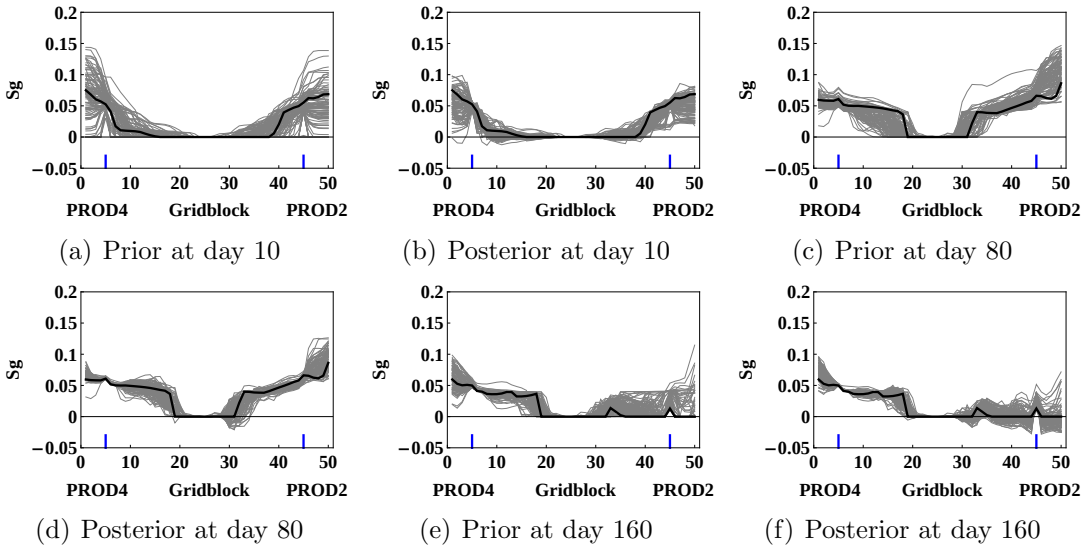
Figs. 4.32 and 4.33 show the prior and the posterior saturation profiles of water




(a) Prior at day 10 (b) Posterior at day 10 (c) Prior at day 80

(d) Posterior at day 80 (e) Prior at day 160 (f) Posterior at day 160

**Figure 4.30:** The water saturation profiles on the diagonal of producers PROD2 and PROD4 at different assimilation times from the standard EnKF. PROD2 is located at gridblock (45,45) and PROD4 is located at gridblock (5,5). The gray curves are the ensemble members and the black curve is the reference.



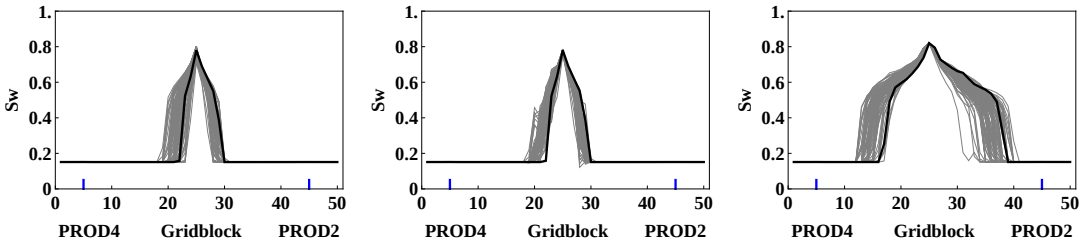

(a) Prior at day 10 (b) Posterior at day 10 (c) Prior at day 80

(d) Posterior at day 80 (e) Prior at day 160 (f) Posterior at day 160

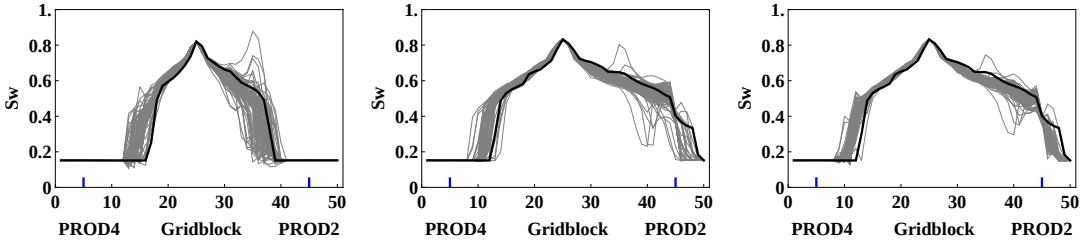
**Figure 4.31:** The gas saturation profiles on the diagonal of producers PROD2 and PROD4 at different assimilation times from the standard EnKF. PROD2 is located at gridblock (45,45) and PROD4 is located at gridblock (5,5). The gray curves are the ensemble members and the black curve is the reference.

and gas, respectively, along the same cross-section (PROD4-PROD2) at different assimilation times for the CEnKF. The gray curves are the ensemble members and the black curve shows the reference saturation profile at the corresponding time. The saturation profiles are updated iteratively in the CEnKF by taking into account the additional constraints on water and gas saturations identified during each iteration. As results indicate, the water and gas saturation profiles updated at different assimilation times from the CEnKF honor the inequality constraints. The ensembles of water and gas saturations from the CEnKF show enough variability and the reference saturation profile lies within the spread of the posterior ensemble. The water saturations updated at day 80 from the CEnKF are reasonable and show similar behavior as the reference water saturation profile. Unlike the standard EnKF (Fig. 4.30(f)), the water saturation profiles updated at day 160 from the CEnKF honor the constraints and the posterior water saturation profiles are similar to the reference saturation profile in the region near producer PROD2. As our emphasis was on the efficiency, we used a small number of iterations for the CEnKF method. Therefore, the updated water and gas saturation profiles obtained from the CEnKF show constraint violations for some ensemble members at a few number of gridblocks, although the magnitude of the constraint violation is not as severe as it was observed for the standard EnKF. Similar to the standard EnKF, the non-physical posterior water and gas saturations from the CEnKF are truncated to their values at the boundary of the feasible region before advancing the simulator to the next assimilation time.

Fig. 4.34 shows the water saturation fields, before and after data assimilation, for realization 68 at the third data assimilation time step (day 80) from the standard EnKF and the CEnKF together with the reference water saturation field at the same time. The posterior water saturation after updating with the standard EnKF is less than the connate water saturation (0.15) in some part of the reservoir between the producer, PROD4, and the injector, INJ (Fig. 4.34(c)). Because that is below the

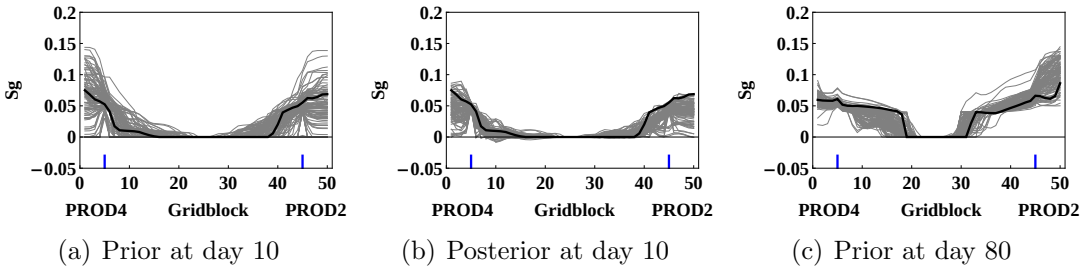



(a) Prior at day 10 (b) Posterior at day 10 (c) Prior at day 80

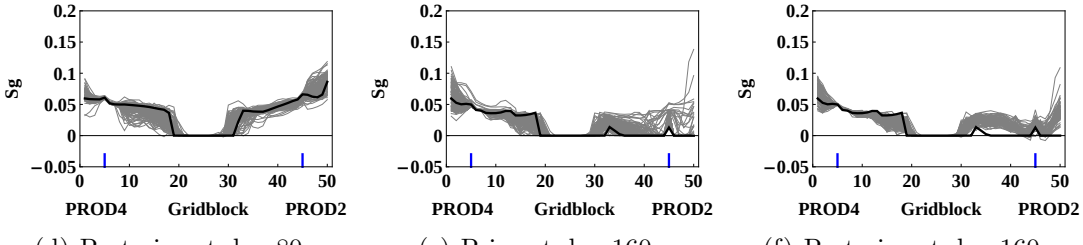



(d) Posterior at day 80 (e) Prior at day 160 (f) Posterior at day 160

**Figure 4.32:** The water saturation profiles on the diagonal of producers PROD2 and PROD4 at different assimilation times from the CEnKF. PROD2 is located at gridblock (45,45) and PROD4 is located at gridblock (5,5). The gray curves are the ensemble members and the black curve is the reference.



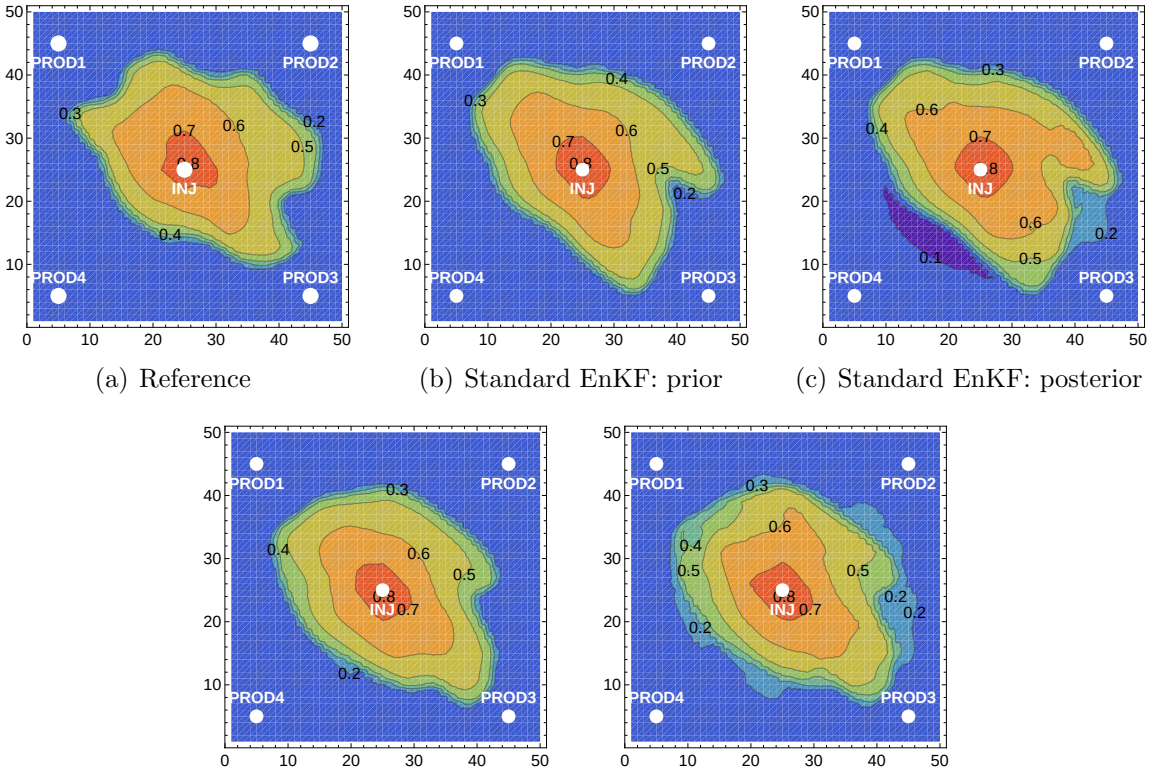

(a) Prior at day 10 (b) Posterior at day 10 (c) Prior at day 80




(d) Posterior at day 80 (e) Prior at day 160 (f) Posterior at day 160

**Figure 4.33:** The gas saturation profiles on the diagonal of producers PROD2 and PROD4 at different assimilation times from the CEnKF. PROD2 is located at gridblock (45,45) and PROD4 is located at gridblock (5,5). The gray curves are the ensemble members and the black curve is the reference.

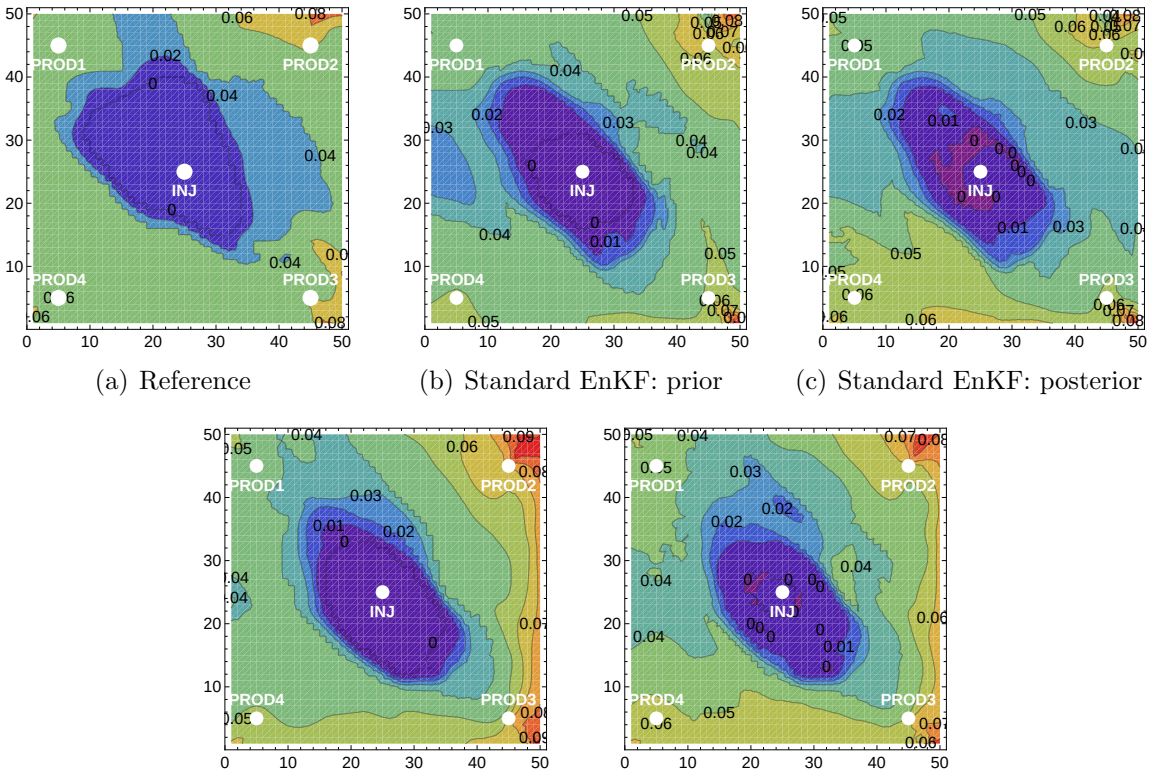
minimum allowable level, all such water saturations in the model would be set to the connate water saturation in a standard EnKF implementation. Although truncation handles the issue of constraint violation, it ignores correlations between other state variables at the same location, or the state variables in the remaining parts of the model. The application of the CEnKF method ensures that the inequality constraints on the water and gas saturations are honored at each data assimilation time step. The posterior water saturations obtained from the CEnKF (Fig. 4.34(e)) are above the connate water saturation (0.15) in all parts of the reservoir. The constraints on phase saturations are not honored during data assimilation using the standard EnKF. These posterior phase saturations, truncated when they are out of physical bounds, are used during the subsequent forecast step. The inconsistent updating of phase saturations can result in wrong forecast for the standard EnKF as shown in Fig. 4.34(b) which indicates that, the water is moving faster between the injector, INJ, and the producer, PROD1, and is moving slowly between the injector, INJ, and the producer, PROD3, compared to the reference model. As phase saturations are updated in a consistent manner during each data assimilation step using the CEnKF, it appears that, the prior and the posterior water saturation fields at day 80 for realization 68 obtained from the CEnKF are able to capture some of the features of the reference water saturation at the same time. The movement of the waterflood-front between the injector, INJ, and the producer, PROD3, from the CEnKF (Fig. 4.34(e)) is similar to the reference model (Fig. 4.34(a)). Fig. 4.35 shows the gas saturation fields, before and after data assimilation, for realization 68 at the third assimilation time using the standard EnKF and the CEnKF together with the reference gas saturation field at the same time.




(a) Reference (b) Standard EnKF: prior (c) Standard EnKF: posterior

(d) CEnKF: prior (e) CEnKF: posterior

**Figure 4.34:** The water saturation fields, before and after data assimilation, for realization 68 at day 80 (third assimilation time) from the standard EnKF and the CEnKF together with the corresponding reference water saturation field.



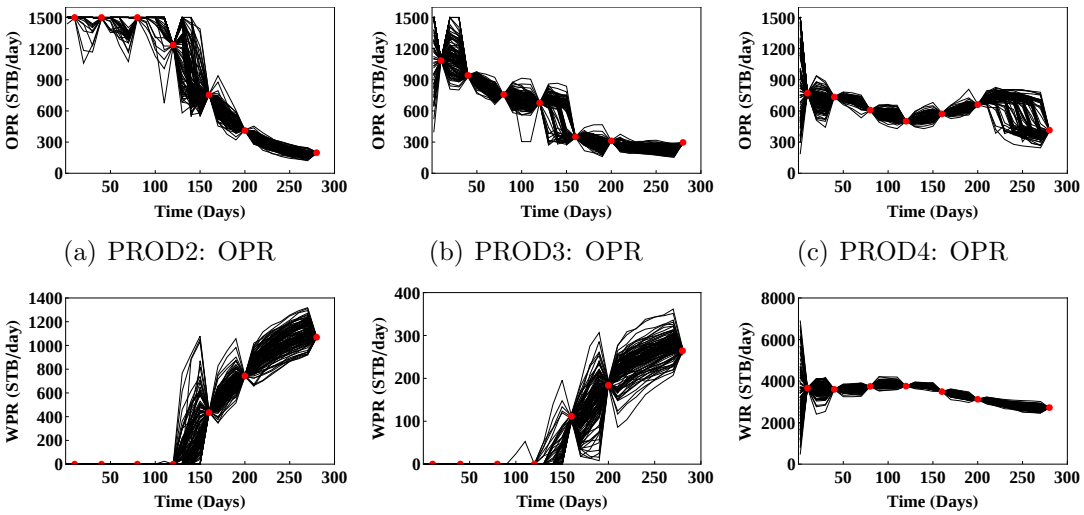

(a) Reference (b) Standard EnKF: prior (c) Standard EnKF: posterior

(d) CEnKF: prior (e) CEnKF: posterior

**Figure 4.35:** The gas saturation fields, before and after data assimilation, for realization 68 at day 80 (third assimilation time) from the standard EnKF and the CEnKF together with the corresponding reference gas saturation field.

#### *4.6.3.2 Matching production data*

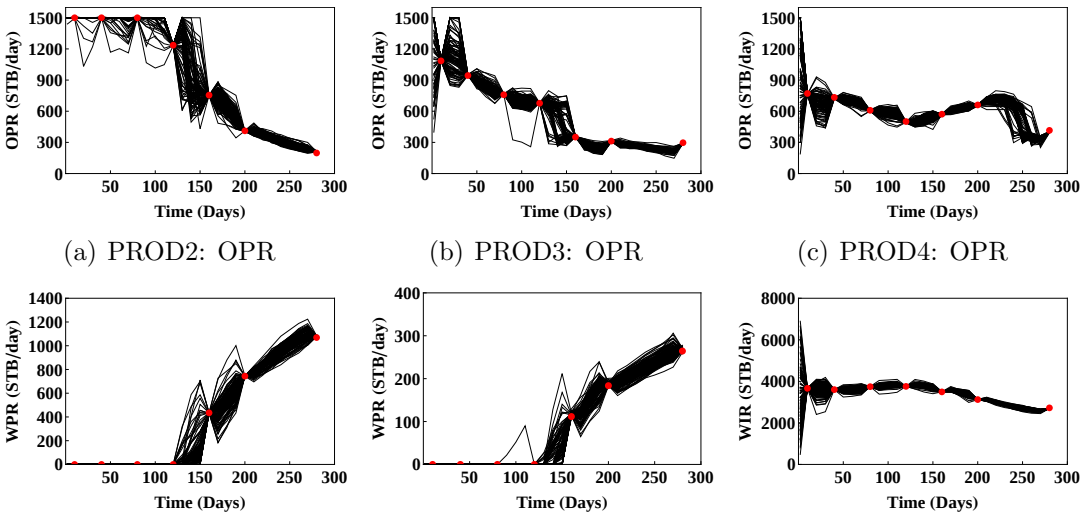
The static model parameters (porosity and permeability) and the dynamic state variables (pressure and phase saturations) are updated sequentially in time using the production data. Although the observations for data assimilation at each assimilation time include oil production rate, producing gas-oil ratio, and water production rate for all four producers, and water injection rate for the injector, Figs. 4.36 and 4.37 show forecast of different types of production data (not all) during the course of the data assimilation process for the standard EnKF and the CEnKF, respectively. The black curves indicate the forecast from the ensemble members and the red dots denote the actual observations from the reference model. We have shown earlier that the updated water and gas saturations at each data assimilation time from the standard EnKF show strong violations of the inequality constraints. The updated saturations out of the physical bounds are truncated to their values at the boundary of the feasible region. The production forecasts at each assimilation time from the standard EnKF, however, suffer from the poor adjustments to phase saturations due to truncation. As a result, the production forecasts from the standard EnKF show poor improvements (Figs. 4.36(c), 4.36(d), and 4.36(e)) and large variability even after data assimilation step 6 (day 200). The inequality constraints on the water and gas saturations are honored and the adjustments to phase saturations are made in a consistent manner during the application of the CEnKF method. The consistency in updating phase saturations at each data assimilation time helps improve the production forecasts from the CEnKF. The production forecasts for the oil production rate of producer PROD4 (Fig. 4.37(c)), and the water production rates of producers PROD2 (Fig. 4.37(d)) and PROD3 (Fig. 4.37(e)) from the CEnKF show significant improvements compared to the standard EnKF.




(a) PROD2: OPR (b) PROD3: OPR (c) PROD4: OPR

(d) PROD2: WPR (e) PROD3: WPR (f) INJ: WIR

**Figure 4.36:** The production forecasts for oil production rate (OPR), water production rate (WPR), and water injection rate (WIR) for different producers and injector from the standard EnKF during the course of the data assimilation. The black curves denote the ensemble members and the red dots denote the true observations.




(a) PROD2: OPR (b) PROD3: OPR (c) PROD4: OPR

(d) PROD2: WPR (e) PROD3: WPR (f) INJ: WIR

**Figure 4.37:** The production forecasts for oil production rate (OPR), water production rate (WPR), and water injection rate (WIR) for different producers and injector from the CEnKF during the course of the data assimilation. The black curves denote the ensemble members and the red dots denote the true observations.

#### *4.6.3.3 Predictability of the updated ensemble*

The ensemble of 100 realizations was continually adjusted by assimilating 13 production data at each data assimilation time step. We had 7 data assimilation time steps and thus, a total of 91 observations were assimilated during the entire history matching process. Once the data assimilation was complete over all the data assimilation time steps, the final updated ensemble of porosity and permeability, together with the initial reservoir equilibrium conditions, was evaluated from the initial time (day 0) up to the end of the production period (day 520) using a commercial reservoir simulator.

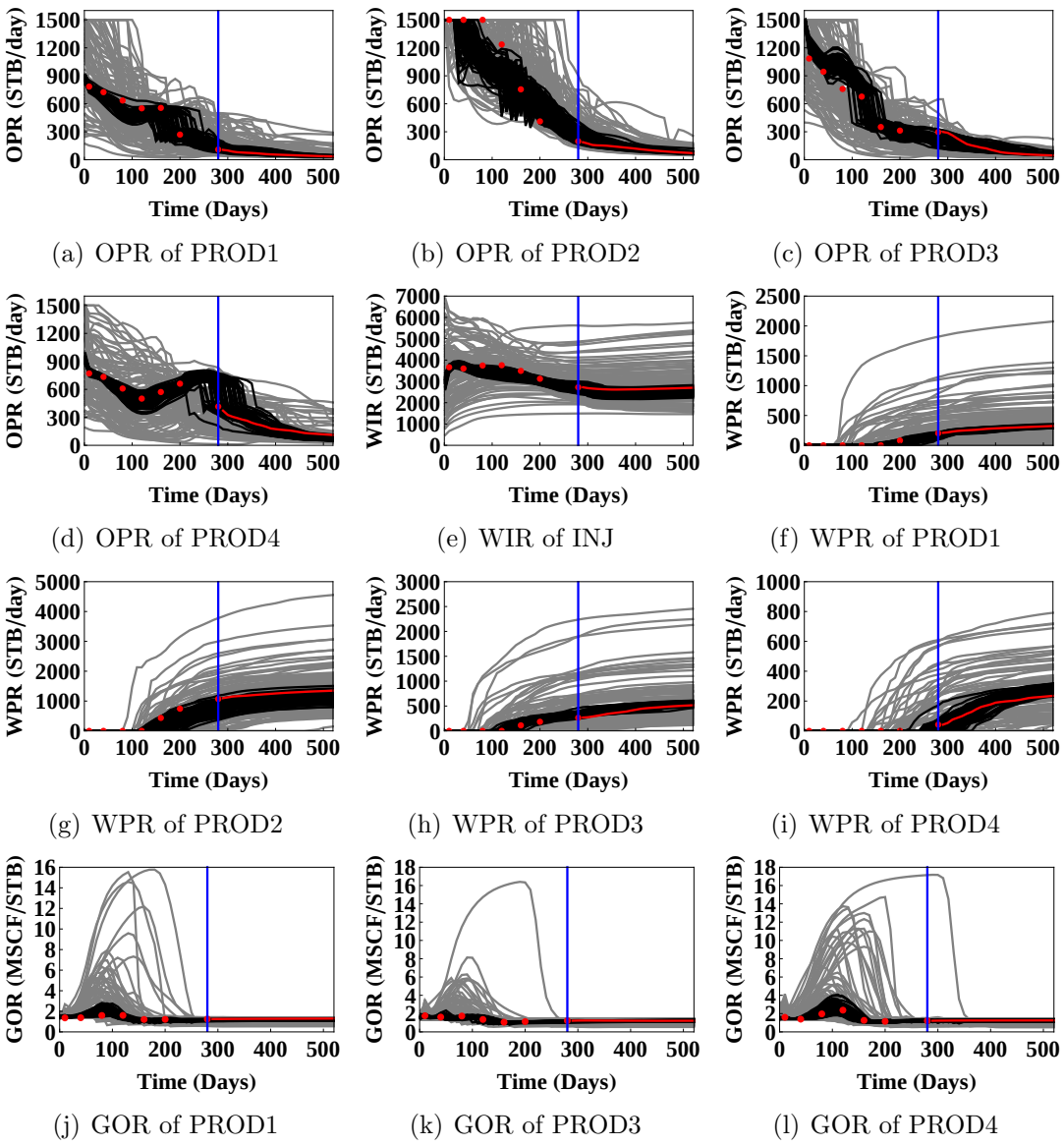
Figs. 4.38 and 4.39 show the results of the production data obtained by rerunning the final updated ensembles from the standard EnKF and the CEnKF methods, respectively. The results include the oil production rate (OPR), water production rate (WPR), and producing gas-oil ratio (GOR) for different producers and the water injection rate (WIR) for the injector. The gray curves represent the initial ensemble, the black curves represent the updated ensemble, the red dots represent the observations from the reference model, the red curve denotes the production data from the reference model in the prediction period, and the vertical blue line indicates the end of the history matching period (day 280). The updated ensemble from the standard EnKF has large variability in the history matching period and shows a poor match to the true observations of oil production rate data for all four producers (Figs. 4.38(a) through 4.38(d)). Similarly, a large fraction of the updated realizations from the standard EnKF show a delay in water-breakthrough at the producer PROD2 compared to the reference model. It is evident that the proposed CEnKF method provided improved history matching results compared to the standard EnKF. The updated ensemble from the CEnKF was able to match water-breakthrough times for different producers (Figs. 4.39(f) through 4.39(i)) with good accuracy. The updated ensemble obtained from the CEnKF performs fairly well in the prediction period (day 280

to day 520) for different types of production data. The predictions of oil and water production rates for the producer, PROD4, however, are biased compared to the reference model.

## 4.7 Mass conservation

The most significant problem that was observed with the application of the standard EnKF to the compositional data assimilation problem was that, excess mass was added during the update step at the first data assimilation time. In general, mass should not be expected to be conserved in data assimilation or history matching, as the mass of a component is typically an unknown quantity that must be estimated. On the other hand, when a substance is injected into the reservoir at a known rate, it would be desirable for that quantity to be conserved during the updating process. In this section, we will investigate experimentally the consequences of the analysis step on mass conservation in a simple model.

We consider 1-D, constant-rate injection of water into a reservoir of length 50 that is initially at irreducible water saturation (0.20). The porosity is assumed to be a correlated random field with mean 0.25 and standard deviation of 0.07. The covariance for porosity is nearly Gaussian with a practical range of 25. At some particular time (560 in dimensionless units), we measure the saturation at the center of the reservoir. Fig. 4.40(a) shows the distribution of forecast water saturation from a large number of reservoir realizations at the measurement time. If the true water saturation at $x = 25$ is 0.2, then the result of a standard EnKF analysis step is as shown in Fig. 4.40(b). Approximately, 10% of the updated realizations contain negative saturations after updating. Fig. 4.40(c) shows the same saturation profiles after updating using the CEnKF method. Constraining the saturation values to lie in the feasible region is not sufficient to ensure that the profiles are monotonic and the advantage over truncation is not obvious because in the two-phase incompressible model, the problems with



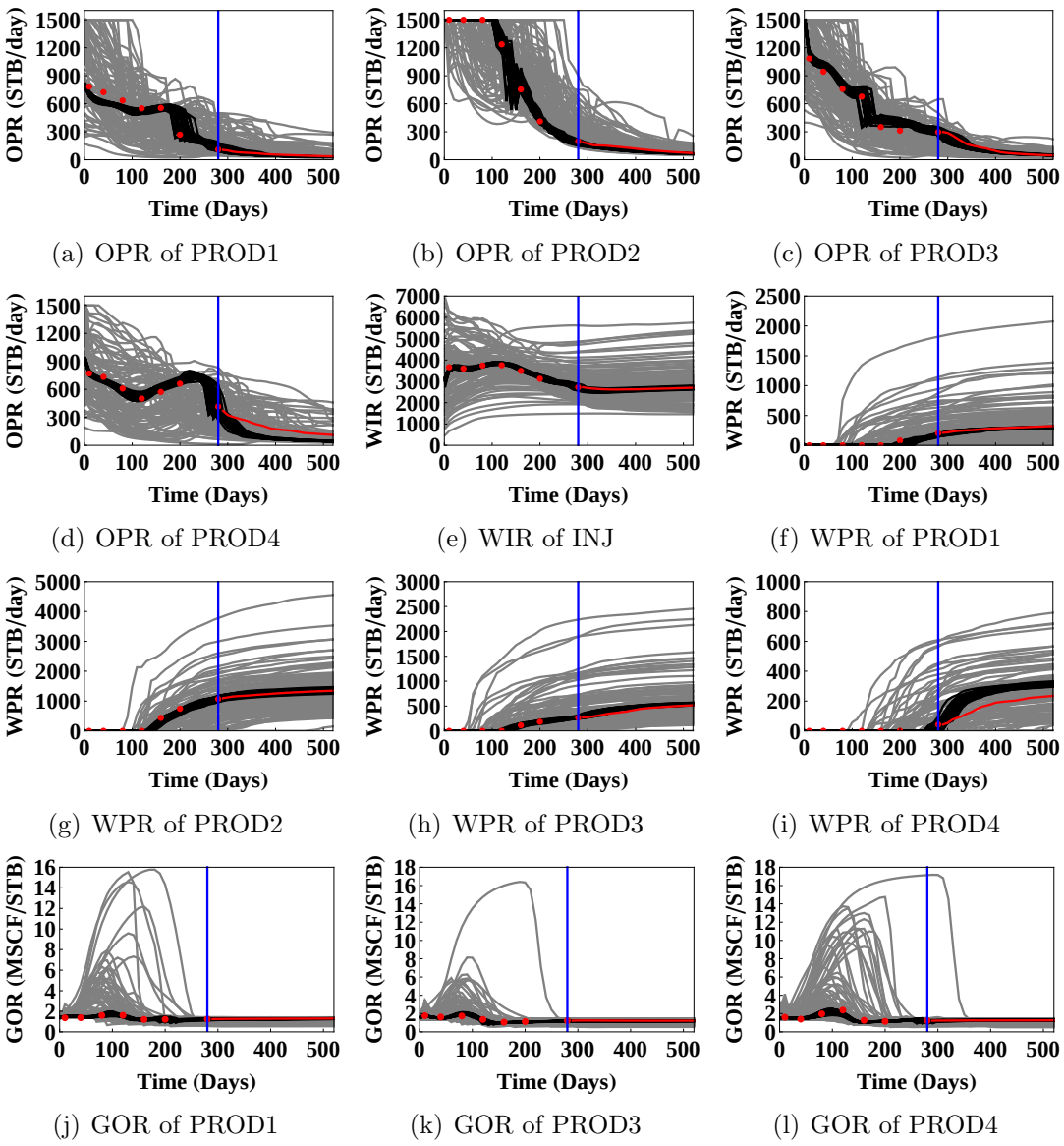

(a) OPR of PROD1 (b) OPR of PROD2 (c) OPR of PROD3

(d) OPR of PROD4 (e) WIR of INJ (f) WPR of PROD1

(g) WPR of PROD2 (h) WPR of PROD3 (i) WPR of PROD4

(j) GOR of PROD1 (k) GOR of PROD3 (l) GOR of PROD4

**Figure 4.38:** The production data predicted from the ensemble of porosity and permeability updated at day 280 by the standard EnKF. The gray curves represent the initial ensemble, the black curves represent the updated ensemble, the red dots represent the observations from the reference model, the red curve denotes the production data from the reference model in the prediction period, and the vertical blue line indicates the end of the history matching period (day 280).




(a) OPR of PROD1 (b) OPR of PROD2 (c) OPR of PROD3

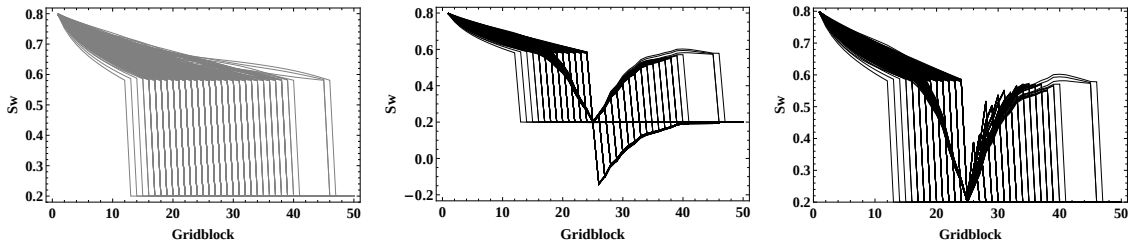
(d) OPR of PROD4 (e) WIR of INJ (f) WPR of PROD1

(g) WPR of PROD2 (h) WPR of PROD3 (i) WPR of PROD4

(j) GOR of PROD1 (k) GOR of PROD3 (l) GOR of PROD4

**Figure 4.39:** The production data predicted from the ensemble of porosity and permeability updated at day 280 by the CEnKF. The gray curves represent the initial ensemble, the black curves represent the updated ensemble, the red dots represent the observations from the reference model, the red curve denotes the production data from the reference model in the prediction period, and the vertical blue line indicates the end of the history matching period (day 280).

volume conservation are not so significant as in the compositional model. Truncation of negative water saturations results in a corresponding reduction in the updated oil saturation, which does not occur in the compositional model.




(a) Forecast (b) Standard EnKF (c) Constrained EnKF

**Figure 4.40:** Water saturation profiles for the 1D reservoir example

While the total volume will not be effected by truncation in this case, it certainly affects the volume of the individual phases. Fig. 4.41 compares the volume of moveable water (water whose saturation is greater than the irreducible water saturation of 0.2) at the end of the forecast with the volume of moveable water after EnKF with truncation and with the CEnKF method. The small variability in the volume at the end of the forecast is a result of discretization of the reservoir and the use of gridblock-centered saturation values. The standard EnKF with truncation shows a bias towards higher water volume because the negative values are truncated. The CEnKF technique does not remove the bias, but does reduce the problem substantially.

If mass is explicitly included in the state vector, and all forecast realizations have the same mass, then the mass in all analyzed realizations will also be identical. The mass of a component in each gridblock must of course be non-negative, but mass will not necessarily be non-negative after an EnKF update. The consequence of truncating negative values of mass is that mass will not be conserved in the analysis step. Application of constraints to the mass variable would necessarily conserve the mass of each component if all realizations initially contained the same mass.

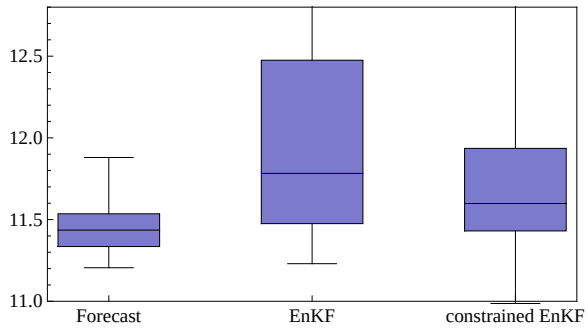
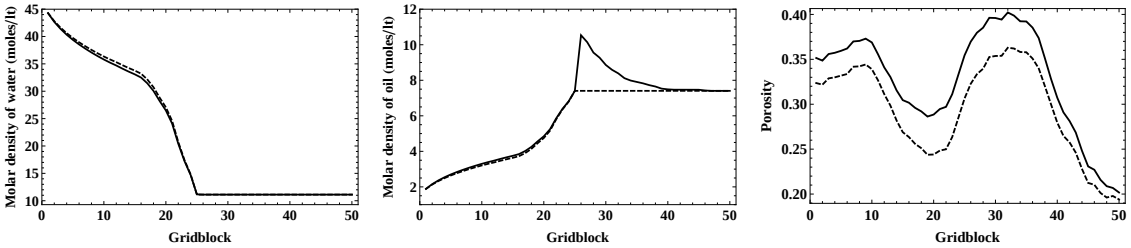



**Figure 4.41:** Distributions of volume of moveable water in the reservoir at the end of several operations: the forecast step, the standard EnKF update with truncation of saturations, and the CEnKF.

In general, however, the variables that are included in the state vectors are saturations, porosities, molar densities, or mass fractions. When saturations or molar densities are used in the state vector, mass is not necessarily conserved even when truncation is not used. When truncation is used, the results are typically much worse. Fig. 4.42 shows results from the same Buckley-Leverett displacement experiment whose results were shown in Fig. 4.40. In this case, however, the molar densities and porosities are included in the state vector instead of saturations and porosities. One consequence of this choice of parameterization is that, unlike saturation, there is no simple upper limit on the molar density. Note that truncation of the low values of molar density of water (solid curve in Fig. 4.42(a)), does not effect the molar density of oil (solid curve in Fig. 4.42(b)). Also, truncation of the molar density of water has no effect on the porosity field (solid curve in Fig. 4.42(c)). As a result, there is 9% more total moles of both components in the realization after truncation than in the realization after the analysis step. If the CEnKF method is used, the molar density of water is quite similar to the molar density after truncation (dashed curve in Fig. 4.42(a)) while the molar density of oil and the porosity field are both reduced substantially (dashed curves in Figs. 4.42(b) and 4.42(c)). The net effect is that the total number of moles is only 2% less than the forecast, so the mass is much more

nearly conserved, even though the mass (number of moles) was not used directly in the state vector.




(a) Molar density of water (b) Molar density of oil (c) Porosity

**Figure 4.42:** Comparison of updated molar densities and porosity for realization 1, from EnKF with truncation (solid) and from CEnKF (dashed).

## 4.8 Summary

In this chapter, we proposed an efficient method for sequential data assimilation that ensures that the updated state variables satisfy inequality constraints. The method is a constrained version of the EnKF that does not require additional model evaluations for constraint application. The validity of the CEnKF method is demonstrated through its application to a synthetic 1D linear model. The second test problem shows the application of the CEnKF technique to a data assimilation problem of a compositional model for which state variables (molar densities of components) have lower limits. The constraints were applied on the $CO_2$ molar density which helped reduce molar densities of other components in a systematic manner. The updating of molar densities of all components in a consistent manner from the CEnKF technique resolved the issues of excess mass and pressure build-up which were observed during the application of the standard EnKF with truncation. In case of black-oil reservoir models involving three-phase flow conditions, adjusting phase saturations, when the updated saturations violate the inequality constraints, is not straightforward. The CEnKF technique presents a viable option for handling three-phase flow conditions such that the updated phase saturations are plausible. The application of

the CEnKF method to a three-phase black-oil model clearly showed the benefits of the constrained solution and obtained a good match to the historical production data. We also showed that in cases where mass conservation of a component is important, the CEnKF gives superior results to the standard truncation scheme.

The proposed CEnKF method can easily be implemented with some modifications to the pre-existing codes for the standard EnKF. Although the CEnKF method is an iterative process, the additional computational cost and time required during each iteration is significantly lower compared to the cost of running the reservoir simulator. Because of the small computational cost and low data-storage requirement, the implementation of the CEnKF method on a real-field reservoir problem with few hundred thousands of gridblocks should not impose any computational or data-storage challenge.

# CHAPTER V

# INTERIOR-POINT METHOD FOR CONSTRAINED ENKF

## 5.1 Introduction

Although a number of attempts have been made for incorporating the inequality constraints into the EnKF framework, most of the techniques depend on the treatment of the unconstrained solution obtained from the standard implementation of the EnKF. None of the previous studies shows the direct involvement of inequality constraints in the mathematical formulation of the EnKF. In Chapter 4, we proposed a constrained form of the EnKF (CEnKF) for updating reservoir models for avoiding the problem of non-physical values of phase saturations and component molar densities. The effectiveness of the CEnKF technique was demonstrated on fluid flow problems (compositional and black-oil) with fairly large number of gridblocks. The CEnKF algorithm is iterative, however, and requires identification of the active constraints during each iteration. In reservoir engineering, the ultimate goal of history matching is to estimate large numbers ($10^5$ to $10^6$) of reservoir static model parameters and dynamic state variables. The application of the CEnKF method to a reservoir model with few hundred thousands of gridblocks might require a large number of iterations for finding a solution that completely lies in the feasible region. Although the computational cost and data-storage requirements for each iteration of the CEnKF method are not significant, the implementation of the CEnKF method with large numbers of iterations may not be a practical approach.

It was discovered in the 1980s that, the algorithms from the field of nonlinear

programming and nonlinear equations can be utilized for efficiently solving many large-scale linear programming problems. One of the important characteristics of these iterative methods was that they required all iterates to strictly satisfy the inequality constraints on the elements of the solution. As a result, these methods became known as interior-point methods. The interest in the research community towards the interior-point method and its application did not increase until 1984 when Karmarkar (1984) presented a successful application of a new interior-point algorithm for linear programming. The algorithm proposed by Karmarkar (1984) was based on nonlinear projective transformations and he reported solution times up to 50 times faster than the simplex method whose computational cost increases exponentially with the problem dimensions. Frisch (1955) proposed the first interior-point method in the form of a logarithmic barrier method that was later extensively analyzed by Fiacco and McCormick (1968) to solve general nonlinearly inequality constrained problems.

The active-set method has been the widely used approach for handling inequality-constrained optimization problems and is suited for small- and medium-sized problems (Nocedal and Wright, 2006). The interior-point method is another branch of algorithms for solving the inequality-constrained optimization problems. It does not require the direct identification of the active constraints and is well suited for large problems. Interior-point methods are usually classified into three main categories (Quintana et al., 2000): (i) projective methods, (ii) affine-scaling methods, and (iii) primal-dual methods. The algorithm proposed by Karmarkar (1984) falls under the category of projective methods. The affine-scaling methods, which were simple and computationally less complex, were proposed as simplifications of projective methods. The primal-dual methods can further be classified as path-following methods and potential reduction methods. Gill et al. (1986) showed a formal equivalence between Karmarkar's projective method and the classical logarithmic barrier method applied

to linear programming. Therefore, the terms "interior-point method" and "barrier method" are now used interchangeably. The first theoretical results for primal-dual path-following methods were presented by Megiddo (1986) who proposed to apply a logarithmic barrier method to the primal and the dual problems simultaneously.

The interior-point method is based on constructing a "barrier" which prevents violation of all the inequality constraints such that the resulting sequence of the iterates remains strictly in the interior of the feasible region defined by the inequality constraints (Baldick, 2008). The barrier is constructed by adding a term to the original objective function that increases rapidly as we approach the boundary of the feasible region from its interior. In the ideal scenario, the iterative solution to the optimization problem moves directly towards the minimizer across the interior of the feasible region, rather than stepping along its boundary as in the active set algorithm. In order to be useful for constrained optimization, a valid barrier function should be differentiable on the interior of the constraint set and should become unbounded as the solution approaches the boundary of the constraint set.

Incorporating fundamental features of a interior-point method into the EnKF framework naturally presents a way for conditioning an ensemble of static model parameters and dynamic state variables to historical production data and simultaneously honoring the inequality constraints on state variables during the analysis step of the EnKF. To our knowledge, we are not aware of any such method that combines a interior-point method with the EnKF for data assimilation. Recently, Bell et al. (2009) presented an application of a interior-point method to a nonlinear Kalman-Bucy smoother problem with inequality constraints. The authors showed the usefulness of the interior-point method by applying it to a low-dimensional nonlinear problem. The direct application of the method proposed by Bell et al. (2009) is not possible for data assimilation in reservoir simulation models with few hundred thousands of gridblocks, however, as it requires the explicit computation and handling of

large-size matrices similar to the covariance matrix of state variables.

In this chapter, we present a new interior-point algorithm for incorporating the inequality constraints into the EnKF methodology (IPCEnKF). The present study draws the basic idea from the earlier work by Bell et al. (2009) on Kalman smoother. The original objective function involved in the data assimilation framework is reformulated by adding a barrier function which takes into account the proximity of state variables, subject to inequality constraints, from the boundaries of the feasible region, and sets a barrier on state variables against leaving the feasible region. By doing so, the original constrained problem is transformed into an unconstrained problem. The solution of the resulting unconstrained problem is presented in the form of a new iterative EnKF scheme which implicitly contains inequality constraints on state variables. The effectiveness of the newly proposed IPCEnKF method is demonstrated through its application to a synthetic 1D linear example and a highly heterogeneous 3D, three-phase black-oil reservoir fluid flow problem. The results indicate the effectiveness of the newly proposed IPCEnKF method in terms of matching the observations and honoring the inequality constraints on state variables. The proposed method is able to achieve a better estimate of state variables than is obtained using only truncation with the standard implementation of the EnKF.

## 5.2 Fundamental concepts

The standard implementation of the EnKF is an optimization problem where the analysis step computes variables that approximately minimize the following objective function

$$J_j(\mathbf{y}) = \frac{1}{2}(\mathbf{y} - \mathbf{y}_j^f)^T(\mathbf{C}_Y^f)^{-1}(\mathbf{y} - \mathbf{y}_j^f) + \frac{1}{2}(\mathbf{H}\mathbf{y} - \mathbf{d}_{\mathrm{obs},j})^T\mathbf{C}_D^{-1}(\mathbf{H}\mathbf{y} - \mathbf{d}_{\mathrm{obs},j}).$$

Since the forecast step of the EnKF uses a reservoir simulator for evolving reservoir models forward in time, the physical constraints on the elements of state vector are honored during the forecast step. The analysis step of the EnKF, however, does not

involve any mechanism for enforcing constraints on state variables. As a result, the state variables can take non-physical values when data are assimilated using the standard EnKF. Under the interior-point formulation, the original objective function of an optimization problem involving constraints, similar to the EnKF, is often reformulated by considering additional term(s). The additional term reflects the constraints and its form depends on the nature of the constraint function. By doing so, explicit handling of the constraints can be avoided and they can be easily treated within the optimization algorithm. In general, the process of finding a minimum (or a maximum) of a given objective function strongly conflicts with the process of satisfying the given constraints and vice-versa. The interior-point formulation exploits this confliction between the two processes. A general observation is that the difficulty in honoring constraints increases (generally more than linearly) with increase in the number of constraints. When a solution that minimizes (or maximizes) the objective function is obtained, and the elements of the solution honor all the constraints on them, the difficulty in the optimization process is less. In a different scenario, it is possible to find a solution that minimizes (or maximizes) the objective function, however, some of the elements of the resulting solution may fail to honor the constraints on them. Under such scenario, the difficulty in the optimization process will be considerably higher. By taking advantage of this behavior, one can incorporate the additional term into the objective function that penalizes the objective function at points that do not satisfy the constraints. There are two basic approaches for incorporating additional term(s) in the objective function. The additional term can be incorporated in the form of a penalty function which penalizes the objective function for values of the solution that violate the constraints. The penalty function can be used in the form of either a partial penalty function where a penalty is applied near the boundary of the feasible region or a global penalty function which is applied throughout the infeasible region. An alternate way of incorporating additional term into the objective function

is to use a barrier function, which is often used for inequality-constrained problems. This approach involves adding a function (often called as barrier) that grows rapidly as the solution approaches the boundary of the feasible region from the interior.

Generally, the interior-point method can be implemented in two forms. The first form is the primal interior-point method which is also known as the primal barrier approach. Under this formulation, the minimization problem is written as

$$\begin{aligned} &\min_{\mathbf{x}} S(\mathbf{x}) \\ \text{subject to} \quad &c(\mathbf{x}) \geq 0 \end{aligned} \tag{5.1}$$

where, $c(\mathbf{x})$ is a vector containing all the scalar inequality constraints $c_j(\mathbf{x}) \geq 0$, $j \in [1, N_{\mathbf{x}}]$. If phase saturation at each grid location is used as a variable in the optimization problem, it is important to have phase saturations that are bounded within an upper and a lower bound at each location at all times. Under such scenario, $N_{\mathbf{x}}$ represents the dimension of the variable $\mathbf{x}$. If the variable $\mathbf{x}$ is under physical constraints as $\mathbf{x}_{\min} \leq \mathbf{x}_j \leq \mathbf{x}_{\max}$, then for each element $j$, the constraint $c_j(\mathbf{x})$ can be expressed as $c_j(\mathbf{x}) = (\mathbf{x}_{\max} - \mathbf{x}_j)(\mathbf{x}_j - \mathbf{x}_{\min})$. In reservoir engineering related applications, porosity and phase saturations are some of the variables that are subject to similar constraints as mentioned here.

To solve the minimization problem with multiple inequality constraints as shown in Eq. 5.1, a barrier function, $f_b(\mathbf{x})$, is added to the original objective function, $S(\mathbf{x})$. The modified objective function is denoted by $O(\mathbf{x})$ and can be expressed as

$$O(\mathbf{x}) = S(\mathbf{x}) + t f_b(\mathbf{x}), \tag{5.2}$$

where, the barrier function $f_b(\mathbf{x})$ prevents the model variables from leaving the feasible region. In Eq. 5.2, $t$ is called the barrier parameter. The two most widely used forms of the barrier functions are the reciprocal function and the negative of the logarithmic function. The logarithmic function is also called as the log-barrier function. In the limit as $t$ goes to 0, the minimizer of Eq. 5.2 approaches the minimizer of Eq. 5.1

for points away from the barrier. In this study, we use the log-barrier function with positive barrier parameter. The log-barrier function can be expressed as

$$f_b(\mathbf{x}) = -\sum_{j=1}^{N_\mathbf{x}} \log c_j(\mathbf{x})\,. \tag{5.3}$$

In case of the model variable with physical constraints of $\mathbf{x}_{\min} \leq \mathbf{x}_j \leq \mathbf{x}_{\max}$, the barrier function can be written as

$$f_b(\mathbf{x}) = -\sum_{j=1}^{N_\mathbf{x}} \log(\mathbf{x}_j - \mathbf{x}_{\min}) - \sum_{j=1}^{N_\mathbf{x}} \log(\mathbf{x}_{\max} - \mathbf{x}_j)\,. \tag{5.4}$$

By introducing the barrier function into the original objective function, the constrained optimization problem with inequality constraints is converted into an unconstrained optimization problem. The modified objective function $O(\mathbf{x})$ is differentiable with continuous partial derivatives. The first order necessary condition for minimizing the objective function $O(\mathbf{x})$ is

$$\nabla_\mathbf{x} S(\mathbf{x}) + t\nabla_\mathbf{x} f_b(\mathbf{x}) = 0\,. \tag{5.5}$$

If the objective function $O(\mathbf{x})$ is twice differentiable and $\mathbf{x}^*$ lies in the interior of the feasible region, then the second-order sufficient conditions for $\mathbf{x}^*$ to be a local minimizer of $O(\mathbf{x})$ include (Antoniou and Lu, 2007): (a) the gradient of $O(\mathbf{x})$ vanishes at $\mathbf{x}^*$, and (b) the Hessian of $O(\mathbf{x})$ is positive definite at $\mathbf{x}^*$. Newton's methods can be used to solve this unconstrained minimization problem. As the modified objective function involves the barrier parameter $t$ through the barrier function, the minimizer $\mathbf{x}^*$ is a function of the barrier parameter. Hence, a path or a sequence, $\mathbf{x}^*(t)$, can be obtained as $t$ is gradually decreased from the initial value to an acceptable small value close to zero. It should be noted that, for each fixed value of $t$, the minimizer $\mathbf{x}^*(t)$ is approached through several iterations until the convergence criteria are satisfied.

The ill-conditioning of the Hessian of the modified objective function is often considered as the root-cause of the shortcomings of the primal interior-point method. In this study, the primal interior-point method is introduced into the iterative EnKF

framework. Thus, even when the barrier parameter takes a small value, the Gauss-Newton method prevents the ill-conditioning since the approximated Hessian matrix is always positive definite.

To improve the conditioning of the Hessian matrix or lessen the nonlinearity of $O(\mathbf{x})$, another form of the interior-point method, called as the primal-dual interior-point method, is popular. In the primal-dual system, a new type of variable, called as the slack variable $p$, is introduced. The slack variable can be defined as $p_j = c_j(\mathbf{x})$, $j \in [1, N_{\mathbf{x}}]$. The modified objective function under the primal-dual interior-point formulation is written as

$$O(\mathbf{x}, p, z) = S(\mathbf{x}) - t \sum_{j=1}^{N_{\mathbf{x}}} \log p_j + (c(\mathbf{x}) - p)^T z, \tag{5.6}$$

where, $z$ is a vector of Lagrange multipliers, $z_j \geq 0$, $j \in [1, N_{\mathbf{x}}]$. The first order necessary condition for minimizing the objective function in Eq. 5.6 involves calculating the derivatives of Eq. 5.6 with respect to the three types of variables including $\mathbf{x}$, $z$, and $p$, and setting them equal to 0 as

$$\nabla_{\mathbf{x}} O(\mathbf{x}, p, z) = \nabla_{\mathbf{x}} S(\mathbf{x}) - A^T(\mathbf{x}) z = 0 \tag{5.7}$$

$$\nabla_z O(\mathbf{x}, p, z) = c(\mathbf{x}) - p = 0 \tag{5.8}$$

$$\nabla_p O(\mathbf{x}, p, z) = Pz - te = 0 \tag{5.9}$$

where, $P = \text{diag}(p_1,\ p_2,\ \ldots,\ p_{N_{\mathbf{x}}})$, $e = [1, 1, \ldots, 1]^T$, and $A^T(\mathbf{x})$ is the Jacobian matrix of the function $c(\mathbf{x})$. Eq. 5.9 is called the complementary slackness condition. The solution to the primal-dual interior-point formulation can be obtained using the Newton-Raphson method. The primal-dual method increases the dimensions of the solution vector, and the subsequent computational cost due to the introduction of slack variables and Lagrange multipliers is significant. Therefore, in the present study, we investigate the primal interior-point algorithm, instead of the primal-dual method, within the EnKF framework.

## 5.3 Formulation

As Oliver et al. (2008) pointed out, a sample $\mathbf{x}$ from the posterior distribution conditioned to observations, can be generated as the value that minimizes the following objective function

$$S_i(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_{pr,i})^T \mathbf{C}_X^{-1}(\mathbf{x} - \mathbf{x}_{pr,i}) + \frac{1}{2}(g(\mathbf{x}) - \mathbf{d}_{\text{obs},i})^T \mathbf{C}_D^{-1}(g(\mathbf{x}) - \mathbf{d}_{\text{obs},i}), \quad (5.10)$$

where, $\mathbf{x}_{pr,i}$ denotes the $i$th ensemble member before data assimilation, $\mathbf{d}_{\text{obs},i}$ is the vector of perturbed observations for the $i$th ensemble member, $g(\cdot)$ is the nonlinear relationship between the model variables and the observations, $\mathbf{C}_X$ represents the prior covariance of model variables, and $\mathbf{C}_D$ denotes the observation-error covariance matrix. If the inequality constraints on the model variables are important, the objective function in Eq. 5.10 is reformulated by adding a term that reflects the constraints on the model variables, $\mathbf{x}$, for the $i$th ensemble member. The modified objective function for the $i$th ensemble member can be expressed as

$$O_i(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_{pr,i})^T \mathbf{C}_X^{-1}(\mathbf{x} - \mathbf{x}_{pr,i}) + \frac{1}{2}(g(\mathbf{x}) - \mathbf{d}_{\text{obs},i})^T \mathbf{C}_D^{-1}(g(\mathbf{x}) - \mathbf{d}_{\text{obs},i}) + t f_b(\mathbf{x}). \quad (5.11)$$

The formulation of the proposed IPCEnKF method presented in this section can be generalized for the $i$th ensemble member. Therefore, in order to reduce the number of superscripts and subscripts, we will omit the index for the ensemble member in the following equations. The gradient of the modified objective function with respect to model variables, $\mathbf{x}$, can be expressed as

$$\nabla_{\mathbf{x}} O(\mathbf{x}) = \mathbf{C}_X^{-1}(\mathbf{x} - \mathbf{x}_{pr}) + \mathbf{G}^T \mathbf{C}_D^{-1}(g(\mathbf{x}) - \mathbf{d}_{\text{obs}}) + t\nabla_{\mathbf{x}} f_b(\mathbf{x}), \quad (5.12)$$

where, $\mathbf{G} = (\nabla g^T)^T$ is the ensemble-based sensitivity which relates the observations to model variables. The Hessian of the modified objective function $O(\mathbf{x})$ is

$$\nabla_{\mathbf{x}}^2 O(\mathbf{x}) = \mathbf{C}_X^{-1} + \mathbf{G}^T \mathbf{C}_D^{-1} \mathbf{G} + (\nabla_{\mathbf{x}} \mathbf{G}^T) \mathbf{C}_D^{-1}(g(\mathbf{x}) - \mathbf{d}_{\text{obs}}) + t\nabla_{\mathbf{x}}^2 f_b(\mathbf{x}). \quad (5.13)$$

We consider the Gauss-Newton method for minimizing the objective function given in Eq. 5.11. Let $\mathbf{x}^l$ denotes the latest approximation to the minimum of the objective function shown in Eq. 5.11. Define $\delta\mathbf{x}$ such that,

$$\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^l.$$

Now, let us consider the second order Taylor series expansion of the objective function, $O(\mathbf{x})$, around $\mathbf{x}^l$ which can be given as

$$O(\mathbf{x}) \approx O(\mathbf{x}^l) + \nabla_{\mathbf{x}} O(\mathbf{x}^l)^T \delta\mathbf{x} + \frac{1}{2}\delta\mathbf{x}^T \nabla^2_{\mathbf{x}} O(\mathbf{x}^l)\delta\mathbf{x}, \tag{5.14}$$

where, $\nabla_{\mathbf{x}} O(\mathbf{x}^l)$ and $\nabla^2_{\mathbf{x}} O(\mathbf{x}^l)$ denote the gradient (Eq. 5.12) and the Hessian (Eq. 5.13), respectively, of the objective function evaluated at $\mathbf{x}^l$. Let $\tilde{O}(\mathbf{x})$ denotes the quadratic approximation of the objective function given in Eq. 5.14. Thus,

$$\tilde{O}(\mathbf{x}) = O(\mathbf{x}^l) + \nabla_{\mathbf{x}} O(\mathbf{x}^l)^T \delta\mathbf{x} + \frac{1}{2}\delta\mathbf{x}^T \nabla^2_{\mathbf{x}} O(\mathbf{x}^l)\delta\mathbf{x}. \tag{5.15}$$

The quadratic approximation $\tilde{O}(\mathbf{x})$ has a unique minimum if and only if the Hessian, $\nabla^2_{\mathbf{x}} O(\mathbf{x}^l)$, is positive definite. According to the first order necessary condition, the gradient of the objective function, $\tilde{O}(\mathbf{x})$, in Eq. 5.15 should vanish at the minimum. This condition results in

$$\nabla_{\mathbf{x}} \tilde{O}(\mathbf{x}) = \nabla_{\mathbf{x}} O(\mathbf{x}^l) + \nabla^2_{\mathbf{x}} O(\mathbf{x}^l)\delta\mathbf{x}^{l+1} = 0,$$

which can also be written as

$$\nabla^2_{\mathbf{x}} O(\mathbf{x}^l)\delta\mathbf{x}^{l+1} = -\nabla_{\mathbf{x}} O(\mathbf{x}^l). \tag{5.16}$$

In Eq. 5.16, the superscript $l$ denotes the iteration index and $\delta\mathbf{x}^{l+1}$ is the search direction for the $(l+1)$th iteration. As Eq. 5.16 indicates, in order to compute the search direction $\delta\mathbf{x}^{l+1}$, we need to compute the gradient and the Hessian of the modified objective function which have already been discussed in Eq. 5.12 and Eq. 5.13,

respectively. For a real reservoir problem, however, computation of the second derivative of $g(\mathbf{x})$ is not feasible. In the current study, the terms involving the second derivative of $g(\mathbf{x})$ and the second derivative of the barrier function from the Hessian (Eq. 5.13) are neglected. The resulting Gauss-Newton approximation of the Hessian matrix can be written as

$$\nabla^2_{\mathbf{x}} O(\mathbf{x}^l) \approx \mathbf{C}_X^{-1} + \mathbf{G}_l^T \mathbf{C}_D^{-1} \mathbf{G}_l. \tag{5.17}$$

Substituting Eq. 5.12 and Eq. 5.17 into Eq. 5.16, we obtain

$$(\mathbf{C}_X^{-1} + \mathbf{G}_l^T \mathbf{C}_D^{-1} \mathbf{G}_l)\delta\mathbf{x}^{l+1} = -\left[\mathbf{C}_X^{-1}(\mathbf{x}^l - \mathbf{x}_{pr}) + \mathbf{G}_l^T \mathbf{C}_D^{-1}(g(\mathbf{x}^l) - \mathbf{d}_{\text{obs}}) + t\nabla_{\mathbf{x}^l} f_b(\mathbf{x}^l)\right]. \tag{5.18}$$

In Eq. 5.18, $\mathbf{x}^l$ is the estimate of the model variables from the $l$th iteration and $\mathbf{G}_l$ is the sensitivity matrix from the $l$th iteration. Rearranging Eq. 5.18,

$$\begin{aligned}\delta\mathbf{x}^{l+1} = &-(\mathbf{C}_X^{-1} + \mathbf{G}_l^T \mathbf{C}_D^{-1} \mathbf{G}_l)^{-1} \\ &\times \left[\mathbf{C}_X^{-1}(\mathbf{x}^l - \mathbf{x}_{pr}) + \mathbf{G}_l^T \mathbf{C}_D^{-1}(g(\mathbf{x}^l) - \mathbf{d}_{\text{obs}}) + t\nabla_{\mathbf{x}^l} f_b(\mathbf{x}^l)\right],\end{aligned} \tag{5.19}$$

which can be written in the following form

$$\begin{aligned}\delta\mathbf{x}^{l+1} = \mathbf{x}_{pr} - \mathbf{x}^l - \mathbf{C}_X \mathbf{G}_l^T (\mathbf{C}_D + \mathbf{G}_l \mathbf{C}_X \mathbf{G}_l^T)^{-1}[g(\mathbf{x}^l) - \mathbf{d}_{\text{obs}} - \mathbf{G}_l(\mathbf{x}^l - \mathbf{x}_{pr})] \\ - t(\mathbf{C}_X^{-1} + \mathbf{G}_l^T \mathbf{C}_D^{-1} \mathbf{G}_l)^{-1} \nabla_{\mathbf{x}^l} f_b(\mathbf{x}^l).\end{aligned} \tag{5.20}$$

Substituting for the term $(\mathbf{C}_X^{-1} + \mathbf{G}_l^T \mathbf{C}_D^{-1} \mathbf{G}_l)^{-1}$ in Eq. 5.20,

$$\begin{aligned}\delta\mathbf{x}^{l+1} = \mathbf{x}_{pr} - \mathbf{x}^l - \mathbf{C}_X \mathbf{G}_l^T (\mathbf{C}_D + \mathbf{G}_l \mathbf{C}_X \mathbf{G}_l^T)^{-1} \left[g(\mathbf{x}^l) - \mathbf{d}_{\text{obs}} - \mathbf{G}_l(\mathbf{x}^l - \mathbf{x}_{pr})\right] \\ - t\left[\mathbf{C}_X - \mathbf{C}_X \mathbf{G}_l^T (\mathbf{C}_D + \mathbf{G}_l \mathbf{C}_X \mathbf{G}_l^T)^{-1} \mathbf{G}_l \mathbf{C}_X\right] \nabla_{\mathbf{x}^l} f_b(\mathbf{x}^l).\end{aligned} \tag{5.21}$$

In ensemble-based methods, the prior covariance matrix $\mathbf{C}_X$ is expressed as

$$\mathbf{C}_X = \frac{1}{(N_e - 1)} \Delta\mathbf{X}^f (\Delta\mathbf{X}^f)^T, \tag{5.22}$$

where, $N_e$ is the ensemble size and $\Delta\mathbf{X}^f$ is a matrix whose columns are the vectors of deviations of model variables from the ensemble mean. Substituting the expression for $\mathbf{C}_X$ from Eq. 5.22 into the last term on the right hand side of Eq. 5.21

$$\begin{aligned}\delta\mathbf{x}^{l+1} = \mathbf{x}_{pr} - \mathbf{x}^l - \mathbf{C}_X\mathbf{G}_l^T(\mathbf{C}_D + \mathbf{G}_l\mathbf{C}_X\mathbf{G}_l^T)^{-1}\left[g(\mathbf{x}^l) - \mathbf{d}_{\text{obs}} - \mathbf{G}_l(\mathbf{x}^l - \mathbf{x}_{pr})\right] \\ - t\left[\frac{1}{(N_e-1)}\Delta\mathbf{X}^f(\Delta\mathbf{X}^f)^T - \mathbf{C}_X\mathbf{G}_l^T(\mathbf{C}_D + \mathbf{G}_l\mathbf{C}_X\mathbf{G}_l^T)^{-1}\right. \\ \left.\times\mathbf{G}_l\frac{1}{(N_e-1)}\Delta\mathbf{X}^f(\Delta\mathbf{X}^f)^T\right]\nabla_{\mathbf{x}^l}f_b(\mathbf{x}^l). \quad (5.23)\end{aligned}$$

Eq. 5.23 can be simplified further as

$$\begin{aligned}\delta\mathbf{x}^{l+1} = \mathbf{x}_{pr} - \mathbf{x}^l - \mathbf{C}_X\mathbf{G}_l^T(\mathbf{C}_D + \mathbf{G}_l\mathbf{C}_X\mathbf{G}_l^T)^{-1}\left[g(\mathbf{x}^l) - \mathbf{d}_{\text{obs}} - \mathbf{G}_l(\mathbf{x}^l - \mathbf{x}_{pr})\right] \\ - \frac{t}{(N_e-1)}\left[\Delta\mathbf{X}^f - \mathbf{C}_X\mathbf{G}_l^T(\mathbf{C}_D + \mathbf{G}_l\mathbf{C}_X\mathbf{G}_l^T)^{-1}\mathbf{G}_l\Delta\mathbf{X}^f\right](\Delta\mathbf{X}^f)^T\nabla_{\mathbf{x}^l}f_b(\mathbf{x}^l). \quad (5.24)\end{aligned}$$

Eq. 5.24 shows the expression for computing the search direction $\delta\mathbf{x}^{l+1}$ for the $(l+1)$th iteration. At the $l$th iteration, let $\Delta\mathbf{D}^l$ represents the deviation of each vector of computed data from the mean vector of computed data and let $\Delta\mathbf{X}^l$ denotes the deviation of each vector of model variables from the current mean. As we know, the ensemble-average sensitivity matrix $\mathbf{G}_l$ is the coefficient matrix relating the changes in model variables to the changes in computed data,

$$\Delta\mathbf{D}^l = \mathbf{G}_l\Delta\mathbf{X}^l. \quad (5.25)$$

As in Gu and Oliver (2007), an approximation to the sensitivity matrix, $\mathbf{G}_l$, can be obtained from Eq. 5.25 using a pseudo-inverse based on the singular value decomposition of $\Delta\mathbf{X}^l$. Knowing the new search direction, $\delta\mathbf{x}^{l+1}$, the iterative estimate of the model variables $\mathbf{x}^{l+1}$ can be computed as

$$\mathbf{x}^{l+1} = \mathbf{x}^l + \beta_l\delta\mathbf{x}^{l+1}. \quad (5.26)$$

In Eq. 5.26, $\beta_l$ is an adjustment to the step length at the $l$th iterate, whose optimal value can be determined using the standard methods (Dennis and Schnabel, 1983).

For the current investigation, the optimal value of the step length is computed using a line search method. Substituting the expression for $\delta\mathbf{x}^{l+1}$ (Eq. 5.24) into Eq. 5.26 and rearranging, we obtain

$$\begin{aligned}\mathbf{x}^{l+1} &= \beta_l \mathbf{x}_{pr} + (1-\beta_l)\mathbf{x}^l \\ &\quad - \beta_l \mathbf{C}_X \mathbf{G}_l^T (\mathbf{C}_D + \mathbf{G}_l \mathbf{C}_X \mathbf{G}_l^T)^{-1} [g(\mathbf{x}^l) - \mathbf{d}_{\text{obs}} - \mathbf{G}_l(\mathbf{x}^l - \mathbf{x}_{pr})] \\ &\quad - \frac{t\beta_l}{(N_e-1)} [\Delta\mathbf{X}^f - \mathbf{C}_X \mathbf{G}_l^T (\mathbf{C}_D + \mathbf{G}_l \mathbf{C}_X \mathbf{G}_l^T)^{-1} \mathbf{G}_l \Delta\mathbf{X}^f] (\Delta\mathbf{X}^f)^T \nabla_{\mathbf{x}^l} f_b(\mathbf{x}^l). \quad (5.27)\end{aligned}$$

Eq. 5.27 shows the expression for the ($l$+1)th iterative estimate of the model variables using the interior-point formulation for the constrained EnKF. Gu and Oliver (2007) proposed an iterative form of the ensemble Kalman filter (EnRML) for data assimilation in nonlinear problems where the primary focus of the iterative approach was to enforce constraints and to ensure that the resulting ensemble is representative of the conditional probability density function. Gu and Oliver (2007) formulated the problem in terms of the static model variables, $\mathbf{m}$. During their study, the dynamic state variables were obtained by running the reservoir simulator. The final form of the iterative estimate of the static model variables obtained by Gu and Oliver (2007) is given in Eq. 5.28.

$$\begin{aligned}\mathbf{m}^{l+1} = \beta_l \mathbf{m}_{pr} \quad &+ \quad (1-\beta_l)\mathbf{m}^l - \beta_l \mathbf{C}_M \mathbf{G}_l^T (\mathbf{C}_D + \mathbf{G}_l \mathbf{C}_M \mathbf{G}_l^T)^{-1} \\ &\times \quad [g(\mathbf{m}^l) - \mathbf{d}_{\text{obs}} - \mathbf{G}_l(\mathbf{m}^l - \mathbf{m}_{pr})]. \quad (5.28)\end{aligned}$$

It can be noticed that Eq. 5.27 has a similar form to the expression for the EnRML method given in Eq. 5.28. Eq. 5.27, however, has an additional term which appears as the contribution from the barrier function that enforces the inequality constraints on the model variables.

As Eq. 5.24 indicates, finding the new search direction would require the computation of the approximation to the sensitivity matrix, $\mathbf{G}_l$, during each iteration. This would involve evaluation of the individual ensemble member with the updated model

variables from the initial time using the forward model. For a real reservoir problem, however, running the reservoir simulator from the initial time is computationally expensive. In order to reduce the computational cost and to make the method practically feasible, we propose to replace $\mathbf{G}_l$ in Eq. 5.24 with the forecast sensitivity. If $\Delta\mathbf{D}_{pr}$ represents the deviation of each vector of forecast data from the mean vector of forecast data and $\Delta\mathbf{X}^f$ is a matrix whose columns are the vectors of deviations of forecast model variables from the ensemble mean, then the forecast sensitivity $\mathbf{G}_{pr}$ relates $\Delta\mathbf{D}_{pr}$ and $\Delta\mathbf{X}^f$ in the following manner

$$\Delta\mathbf{D}_{pr} = \mathbf{G}_{pr}\Delta\mathbf{X}^f. \tag{5.29}$$

As discussed earlier, an approximation to the forecast sensitivity matrix, $\mathbf{G}_{pr}$, can be obtained from Eq. 5.29 using a pseudo-inverse based on the singular value decomposition of $\Delta\mathbf{X}^f$. Under this assumption, Eq. 5.27 becomes

$$\begin{aligned}\mathbf{x}^{l+1} = \beta_l\mathbf{x}_{pr} + (1-\beta_l)\mathbf{x}^l \\ - \beta_l\mathbf{C}_X\mathbf{G}_{pr}^T(\mathbf{C}_D + \mathbf{G}_{pr}\mathbf{C}_X\mathbf{G}_{pr}^T)^{-1}\left[g(\mathbf{x}^l) - \mathbf{d}_{\text{obs}} - \mathbf{G}_{pr}(\mathbf{x}^l - \mathbf{x}_{pr})\right] \\ - \frac{t\beta_l}{(N_e-1)}\left[\Delta\mathbf{X}^f - \mathbf{C}_X\mathbf{G}_{pr}^T(\mathbf{C}_D + \mathbf{G}_{pr}\mathbf{C}_X\mathbf{G}_{pr}^T)^{-1}\mathbf{G}_{pr}\Delta\mathbf{X}^f\right](\Delta\mathbf{X}^f)^T\nabla_{\mathbf{x}^l}f_b(\mathbf{x}^l).\end{aligned} \tag{5.30}$$

The simulated data during each iteration can be approximated using the forecast sensitivity, $\mathbf{G}_{pr}$, in the following manner

$$g(\mathbf{x}^l) \approx g(\mathbf{x}_{pr}) + \mathbf{G}_{pr}(\mathbf{x}^l - \mathbf{x}_{pr}). \tag{5.31}$$

Using the forecast sensitivity, $\mathbf{G}_{pr}$, and the approximation stated in Eq. 5.31, the expression for the search direction for the $(l+1)$th iteration can be updated as

$$\begin{aligned}\delta\mathbf{x}^{l+1} = \mathbf{x}_{pr} - \mathbf{x}^l + \mathbf{C}_X\mathbf{G}_{pr}^T(\mathbf{C}_D + \mathbf{G}_{pr}\mathbf{C}_X\mathbf{G}_{pr}^T)^{-1}[\mathbf{d}_{\text{obs}} - g(\mathbf{x}_{pr})] \\ - \frac{t}{(N_e-1)}\left[\Delta\mathbf{X}^f - \mathbf{C}_X\mathbf{G}_{pr}^T(\mathbf{C}_D + \mathbf{G}_{pr}\mathbf{C}_X\mathbf{G}_{pr}^T)^{-1}\mathbf{G}_{pr}\Delta\mathbf{X}^f\right] \\ \times(\Delta\mathbf{X}^f)^T\nabla_{\mathbf{x}^l}f_b(\mathbf{x}^l).\end{aligned} \tag{5.32}$$

Substituting Eq. 5.32 into Eq. 5.26, we obtain

$$\mathbf{x}^{l+1} = \beta_l \mathbf{x}_{pr} + (1-\beta_l)\mathbf{x}^l + \beta_l \mathbf{C}_X \mathbf{G}_{pr}^T (\mathbf{C}_D + \mathbf{G}_{pr}\mathbf{C}_X\mathbf{G}_{pr}^T)^{-1}[\mathbf{d}_{\text{obs}} - g(\mathbf{x}_{pr})] - \frac{t\beta_l}{(N_e-1)}\left[\Delta\mathbf{X}^f - \mathbf{C}_X\mathbf{G}_{pr}^T(\mathbf{C}_D + \mathbf{G}_{pr}\mathbf{C}_X\mathbf{G}_{pr}^T)^{-1}\mathbf{G}_{pr}\Delta\mathbf{X}^f\right] \times (\Delta\mathbf{X}^f)^T \nabla_{\mathbf{x}^l} f_b(\mathbf{x}^l). \quad (5.33)$$

Eq. 5.33 shows the final form of the expression for computing the iterative estimate of the model variables used in this study. As Eq. 5.33 indicates, the iterative estimate of the model variables involves a term comprising of the Kalman gain that appears during the standard implementation of the EnKF. The expression for the Kalman gain is given in Eq. 5.34.

$$\mathbf{K} = \mathbf{C}_X\mathbf{G}_{pr}^T(\mathbf{C}_D + \mathbf{G}_{pr}\mathbf{C}_X\mathbf{G}_{pr}^T)^{-1}. \quad (5.34)$$

As Eq. 5.33 indicates, one does not have to handle large matrices in order to compute the iterative estimate of the model variables. The only significant computations involved in solving Eq. 5.33 are the matrix-vector product type of computations, such as $\mathbf{G}_{pr}\Delta\mathbf{X}^f$ and $(\Delta\mathbf{X}^f)^T\nabla_{\mathbf{x}^l} f_b(\mathbf{x}^l)$, which are computationally inexpensive.

### 5.3.1 Reduced form of objective function

In an ideal situation, we would like to minimize the objective function given in Eq. 5.11. For most of the reservoir engineering related applications, however, the state vector often contains a large number of model parameters and state variables. Under such condition, computations of the covariance matrix, $\mathbf{C}_X$, and its inverse, $\mathbf{C}_X^{-1}$, are not feasible. Hence, for practical purposes, a reduced form of the objective function is evaluated for each ensemble member during each iteration of the IPCEnKF method to check for convergence. The reduced form of the objective function for the $i$th ensemble member involves the data mismatch term and the barrier function as shown below

$$\overline{O}(\mathbf{x}_i^{l+1}) = \frac{1}{2}(g(\mathbf{x}_i^{l+1}) - \mathbf{d}_{\text{obs},i})^T\mathbf{C}_D^{-1}(g(\mathbf{x}_i^{l+1}) - \mathbf{d}_{\text{obs},i}) + tf_b(\mathbf{x}_i^{l+1}). \quad (5.35)$$

In Eq. 5.35, $g(\mathbf{x}_i^{l+1})$ denotes the simulated data at the $(l+1)$th iteration for the $i$th ensemble member which can be approximated as

$$g(\mathbf{x}_i^{l+1}) \approx \mathbf{d}_{pr,i} + \mathbf{G}_{pr}(\mathbf{x}_i^{l+1} - \mathbf{x}_{pr,i}), \tag{5.36}$$

where, $\mathbf{d}_{pr,i}$ represents the simulated data obtained from the prior state vector for the $i$th ensemble member. For each value of the barrier parameter $t$, it is important to attain the convergence for the entire ensemble before updating the barrier parameter. A reduced form of the objective function, similar to Eq. 5.35, is evaluated for the entire ensemble during each iteration in order to determine if the solution has converged for the given value of the barrier parameter. The reduced form of the objective function for the entire ensemble can be represented as

$$\overline{O}_{ens}(\mathbf{X}^{l+1}) = \frac{1}{N_e}\sum_{i=1}^{N_e}\left[\frac{1}{2}(g(\mathbf{x}_i^{l+1}) - \mathbf{d}_{\text{obs},i})^T\mathbf{C}_D^{-1}(g(\mathbf{x}_i^{l+1}) - \mathbf{d}_{\text{obs},i}) + tf_b(\mathbf{x}_i^{l+1})\right]. \tag{5.37}$$

### 5.3.2 Updating barrier parameter

In the current study, the algorithm for the IPCEnKF method is initialized with a positive value of the barrier parameter, $t$. It is, however, necessary to decrease the barrier parameter and to seek the minimizer at the updated value of the barrier parameter. During each iteration of the IPCEnKF method, Eq. 5.37 is used to evaluate the objective function for the entire ensemble. Following criterion is used to determine if a minimizer is attained for the given value of the barrier parameter,

$$\left|\frac{\overline{O}_{ens}(\mathbf{X}^l) - \overline{O}_{ens}(\mathbf{X}^{l+1})}{\overline{O}_{ens}(\mathbf{X}^l)}\right| < \epsilon_1 \quad \text{with} \quad \epsilon_1 = 0.05. \tag{5.38}$$

When the objective function evaluated at the current iteration satisfies the condition given in Eq. 5.38, the barrier parameter for the next iteration is updated as

$$t^{l+1} = \frac{t^l}{\alpha} \quad \text{with} \quad \alpha = 1.25. \tag{5.39}$$

### 5.3.3 Algorithm description for the IPCEnKF

Following is a brief generalized algorithm for implementing the IPCEnKF method. This algorithm describes various steps which need to be carried out in order to obtain a solution for each ensemble member that satisfies the inequality constraints on any type of state variable.

- At each assimilation time, initialize the barrier parameter, $t$, with a positive value. For each ensemble member $i$, $(i = 1, \ldots, N_e)$, use the reservoir simulator to compute the forecast values of state variables at assimilation time $t_k$, $(k = 1, \ldots, k_{\max})$, based on the updated values of model parameters and state variables at time $t_{k-1}$.

    1. Obtain the vectors of deviations of simulated data, $\Delta\mathbf{D}_{pr}$, and the vectors of deviations of the model and state variables, $\Delta\mathbf{X}^f$, from their respective ensemble means.
    2. Compute the ensemble approximation to the forecast sensitivity, $\mathbf{G}_{pr}$, using Eq. 5.29.
    3. For ensemble member $i$, compute the search direction $\delta\mathbf{x}^{l+1}$ using Eq. 5.32.
    4. For ensemble member $i$, compute the optimal value of the step size $\beta_l$ using the search direction computed in step 3 and a line search method. The objective function for ensemble member $i$ is evaluated using Eq. 5.35 during the line search approach.
    5. Compute the ensemble-based objective function using Eq. 5.37.
    6. Evaluate the criterion (Eq. 5.38) for updating the barrier parameter. If the ensemble-based objective function at the current iteration satisfies Eq. 5.38, update the barrier parameter using Eq. 5.39.

7. If the convergence criteria are satisfied, exit and go to the next assimilation time; otherwise go to step 3 and iterate.

The following criteria are used to determine if the algorithm should terminate:

- $|\overline{O}_{ens}(\mathbf{X}^{l+1}) - \overline{O}_{ens}(\mathbf{X}^{l})| < \epsilon_2 \, \overline{O}_{ens}(\mathbf{X}^{l})$ with $\epsilon_2 = 0.0001$
- $\overline{O}_{ens}(\mathbf{X}^{l+1}) < N_D$ where $N_D$ is the number of observations at the current data assimilation time step.
- Iteration exceeds the pre-set maximum number of iterations.

The computational cost during each iteration of the IPCEnKF method is significantly less than the cost of running the reservoir simulator. Therefore, we set a fairly large limit (30) on the maximum number of iterations during the implementation of the IPCEnKF method.

We consider two synthetic examples to demonstrate the performance of the IPCEnKF. The results obtained from the IPCEnKF are compared with the results obtained from the standard EnKF. The newly proposed IPCEnKF method is tested first on a synthetic 1D linear example to demonstrate its effectiveness when the nonlinearity is not an issue. The synthetic 1D linear example has been discussed earlier and the model description has been given in Section 4.4.1. The second test problem is a more realistic 3D, three-phase reservoir flow assimilation problem based on the modified SPE9 model.

## 5.4 1D linear example

In this section, we describe in detail the application of the IPCEnKF method to a synthetic 1D linear example studied in Chapter 4. The reference model description and the initial ensemble of model variables have been discussed in Section 4.4.2. The parameters used for data assimilation with the standard EnKF are mentioned in Section 4.4.3.

### 5.4.1 Results and discussion

As discussed in Section 4.4.4, the data mismatch objective function (Eq. 4.7) evaluated for the updated ensemble obtained from the standard EnKF has a value of 0.053. The updated model variables for some ensemble members, however, show constraint violations at a number of locations (Fig. 4.4(a)). The model variable $\mathbf{x}_j$ is subject to the inequality constraints defined by the boundaries of the feasible region which can be written as

$$1 \ \geq \ \mathbf{x}_j \ \geq \ 0.$$

Since the model variable is subject to two inequality constraints, following form of the barrier function is chosen for the implementation of the IPCEnKF method. The barrier function can be expressed as

$$f_b(\mathbf{x}) = -\sum_{j=1}^{N_x} \log(\mathbf{x}_j(1-\mathbf{x}_j)). \tag{5.40}$$

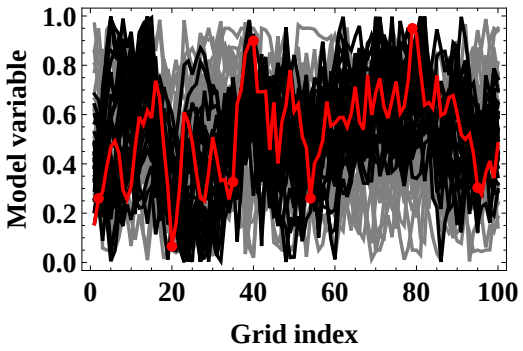
The barrier parameter, $t$, associated with the barrier function is initialized with a value of 1 at the beginning of the IPCEnKF algorithm. The adjustment to the step length, $\beta_l$, for different ensemble members is obtained using a line search method. The values of $\beta_l$ for realizations 1 and 15 obtained during different iterations of the IPCEnKF method are given in Table 5.1. The results for the updated ensemble of model variables at different iterations of the IPCEnKF method are shown in Fig. 5.1. The initial ensemble of model variables is shown by the gray curves, the updated realizations obtained during each iteration, where the barrier parameter ($t$) is updated, are shown by the black curves, the red curve is the reference model, and the true observations from the reference model are denoted by red dots. As Fig. 5.1(a) shows, the model variables for some of the updated realizations have values close to the boundaries of the feasible region. None of the updated realizations, however, contains model variables which violate the inequality constraints.

**Table 5.1:** The values of $\beta_l$ for realizations 1 and 15 obtained during different iterations of the IPCEnKF method

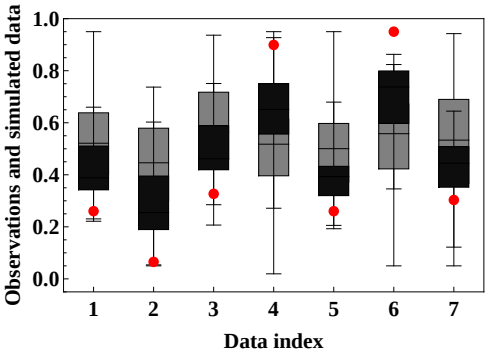
| Iteration index ($l$) | Realization 1 | Realization 15 |
|---|---|---|
| 1 | 0.2326 | 0.1567 |
| 10 | 0.1176 | 0.3287 |
| 20 | 0.0265 | 0.0871 |
| 30 | 0.0000 | 0.00005 |

Fig. 5.1 also shows the results for the data-match obtained from the updated ensemble of realizations at different iterations of the IPCEnKF method. The gray and the black boxes denote the statistical distributions of the simulated data from the initial ensemble and the simulated data from the updated ensemble obtained from the IPCEnKF method, respectively, and the red dots represent the true observations (without noise) from the reference model. As Fig. 5.1(b) shows, the data-match between the simulated data and the true observations during the first iteration of the IPCEnKF method is poor. The data-match, however, improves with each iteration. The simulated data obtained from the final updated ensemble of realizations at iteration 30 show a good match to the true observations as shown in Fig. 5.1(f).

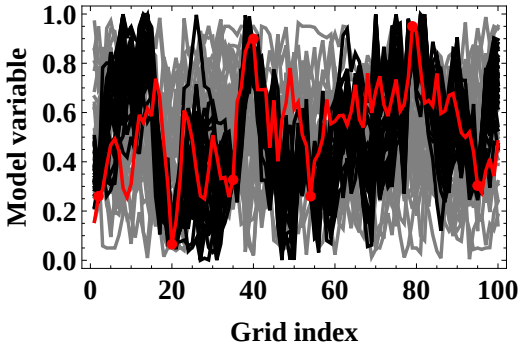
As discussed in Chapter 4, the CEnKF method is implemented for enforcing the inequality constraints (in the form of the equality constraints) on model variable estimates. The results obtained from the application of the CEnKF to the synthetic 1D linear example are discussed in Section 4.4.5. The unconstrained solutions (standard EnKF) and the constrained solutions (IPCEnKF and CEnKF) obtained for realizations 1 and 15 are compared in Fig. 5.2. From the results, it is clear that, the proposed IPCEnKF method has been successful in enforcing the inequality constraints on the model variables and thereby, has improved the estimates compared to the standard EnKF which suffers from the issue of constraint violations. As Fig. 5.2 shows, the final updated estimates of model variables obtained from the IPCEnKF and the CEnKF methods are nearly indistinguishable for realizations 1 and 15. The



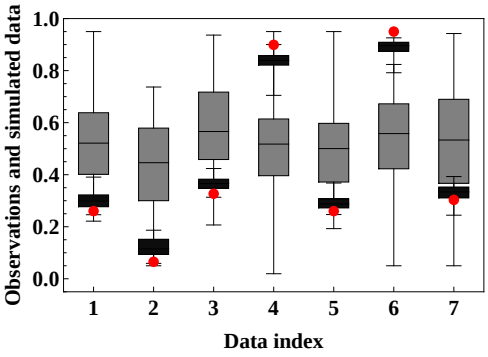

(a) Iteration 1: estimate of model variables



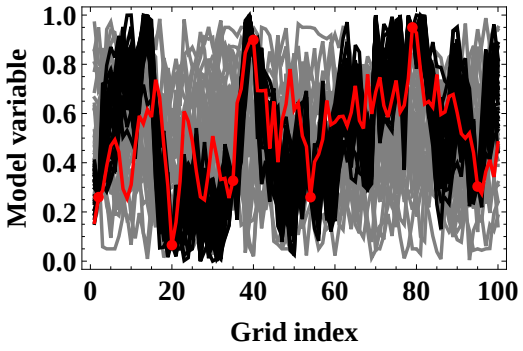

(b) Iteration 1: data-match



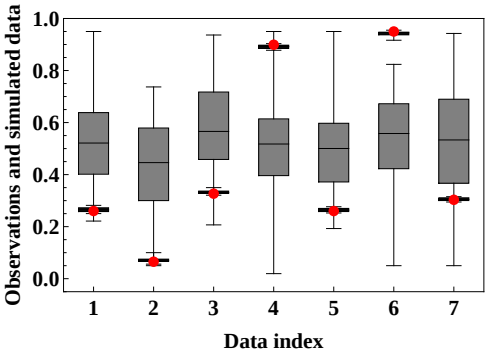

(c) Iteration 10: estimate of model variables




(d) Iteration 10: data-match



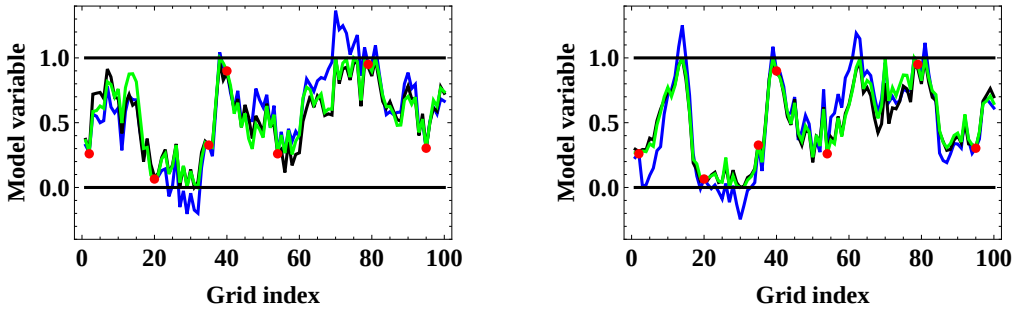

(e) Iteration 30: estimate of model variables




(f) Iteration 30: data-match

**Figure 5.1:** The updated ensemble of model variables (black curves) and the data-match (black boxes) obtained at different iterations of the IPCEnKF method. The red curve is the reference model and the red dots are the true observations.

CEnKF method, however, requires identification of the active constraints and computation of a separate Kalman gain for each ensemble member during each iteration of the algorithm which makes the method computationally expensive compared to the IPCEnKF method.




(a) Realization 1 (b) Realization 15

**Figure 5.2:** Comparison of the unconstrained solutions (standard EnKF) and the constrained solutions (IPCEnKF, iteration 30 and CEnKF, iteration 10) for realizations 1 and 15. The blue curves denote the unconstrained solutions, the black curves denote the constrained solutions from the IPCEnKF, and the green curves denote the constrained solutions from the CEnKF. The red dots are the true observations from the reference model.

The IPCEnKF method can be understood by analyzing the effect of incorporating the inequality constraints during the data assimilation process. As shown in Eq. 5.33, the iterative estimate of model variables, $\mathbf{x}^{l+1}$, is made up of a number of terms. The expression for calculating the iterative estimate of model variables is given as

$$\begin{aligned}\mathbf{x}^{l+1} = \beta_l \mathbf{x}_{pr} + (1-\beta_l)\mathbf{x}^l + \beta_l \mathbf{C}_X \mathbf{G}_{pr}^T (\mathbf{C}_D + \mathbf{G}_{pr}\mathbf{C}_X\mathbf{G}_{pr}^T)^{-1}[\mathbf{d}_{\text{obs}} - g(\mathbf{x}_{pr})] \\ - \frac{t\beta_l}{(N_e-1)}\left[\Delta\mathbf{X}^f - \mathbf{C}_X\mathbf{G}_{pr}^T(\mathbf{C}_D + \mathbf{G}_{pr}\mathbf{C}_X\mathbf{G}_{pr}^T)^{-1}\mathbf{G}_{pr}\Delta\mathbf{X}^f\right] \\ \times (\Delta\mathbf{X}^f)^T \nabla_{\mathbf{x}^l} f_b(\mathbf{x}^l).\end{aligned}$$

As the above expression shows, the iterative estimate of model variables consists of four elements: (1) prior knowledge of model variables, $\mathbf{x}_{pr}$, (2) the iterative estimate obtained at the previous iteration, $\mathbf{x}^l$, (3) contribution from the standard update (involves data mismatch), and (4) contribution from the barrier function. The terms

representing the contributions from the standard update and the barrier function can be expressed as

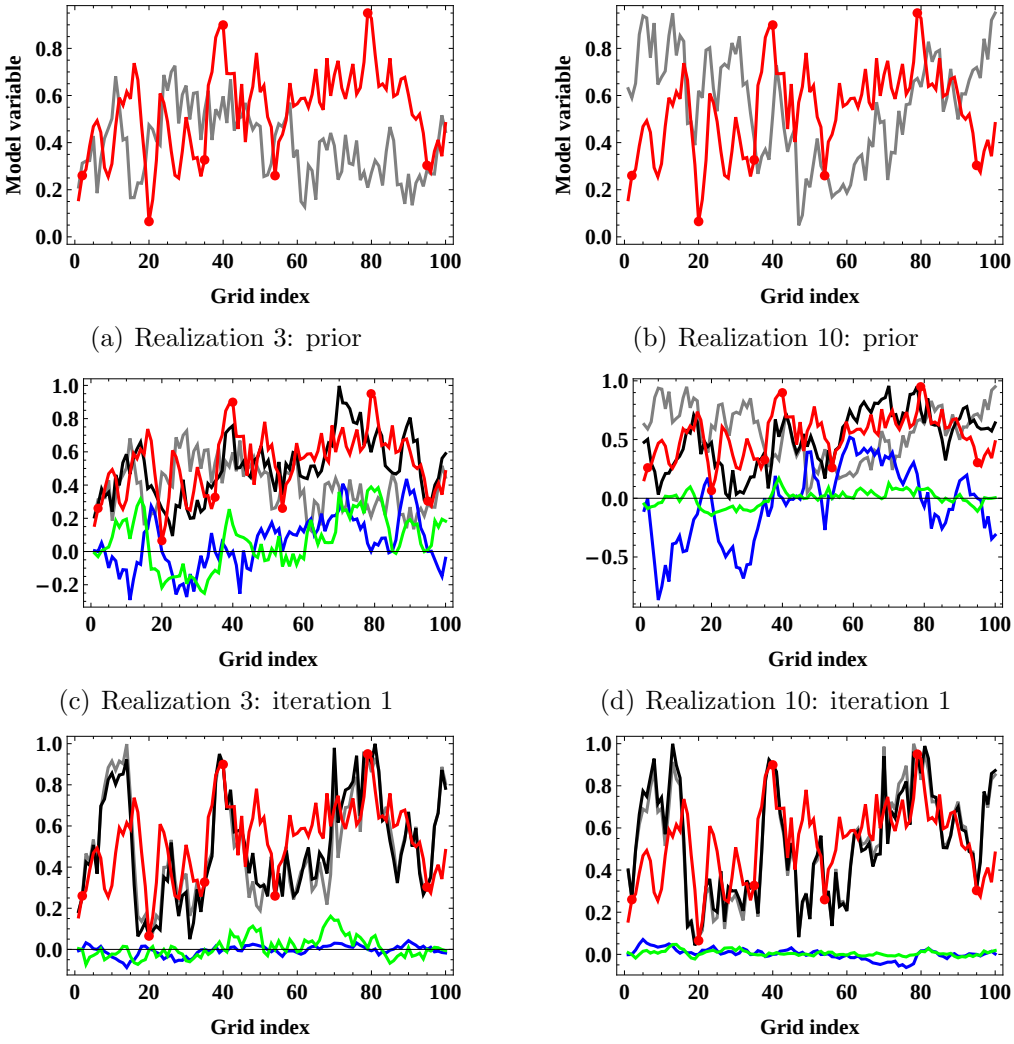
$$\text{standard contribution} = \beta_l \mathbf{C}_X \mathbf{G}_{pr}^T (\mathbf{C}_D + \mathbf{G}_{pr} \mathbf{C}_X \mathbf{G}_{pr}^T)^{-1} [\mathbf{d}_{\text{obs}} - g(\mathbf{x}_{pr})], \quad (5.41)$$

and

$$\text{barrier contribution} = -\frac{t\beta_l}{(N_e - 1)} \left[ \Delta \mathbf{X}^f - \mathbf{C}_X \mathbf{G}_{pr}^T (\mathbf{C}_D + \mathbf{G}_{pr} \mathbf{C}_X \mathbf{G}_{pr}^T)^{-1} \mathbf{G}_{pr} \Delta \mathbf{X}^f \right] \times (\Delta \mathbf{X}^f)^T \nabla_{\mathbf{x}^l} f_b(\mathbf{x}^l). \quad (5.42)$$

The term representing the barrier function contribution (Eq. 5.42) involves the gradient of the barrier function, $\nabla_{\mathbf{x}^l} f_b(\mathbf{x}^l)$. Recall that the logarithmic barrier function (Eq. 5.40) is used in the current study. As a result of this choice of the barrier function, the term representing the contribution from the barrier function shows either a high positive value or a high negative value at a location where the model variable approaches the boundary of the feasible region. The results from this analysis are given in Fig. 5.3 which shows the updated solutions for realizations 3 and 10 at different iterations of the IPCEnKF method. The terms representing the contributions from the standard update (Eq. 5.41) and the barrier function (Eq. 5.42) at different iterations are shown by the green and the blue curves, respectively. The red curve represents the reference model, the red dots denote the true observations from the reference model, and the black curve shows the estimate of model variables at the current iteration. In Figs. 5.3(a) through 5.3(d), the gray curves denote the prior estimates of realizations 3 and 10. The updated estimates of model variables from the previous iteration for realizations 3 and 10 are also shown by gray curves in Figs. 5.3(e) and 5.3(f).

As Fig. 5.3(b) shows, realization 10 contains some model variables which have values close to the boundaries of the feasible region. The model variable at gridblock 47 has the lowest value of 0.0487 which is close to the boundary of the feasible region




(a) Realization 3: prior

(b) Realization 10: prior

(c) Realization 3: iteration 1

(d) Realization 10: iteration 1

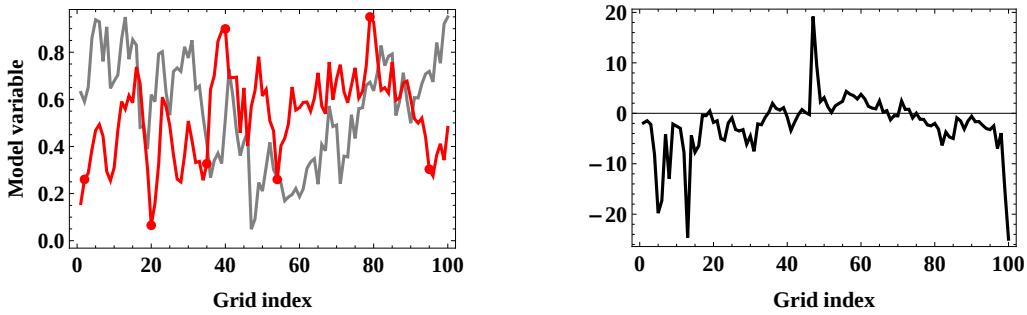
(e) Realization 3: iteration 20

(f) Realization 10: iteration 20

**Figure 5.3:** The updated solutions (black curve) for realizations 3 and 10 at different iterations of the IPCEnKF method. The green curves show the contribution from the standard update and the blue curves denote the contribution from the barrier function at different iterations. The prior realizations at different iterations are shown by the gray curves, the reference model is represented by the red curve, and the red dots denote the true observations from the reference model.

on the lower side (0). Realization 10 contains a number of model variables which have values close to the boundary of the feasible region on the higher side (1). The model variables at grid locations 5, 13, and 100 have values of 0.9377, 0.9494, and 0.9500, respectively. The negative of the gradient of the barrier function, $-\nabla_{\mathbf{x}^l} f_b(\mathbf{x}^l)$, for realization 10 at iteration 1 is plotted in Fig. 5.4(b). Due to the presence of model variables with values close to the boundaries of the feasible region, the negative of the gradient of the barrier function at iteration 1 shows peaks (high positive and negative values) at a number of locations. The negative of the gradient has the largest positive value for the model variable at gridblock 47 and has the largest negative value for the model variable at the gridblock 100. The term representing the contribution from the barrier function (Eq. 5.42) involves the gradient of the barrier function. As a result, the blue curve in Fig. 5.3(d) has similar features. As Eq. 5.42 shows, however, the gradient of the barrier function is pre-multiplied by a number of terms. As a result, the behavior of the term representing the contribution from the barrier function slightly differs from the behavior of the gradient of the barrier function. During iteration 1, the contribution from the barrier function appears to dominate the overall contribution to the solution and the contribution from the standard update (green curve in Fig. 5.3(d)) appears to be negligible for realization 10. The large contribution from the barrier function ensures that the resulting updated solution stays within the interior of the feasible region.

Fig. 5.3(a) shows the prior for realization 3. Note that none of the model variables for realization 3 has extreme values that are close to the boundaries of the feasible region. As a result, the contributions from the standard update (green curve) and the barrier function (blue curve), shown in Fig. 5.3(c), have comparable magnitudes during iteration 1. Once the solution lies in the feasible region, the contributions from the standard update and the barrier function show comparable magnitudes during all subsequent iterations. Such type of behavior can be seen from the results presented

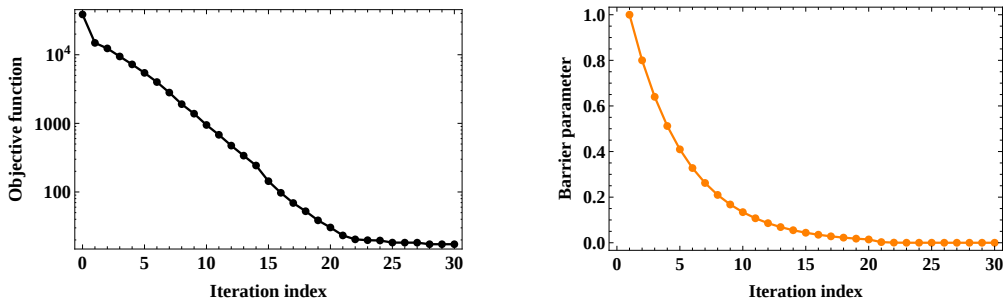



(a) Prior (b) $-\nabla_{\mathbf{x}^l} f_b(\mathbf{x}^l)$

**Figure 5.4:** The prior (gray curve) and the negative of the gradient of the barrier function at iteration 1 ($l$=1) of the IPCEnKF for realization 10. The red curve represents the reference model.

in Figs. 5.3(e) and 5.3(f).

The data mismatch objective function (Eq. 4.7) is evaluated for the simulated data obtained from the updated ensemble of realizations at the end of each iteration of the IPCEnKF method. The results of the data mismatch objective function are plotted on a semi-log scale and are shown in Fig. 5.5(a). The results show a continuous decrease in the objective function with increase in the number of iterations. The data mismatch objective function evaluated for the initial ensemble (iteration 0) results in a value of 38675. The continuous decrease in the data mismatch objective function indicates that the match between the simulated data and the true observations improves during each iteration. The objective function for the updated ensemble of realizations at iteration 30 has a value of 17.31.

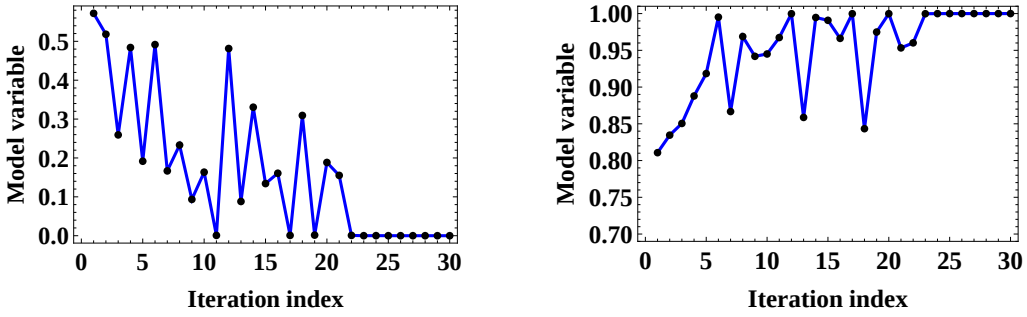



(a) Data mismatch objective function (b) Barrier parameter

**Figure 5.5:** The data mismatch objective function and the barrier parameter at different iterations of the IPCEnKF method.

One of the critical elements for the success of the interior-point formulation is updating the barrier parameter during the iterative process and seeking the minimizer of the objective function at every value of the barrier parameter. In an ideal situation, we would like to have the value of the barrier parameter as close as possible to 0. In the current study, the barrier parameter is updated based upon the decrease in the ensemble-based objective function. The strategy for updating the barrier parameter has been discussed in the Section 5.3.2. The evolution of the barrier parameter during different iterations of the IPCEnKF method is shown in Fig. 5.5(b). The value of the barrier parameter at the end of the iterative process (iteration 30) is $4.61 \times 10^{-6}$.

At the end of the data assimilation process using the IPCEnKF method (iteration 30), the model variable 27 of realization 8 has the minimum value of $1.89 \times 10^{-6}$ among all the model variables. Similarly, model variable 39 of realization 30 shows the maximum value of 0.99999 among all the model variables. Fig. 5.6 shows the iterative decrease and increase in the magnitudes of the model variables 27 (realization 8) and 39 (realization 30), respectively. In this particular test problem, we observed fairly rapid decrease or increase in the magnitudes of the model variable estimates during the first 5 to 10 iterations. This rapid decrease or increase, however, is followed by an intermittent oscillatory behavior for the next few iterations before the estimate is converged to its final solution.




(a) Realization 8: model variable 27 (b) Realization 30: model variable 39

**Figure 5.6:** The minimum and the maximum value of model variables at different iterations of the IPCEnKF method.

## 5.5 Modified SPE9 model

In the previous section, the application of the IPCEnKF to a synthetic 1D linear exmple was illustrated. For reservoir fluid flow problems, the implementation of the IPCEnKF becomes somewhat more challenging. In this section, we describe the implementation of the standard EnKF and the IPCEnKF methods on a synthetic 3D, three-phase reservoir model and compare the results obtained. A detailed discussion of the reservoir model description is given below.

### 5.5.1 Reservoir model description

We consider a 3D, three-phase synthetic reservoir model in this section. There is no initial gas cap and the connate water saturation is 0.15. The reservoir model is divided into a non-uniform grid-system of dimensions $24 \times 25 \times 15$. The areal dimensions of each gridblock are 300 ft $\times$ 300 ft and the layer thickness varies in different layers between the range of 8 ft to 100 ft. The size of the reservoir model is 7200 ft $\times$ 7500 ft $\times$ 359 ft. We assume that all sides of the reservoir model are no-flow boundaries, and the phases present in the reservoir are oil, water, and gas. The capillary effects are neglected during the fluid flow and transport process. Fig. 5.7 shows the structure of the SPE9 model which has a dip in the $x$-direction. The depths to the top of the first layer on the west side and the east side are defined to be 9000 ft and 10216 ft, respectively. There is an aquifer on the east side of the model and the water-oil contact (WOC) is defined at a depth of 9950 ft. The initial reservoir pressure at a reference depth of 9035 ft is assigned a value of 3600 psia, which is also the bubble point pressure.

There are six producers and three injectors in the reservoir simulation model. All six producers are completed between the second and the fourth layers. The injectors are completed in different layers; injector INJ1 is completed between layers twelve and fifteen, injector INJ2 is completed between layers eight and ten, and injector INJ3

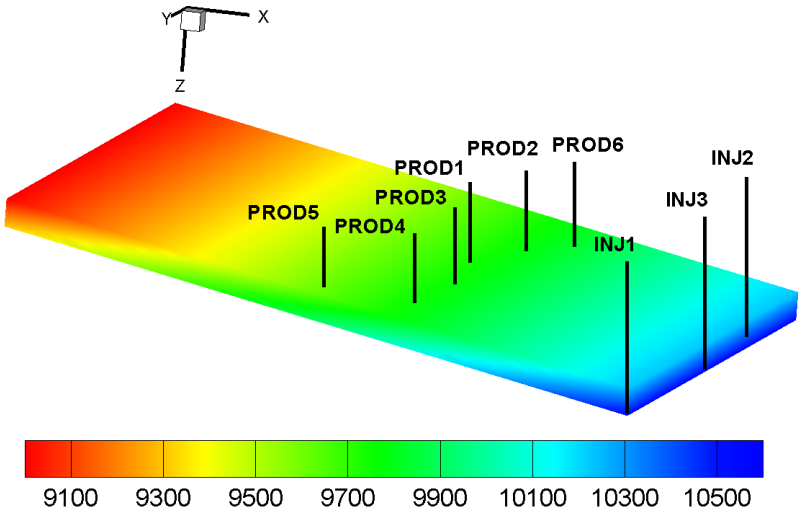



**Figure 5.7:** Structure map of the SPE9 reservoir model.

is completed between layers ten and twelve. The production from all six producers starts from day 0. We also assume that the water injection is started from day 0 in all three injectors and is continued throughout the life of the water-flooding project (day 2000). All six producers are operated with a target oil production rate of 15000 STB/day and a minimum BHP of 800 psia as a secondary constraint. All three injectors are operated with a target water injection rate of 10000 STB/day and a maximum BHP of 6000 psia as a secondary constraint.

The porosity and permeability in each gridblock are assumed to be the uncertain model parameters. The log-permeability, instead of permeability, is used in the state vector during data assimilation. The porosity field has a mean of 0.18 and a standard deviation of 0.03. The log-permeability field is generated with a mean and a standard deviation of 4.5 and 1.2, respectively. The log-permeability is correlated to porosity with a correlation coefficient of 0.60. An isotropic Gaussian variogram model with a practical range of 10 is used to generate the static model parameter fields. Fig. 5.8 shows the porosity and log-permeability for different layers together with well-locations for the reference case. The reference reservoir model is produced for a total of 2000 days. Figs. 5.9 and 5.10 show the profiles of the production data

from the reference model which include the oil production rate, water cut, producing gas-oil ratio, and bottom hole pressure data for the six producers, and the water injection rate and bottom hole pressure data for the three injectors.

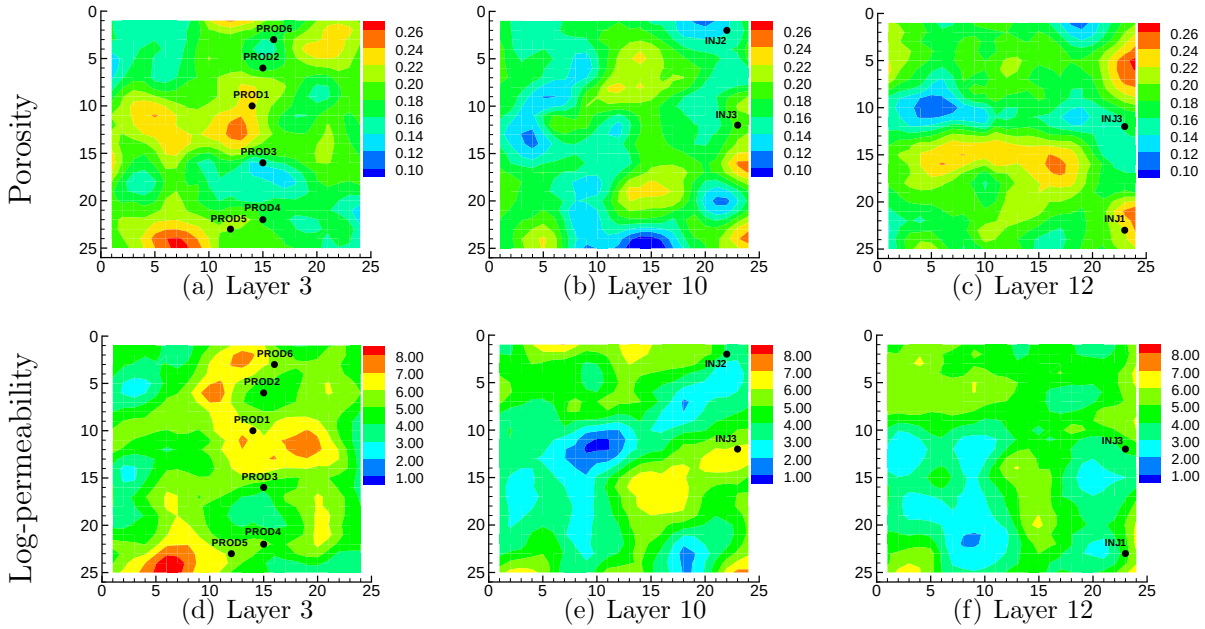


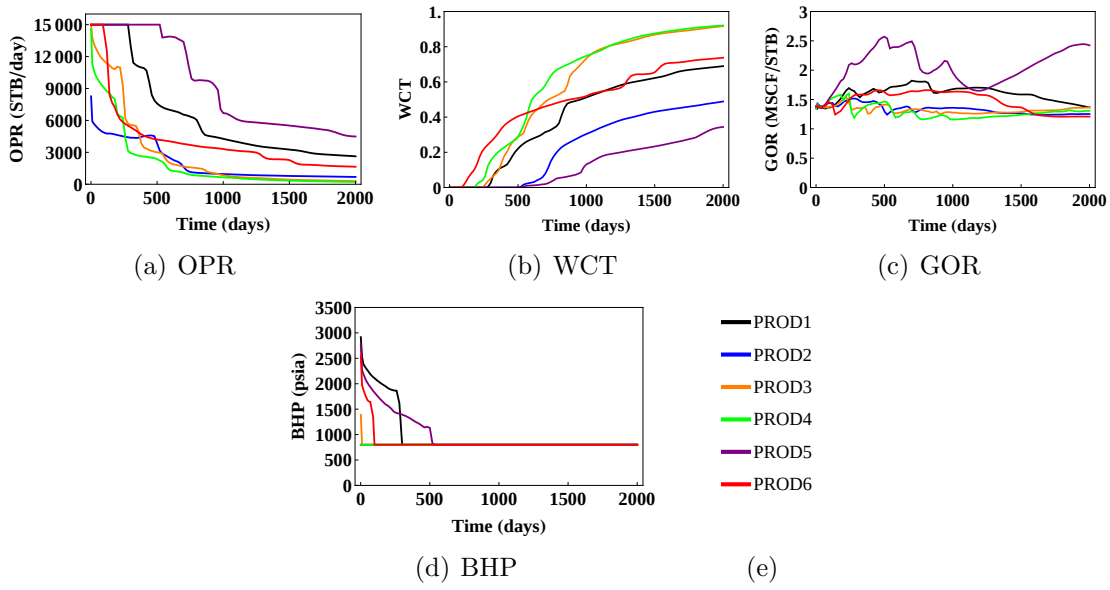

**Figure 5.8:** The porosity and log-permeability for different layers of the SPE9 model (reference).

### 5.5.2 Data assimilation setup

For this history matching problem, we consider two cases for evaluating the performance of the standard EnKF and the IPCEnKF methods. The porosity and log-permeability fields for the initial ensemble are generated using the same geostatistical model parameters as the reference model. The initial ensemble is not conditioned to values of porosity and log-permeability at well locations from the reference model and the ensemble size is 90 for both cases.

#### *5.5.2.1 Case 1: the initial ensemble is unbiased*

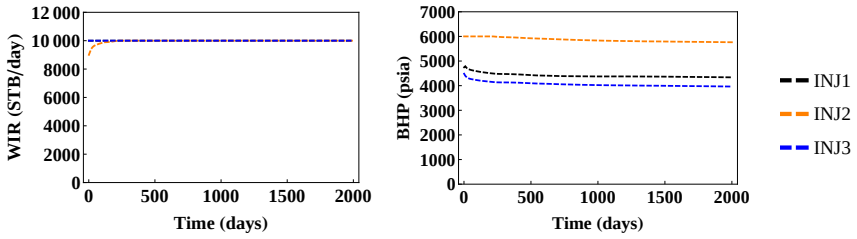
The initial ensemble of porosity and log-permeability fields is generated using sequential Gaussian simulation. For Case 1, we assume that the mean values of the porosity




(a) OPR (b) WCT (c) GOR

(d) BHP (e)

**Figure 5.9:** The profiles of the oil production rate (OPR), water cut (WCT), producing gas-oil ratio (GOR), and bottom hole pressure (BHP) for the six producers from the SPE9 model (reference).




(a) WIR (b) BHP (c)

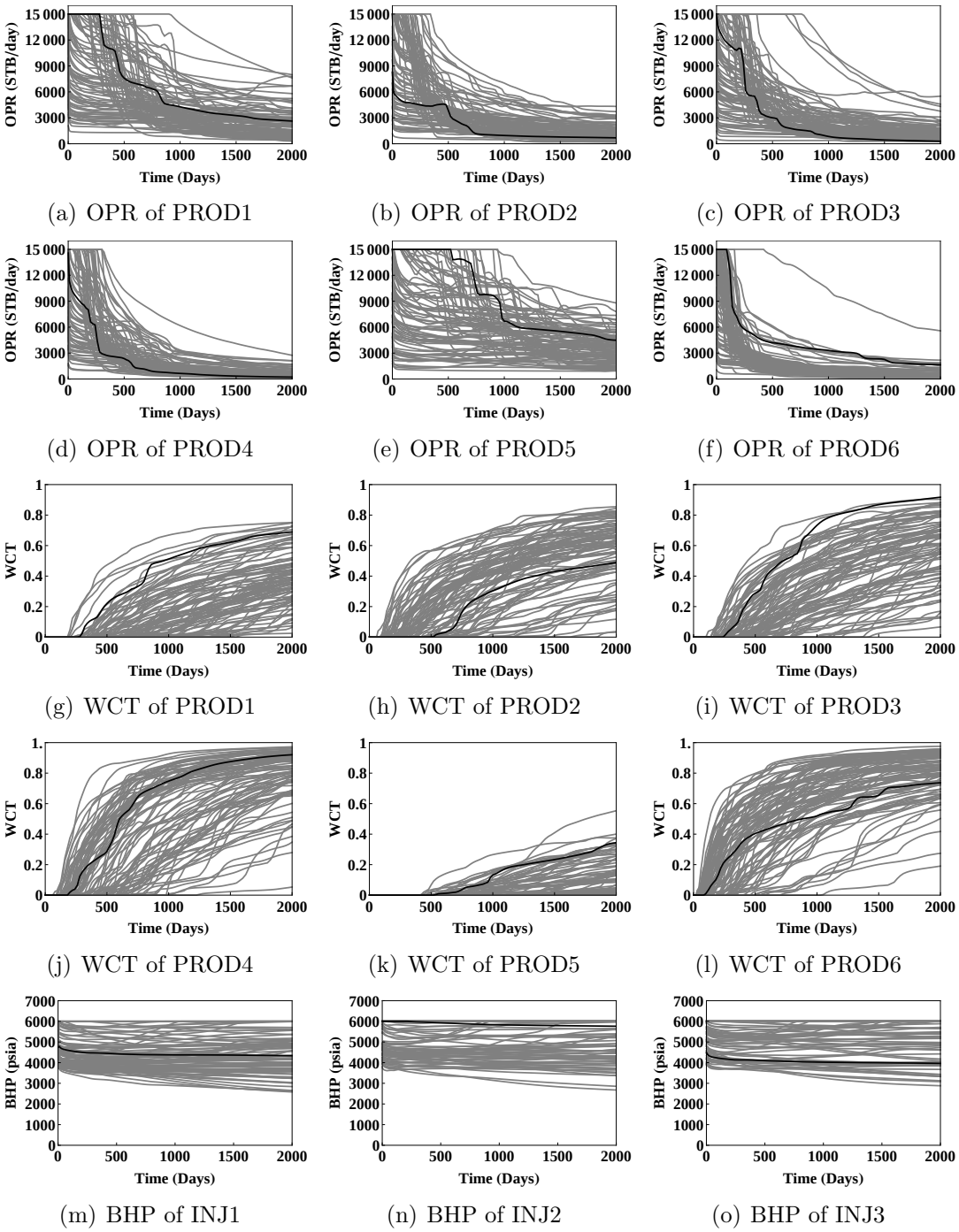
**Figure 5.10:** The profiles of the water injection rate (WIR) and bottom hole pressure (BHP) for the three injectors from the SPE9 model (reference).

and log-permeability fields are known correctly. Fig. 5.11 shows the production data obtained from the initial ensemble together with the production data from the reference model. The gray curves indicate the ensemble forecasts and the black curve shows the production data for the reference model. The different types of production data shown in Fig. 5.11 include oil production rate (OPR) and water cut (WCT) for six producers, and bottom hole pressure (BHP) for three injectors. Since the initial ensemble is not conditioned to the static data at well-locations, a large variability can be observed among ensemble members for different types of production data. The oil production rate for different producers from the reference model is contained within the spread of the initial ensemble (Figs. 5.11(a) through 5.11(f)). Each producer is completed between three layers (2 through 4) and the water breakthrough in different layers can be observed as a slight change in water cut profiles for different producers.

The static model parameters, such as porosity and log-permeability, and the dynamic state variables, such as pressure, water saturation, and gas saturation are included into the state vector. We assume the period of the first 1500 days as the production history and the period from day 1501 to day 2000 is treated as the prediction period. The types of observations available for data assimilation are oil production rate, water cut, producing gas-oil ratio, water injection rate, and bottom hole pressure (for producers and injectors) at days 300, 600, 900, 1250, and 1500. Thus, there are a total of 5 data assimilation steps and 30 observations are assimilated at each assimilation time. The measurement noise associated with the production data is 2% of the true measurement value. The same set of noisy observations are used in evaluating the performance of the standard EnKF and the IPCEnKF methods.

#### *5.5.2.2 Case 2: the initial ensemble is biased*

Similar to Case 1, the initial ensemble of porosity and log-permeability fields is generated using sequential Gaussian simulation. For Case 2, however, we assume that




(a) OPR of PROD1 (b) OPR of PROD2 (c) OPR of PROD3

(d) OPR of PROD4 (e) OPR of PROD5 (f) OPR of PROD6

(g) WCT of PROD1 (h) WCT of PROD2 (i) WCT of PROD3

(j) WCT of PROD4 (k) WCT of PROD5 (l) WCT of PROD6

(m) BHP of INJ1 (n) BHP of INJ2 (o) BHP of INJ3

**Figure 5.11:** The production forecasts from the initial ensemble (gray curves) and the reference production data (black curve). The initial ensemble is not biased (Case 1).

the mean values of the porosity and log-permeability fields are known incorrectly. The initial ensemble of porosity fields is generated with mean equal to 0.28, which is higher than that of the reference field (0.18), and standard deviation equal to 0.03. The initial ensemble of log-permeability fields is generated with mean equal to 3.5, which is lower than that of the reference field (4.5), and standard deviation equal to 1.2. Fig. 5.12 shows the production data obtained from the initial ensemble together with the reference production data. The gray curves indicate the ensemble forecasts and the black curve shows the production data for the reference model. Since the initial ensemble is generated using a higher and a lower value for the mean porosity and log-permeability field, respectively, the initial ensemble shows a delay in the water breakthrough compared to the reference model. As Fig. 5.12 indicates, the water cut at producers PROD1, PROD3, PROD4, and PROD5 from the reference model are higher than those for the initial ensemble in the entire production period.

For Case 2, the elements of the state vector, the total production period, the production history as well as the prediction periods, and the types of observations available for data assimilation are similar to those used in Case 1. The standard EnKF is known to show a poor performance for data assimilation in a situation when the reference model has water breakthrough, but none of the realizations from the initial ensemble show water cut or a few realizations have water breakthrough which is significantly lower than the actual observations. Case 2 is designed to study the performance of the standard EnKF and the IPCEnKF under a similar situation. The first data assimilation is carried out at day 500. As Fig. 5.12 shows, at the first assimilation time (day 500), a large number of realizations from the initial ensemble either show no water breakthrough or have small water cut values compared to the reference model at producers PROD1 and PROD3. Subsequently, the observations are assimilated at days 740, 900, 1250, and 1500. Thus, there are a total of 5 data assimilation steps and 30 observations are assimilated at each assimilation step. The

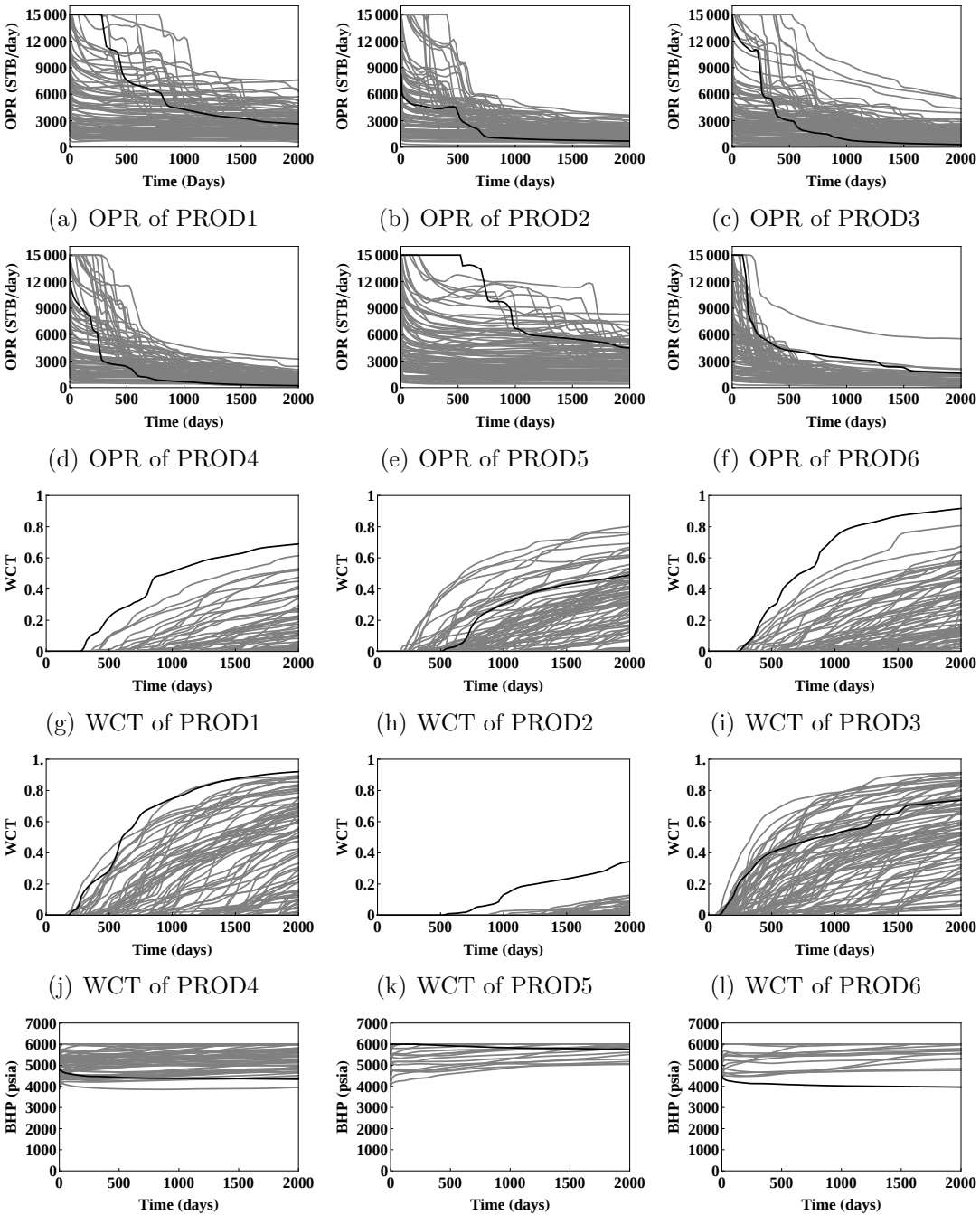
measurement noise associated with the production data is 2% of the true measurement value. Similar to Case 1, the same set of noisy observations are used in implementing the standard EnKF and the IPCEnKF.

### 5.5.3 Results and discussion

#### *5.5.3.1 Estimation of saturation fields*

Figs. 5.13 and 5.14 show the water saturation profiles before and after data assimilation at the first data assimilation time (day 300) for Case 1 along different cross-sections from different layers of the simulation model. The cross-sections are perpendicular to the direction of water movement in the reservoir and are parallel to the $y$-axis. Each cross-section is identified by its parent-layer and the $x$-index (e.g. Layer 12, $x$=15). The thin gray curves show the water saturation profiles for the ensemble members, the thick black curve is the reference, and the thick green curve is the ensemble mean. The connate water saturation for this reservoir model is 0.15 which is denoted by a thin horizontal black line. The realizations from the initial ensemble show a large variability in the forecast water saturation profiles. The ensemble mean (thick green curve) water saturation computed from the initial ensemble is significantly different from the reference model (thick black curve).

The saturation profiles are adjusted in the standard EnKF by assimilating the production data at each data assimilation time. The updated water saturation profiles from the standard EnKF show constraint violations and the posterior water saturation is less than the connate water saturation (0.15) in some region of the reservoir model. In the standard implementation of the EnKF, the updated saturations are truncated to their values at the boundary of the feasible region (e.g. 0.15 and 0.88 for water saturation) if they violate the inequality constraints on them and lie outside the feasible region. In the IPCEnKF method, the saturation profiles are adjusted by additionally considering the proximity of the prior saturation from the boundary of the feasible region. As Fig. 5.14 shows, the variability of the ensemble is reduced




(a) OPR of PROD1 (b) OPR of PROD2 (c) OPR of PROD3

(d) OPR of PROD4 (e) OPR of PROD5 (f) OPR of PROD6

(g) WCT of PROD1 (h) WCT of PROD2 (i) WCT of PROD3

(j) WCT of PROD4 (k) WCT of PROD5 (l) WCT of PROD6

(m) BHP of INJ1 (n) BHP of INJ2 (o) BHP of INJ3

**Figure 5.12:** The production forecasts from the initial ensemble (gray curves) and the reference production data (black curve). The initial ensemble is biased (Case 2).

after assimilating data using the IPCEnKF method. The posterior saturation profiles, however, do not resemble the reference in some part of the reservoir model. As we can see in Fig. 5.14(d), there are saturations out of the physical bound (less than 0.15) after the updating of the IPCEnKF. This is due to the fact that the computation of the iterative solution using the IPCEnKF involves the gradient of the barrier function with respect to the water saturation at each gridblock. In case of a reservoir model subject to water flooding, it is reasonable to have a large number of gridblocks with water saturation equal to the connate water saturation. The gradient of the barrier function evaluated at a gridblock with connate water saturation will have a large magnitude which can subsequently result in significantly small step-size for updating the state vector. Therefore, a different threshold, which is lower than the connate water saturation, is used for constraining the water saturations during the updating of the IPCEnKF. In the present study, we use a constraint of 0.10 on the water saturation.

For Case 1, we do not see a significant difference between the posterior water saturation profiles obtained from the standard EnKF and the IPCEnKF (Figs. 5.13 and 5.14) at the first assimilation time, although the ensemble updated using the standard EnKF shows a larger variability compared to the IPCEnKF. The updated water saturation profiles from the standard EnKF suggest that, using a truncation approach to truncate the non-physical water saturations can be a reasonable solution for handling the constraint violations in Case 1. The ensemble mean computed after truncating the non-physical water saturations would not be significantly different from the reference in the region of constraint violations. It appears that, using the truncation approach with the standard EnKF would not cause significant harm during the subsequent forecast step.

Fig. 5.15 shows the water saturation fields before and after data assimilation for different model-layers of realization 15 for Case 1 at the first data assimilation time

(day 300) obtained from the standard EnKF and the IPCEnKF. The corresponding saturation fields from the reference model are also shown (Figs. 5.15(j) through 5.15(l)). As mentioned earlier, the water saturations were constrained to a value of 0.10 during the implementation of the IPCEnKF. In Fig. 5.15, the region of the reservoir model where the updated water saturations are observed to take values less than this threshold are shown by white color. As Figs. 5.15(d) through 5.15(f) show, the updated water saturations from the standard EnKF show constraint violations in some region of the reservoir model. The minimum values of the posterior water saturations from the standard EnKF for layers 2, 3, and 10 are 0.0005, 0.0791, and 0.0241, respectively. The posterior water saturation fields obtained from the IPCEnKF are similar to those obtained from the standard EnKF. The IPCEnKF, however, is able to improve the estimates of the water saturation in the entire reservoir model.

Fig. 5.16 shows the prior and the posterior water saturation profiles at the first data assimilation time (day 500) for Case 2 along different cross-sections from different simulation model-layers obtained from the standard EnKF. The initial ensemble of porosity and log-permeability fields was generated using different statistical parameter (mean) compared to the reference model. As a result, the prior water saturation profiles along different cross-sections from the initial ensemble do not show resemblance to the saturation profiles from the reference model. For Case 2, a large fraction of the realizations from the initial ensemble either predict no water production or has very small water production while the observed data already show significant water breakthrough at the producers. Therefore, the standard implementation of the EnKF makes large changes to the saturation fields during the analysis step at the first data assimilation time. The posterior water saturations from the standard EnKF show strong constraint violations on both sides of the feasible region in some part of the reservoir model. As Figs. 5.16(d), 5.16(e), and 5.16(f) show, the water saturations in some region of the model take non-physical values after data assimilation and are

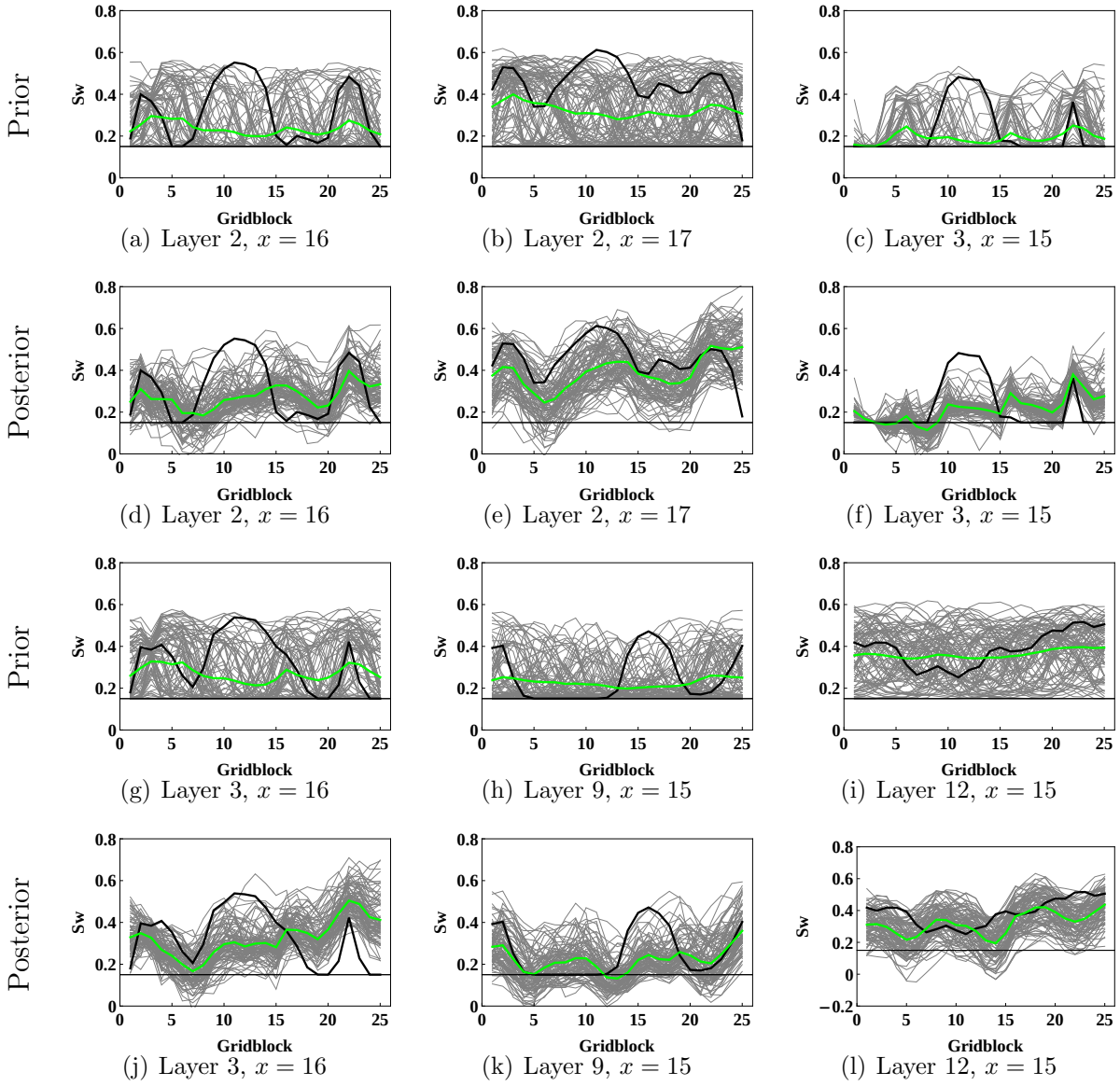
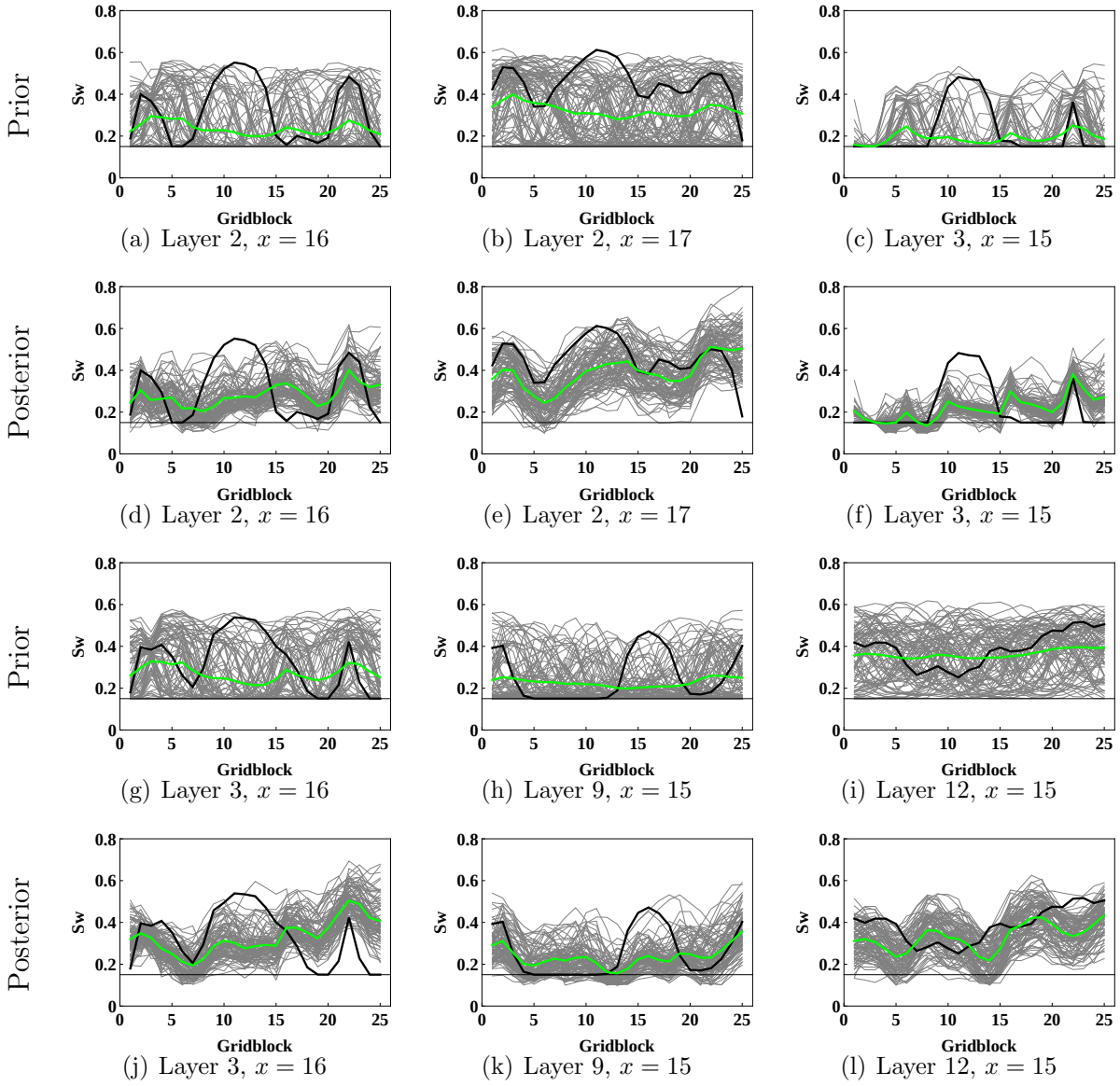



**Figure 5.13:** The water saturation profiles, before and after data assimilation, along different cross-sections at the first data assimilation time (day 300) for Case 1 from the standard EnKF. The cross-sections are considered in the $y$-direction and are denoted by the $x$-indices.




(a) Layer 2, $x = 16$ (b) Layer 2, $x = 17$ (c) Layer 3, $x = 15$

(d) Layer 2, $x = 16$ (e) Layer 2, $x = 17$ (f) Layer 3, $x = 15$

(g) Layer 3, $x = 16$ (h) Layer 9, $x = 15$ (i) Layer 12, $x = 15$

(j) Layer 3, $x = 16$ (k) Layer 9, $x = 15$ (l) Layer 12, $x = 15$

**Figure 5.14:** The water saturation profiles, before and after data assimilation, along different cross-sections at the first data assimilation time (day 300) for Case 1 from the IPCEnKF. The cross-sections are considered in the $y$-direction and are denoted by the $x$-indices.

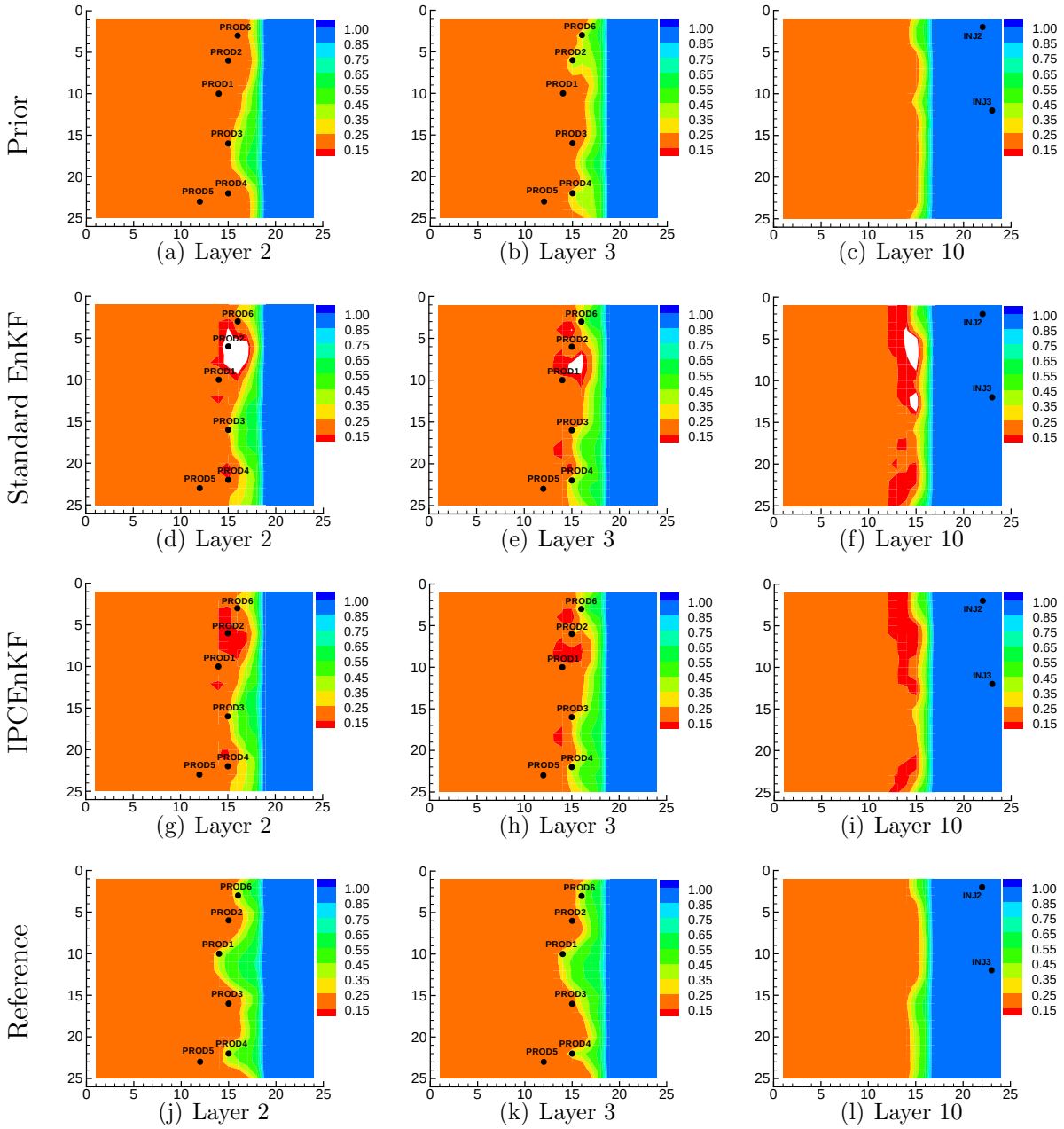



**Figure 5.15:** The prior and posterior water saturation fields for different model-layers of realization 15 at the first data assimilation time (day 300) from the standard EnKF and the IPCEnKF (Case 1) together with the corresponding reference saturation fields.

greater than 0.88. Similarly, as Figs. 5.16(d) and 5.16(j) indicate, the posterior water saturations from the standard EnKF in some region show strong constraint violations and are significantly lower than the connate water saturation (0.15). Similar to Case 1, the truncation approach is used with the standard EnKF for handling constraint violations. The posterior water saturations that are out of physical bounds (greater than 0.88 or less than 0.15) are truncated to their values at the boundary of the feasible region. Although the truncation avoids the non-physical saturations obtained from the analysis step of the standard EnKF, it affects the behavior of the ensemble members in that region of the reservoir model during the subsequent forecast step. In the region of truncated saturations, the ensemble members show significantly different flow characteristics compared to the reference model. For example, as Fig. 5.16(f) shows, the posterior water saturations are out of physical bound (greater than 0.88) in some part of the model. All such water saturations are truncated to a value of 0.88. As it can be noticed, the reference model shows water saturation of approximately 0.60 for gridblocks in this region while a large fraction of the realizations will have water saturation of 0.88 after truncation which will affect the performance of those realizations during the subsequent forecast step.

Fig. 5.17 shows the prior and the posterior water saturation profiles at the first data assimilation time (day 500) for Case 2 along different cross-sections from different simulation model-layers obtained from the IPCEnKF. From the results presented here, it is clear that, the posterior saturation profiles obtained from the IPCEnKF are better than those obtained from the standard EnKF. The posterior ensemble of saturations from the IPCEnKF has enough variability and the reference saturation profile is contained within the ensemble-spread. The mean of the posterior ensemble of saturations shows similar behavior as the reference model in most part of the reservoir model. As Figs. 5.17(d), 5.17(e), and 5.17(f) show, the updated water saturations from the IPCEnKF, unlike the standard EnKF, honor the inequality constraint and

are less than 0.88 in all parts of the reservoir model. The posterior saturation profiles obtained from the IPCEnKF (e.g. Figs. 5.17(e) and 5.17(i)) represent a better approximation to the reference model compared to those obtained from the standard EnKF (e.g. Figs. 5.16(e) and 5.16(i)).

The saturations out of the physical bounds are truncated for both the standard EnKF and the IPCEnKF before advancing the simulator to the next data assimilation time. Fig. 5.18 shows the prior and the posterior water saturation profiles along different cross-sections at the second assimilation time (day 740) for Case 2 from the two methods. At day 740, the water saturations predicted by the IPCEnKF (e.g. Figs. 5.18(g) and 5.18(h)) are closer to the reference than the saturations predicted by the standard EnKF (e.g. Figs. 5.18(a) and 5.18(b)). A large fraction of the gridblocks shown in Fig. 5.18 are in the water flood-front area at day 740. The saturations updated by the standard EnKF are unreasonable (e.g. Figs. 5.18(e) and 5.18(f)), while the saturations updated by the IPCEnKF are plausible (e.g. Figs. 5.18(k) and 5.18(l)) and generally closer to the reference. After truncating the non-physical saturations at the second assimilation time, the simulator is advanced to the next data assimilation time. At each subsequent assimilation time, the standard EnKF suffers from the poor saturations inherited from the previous assimilation time.

Fig. 5.19 shows the water saturation fields before and after data assimilation for different model-layers of realization 6 at day 500 from the standard EnKF and the IPCEnKF for Case 2. The corresponding saturation fields from the reference model are shown in Figs. 5.19(j) through 5.19(l). Because the initial ensemble of porosity and log-permeability was generated using different statistical parameter (mean) compared to the reference model, the prior saturation fields for different model-layers of realization 6 are significantly different from the corresponding reference saturation fields. The saturation fields for the reference model indicate that, at day 500, a larger area has been swept by water compared to realization 6. The saturation fields updated

by the standard EnKF show strong constraint violations and have values less than the connate water saturation (shown by white region) in some part of the reservoir model. The minimum posterior water saturations obtained from the standard EnKF for layers 2, 10, and 12 are -0.0682, -0.1142, and -0.4870, respectively. As we can see in Fig. 5.19(d), there is a region in layer 2, away from the aquifer and close to PROD1, where the updated water saturations are close to 1. The water saturations updated from the IPCEnKF are reasonable and closer to the reference saturation fields.

Fig. 5.20 shows the water saturation fields before and after data assimilation for different model-layers of realization 51 at day 740 from the standard EnKF and the IPCEnKF for Case 2 together with the corresponding reference saturation fields. Because the posterior saturations out of the physical bounds at day 500 are truncated in the standard EnKF, the prior saturation fields obtained at day 740 from the two methods are different. For the standard EnKF, the prior saturation fields for realization 51 (Figs. 5.20(a) through 5.20(c)) are different from the reference saturation fields (Figs. 5.20(m) through 5.20(o)). The updated saturations from the standard EnKF are unreasonable. The saturations updated from the IPCEnKF are plausible and closer to the reference saturation fields.

#### *5.5.3.2 Predictability of the updated ensemble*

For Cases 1 and 2, the ensemble of 90 realizations was continually adjusted by assimilating 30 production data at each assimilation time step. There were 5 data assimilation steps and thus, a total of 150 observations were assimilated during the entire history matching process. Once the data assimilation was complete over all the data assimilation time steps, the reservoir models were initialized with the final updated porosity and permeability fields together with the initial reservoir equilibrium conditions. The reservoir models were then evaluated from the initial time (day 0) up to the end of the production period (day 2000) using a commercial reservoir

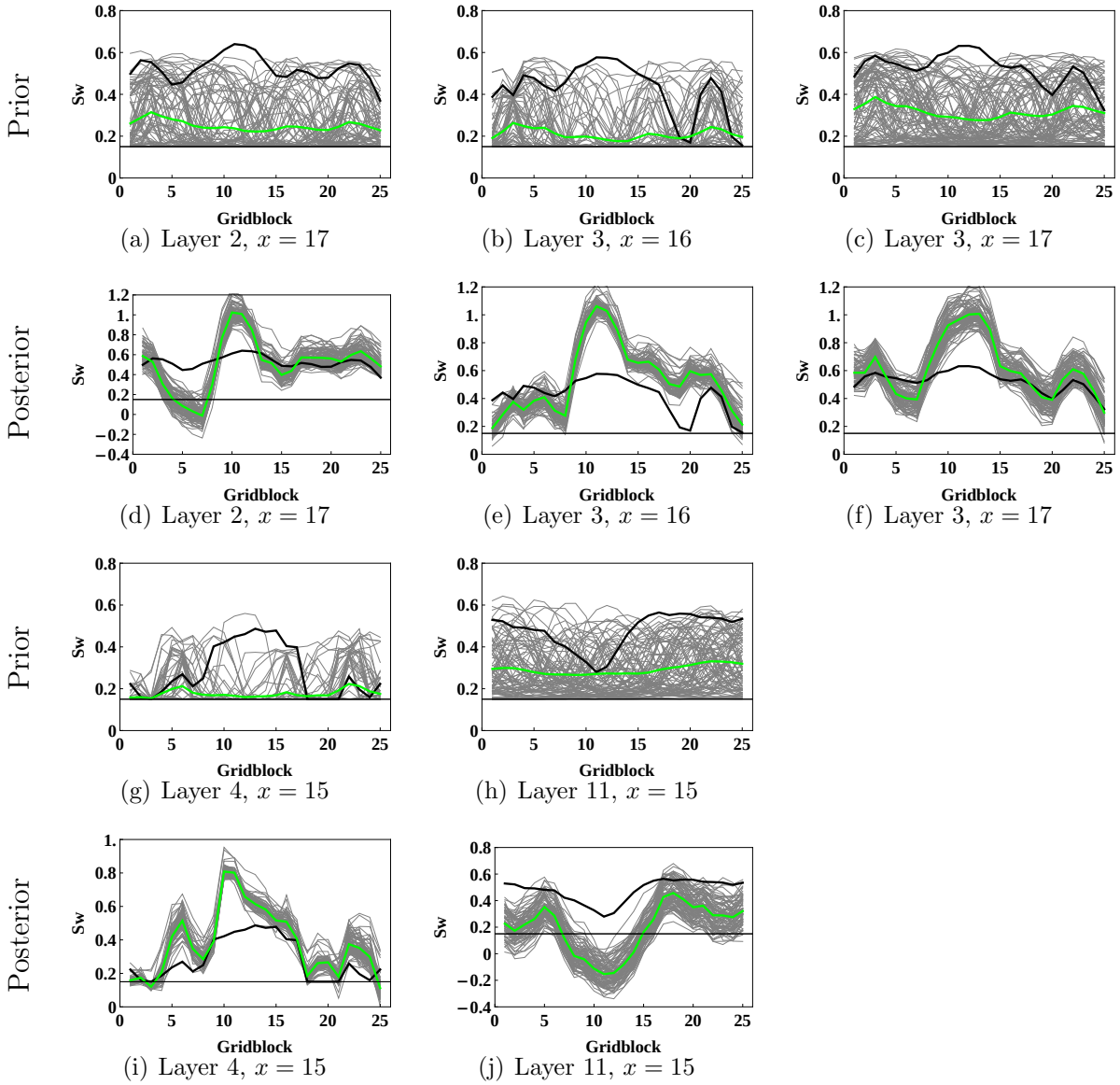
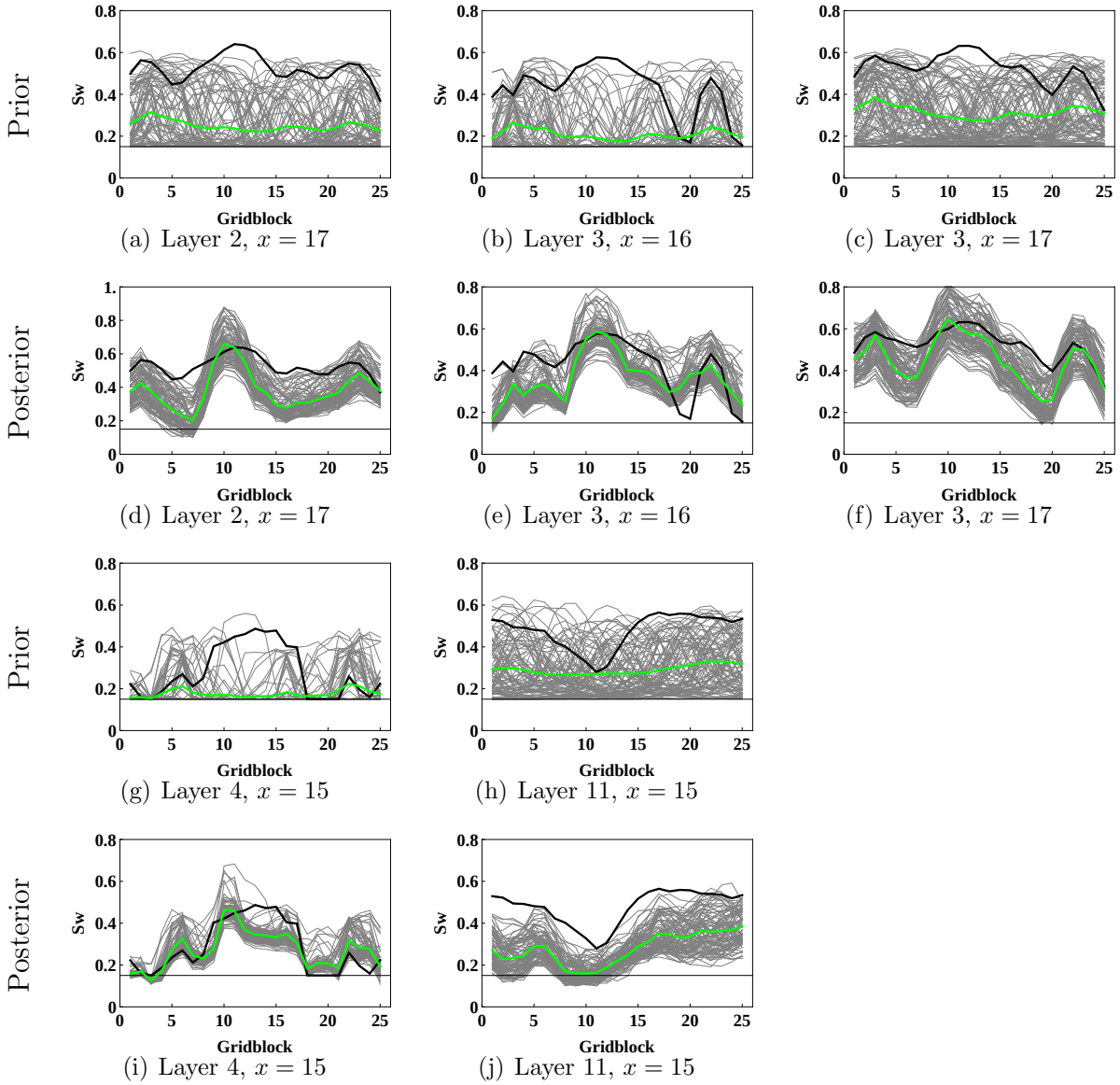



**Figure 5.16:** The water saturation profiles, before and after data assimilation, along different cross-sections at the first data assimilation time (day 500) for Case 2 from the standard EnKF. The cross-sections are considered in the $y$-direction and are denoted by the $x$-indices.



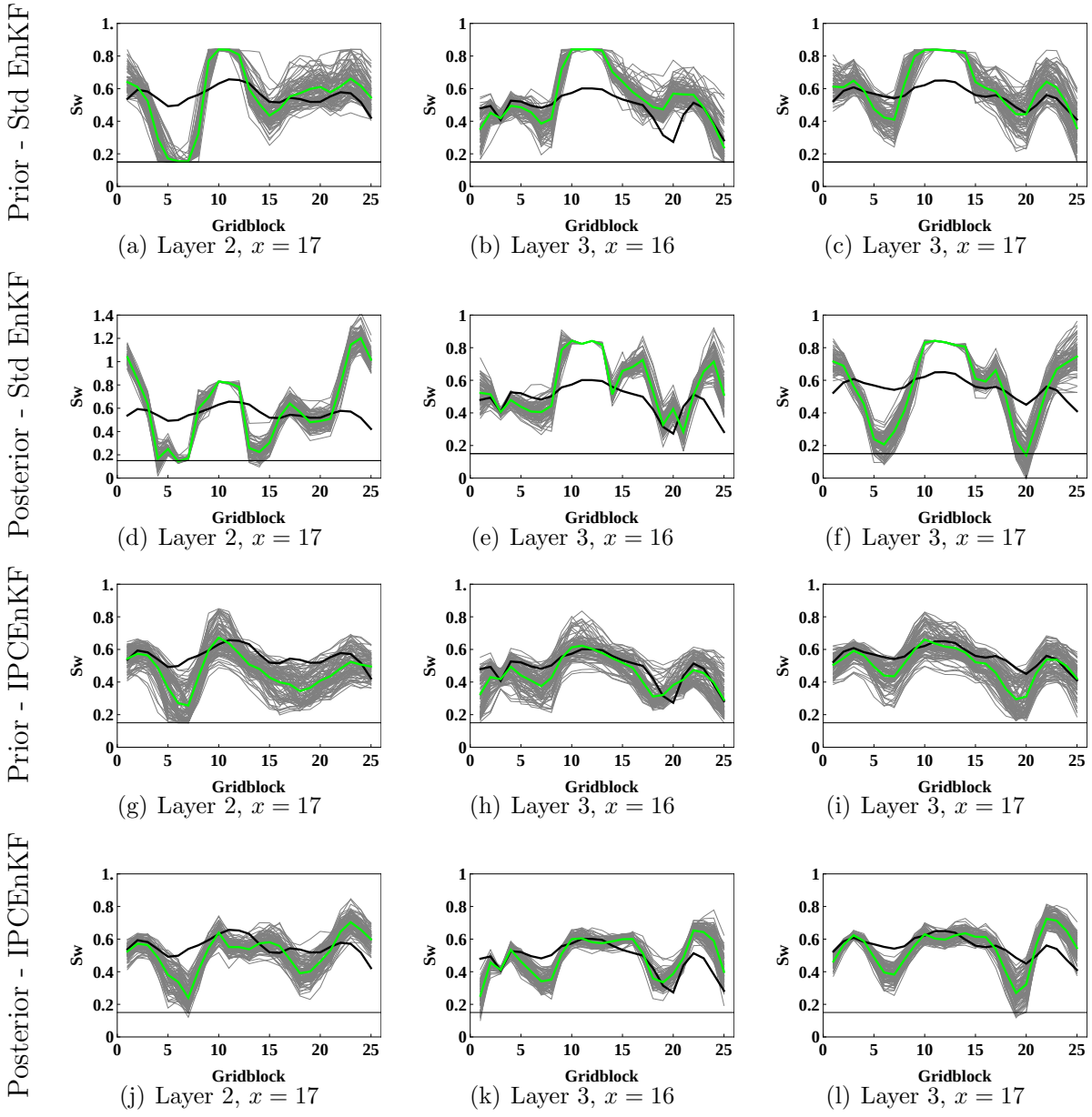

(a) Layer 2, $x = 17$ (b) Layer 3, $x = 16$ (c) Layer 3, $x = 17$

(d) Layer 2, $x = 17$ (e) Layer 3, $x = 16$ (f) Layer 3, $x = 17$

(g) Layer 4, $x = 15$ (h) Layer 11, $x = 15$

(i) Layer 4, $x = 15$ (j) Layer 11, $x = 15$

**Figure 5.17:** The water saturation profiles, before and after data assimilation, along different cross-sections at the first data assimilation time (day 500) for Case 2 from the IPCEnKF. The cross-sections are considered in the $y$-direction and are denoted by the $x$-indices.




(a) Layer 2, $x = 17$ (b) Layer 3, $x = 16$ (c) Layer 3, $x = 17$

(d) Layer 2, $x = 17$ (e) Layer 3, $x = 16$ (f) Layer 3, $x = 17$

(g) Layer 2, $x = 17$ (h) Layer 3, $x = 16$ (i) Layer 3, $x = 17$

(j) Layer 2, $x = 17$ (k) Layer 3, $x = 16$ (l) Layer 3, $x = 17$

**Figure 5.18:** The water saturation profiles, before and after data assimilation, along different cross-sections at the second data assimilation time (day 740) for Case 2 from the standard EnKF and the IPCEnKF. The cross-sections are considered in the $y$-direction and are denoted by the $x$-indices.

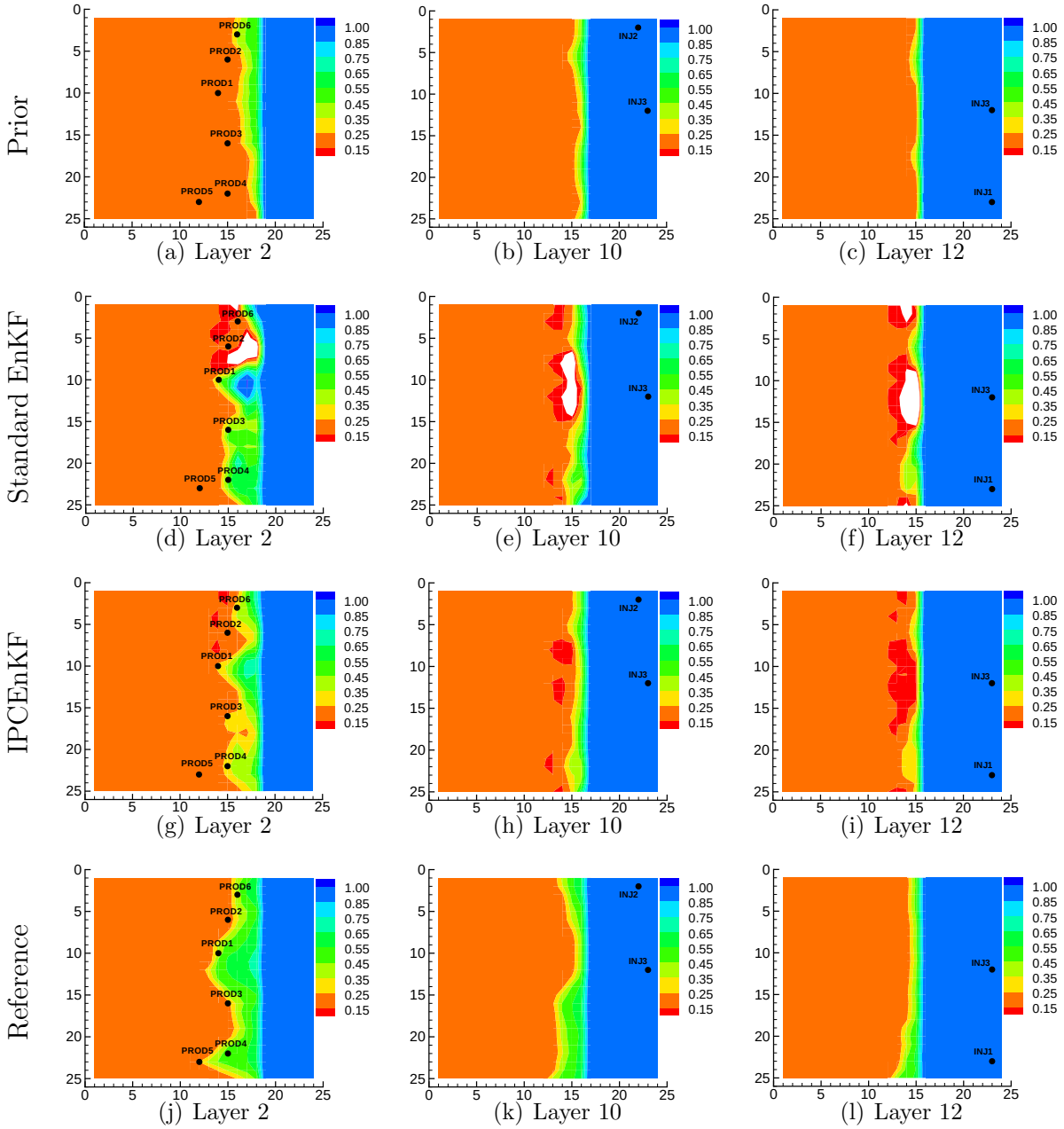



**Figure 5.19:** The prior and posterior water saturation fields for different model-layers of realization 6 at the first data assimilation time (day 500) from the standard EnKF and the IPCEnKF (Case 2) together with the corresponding reference saturation fields.

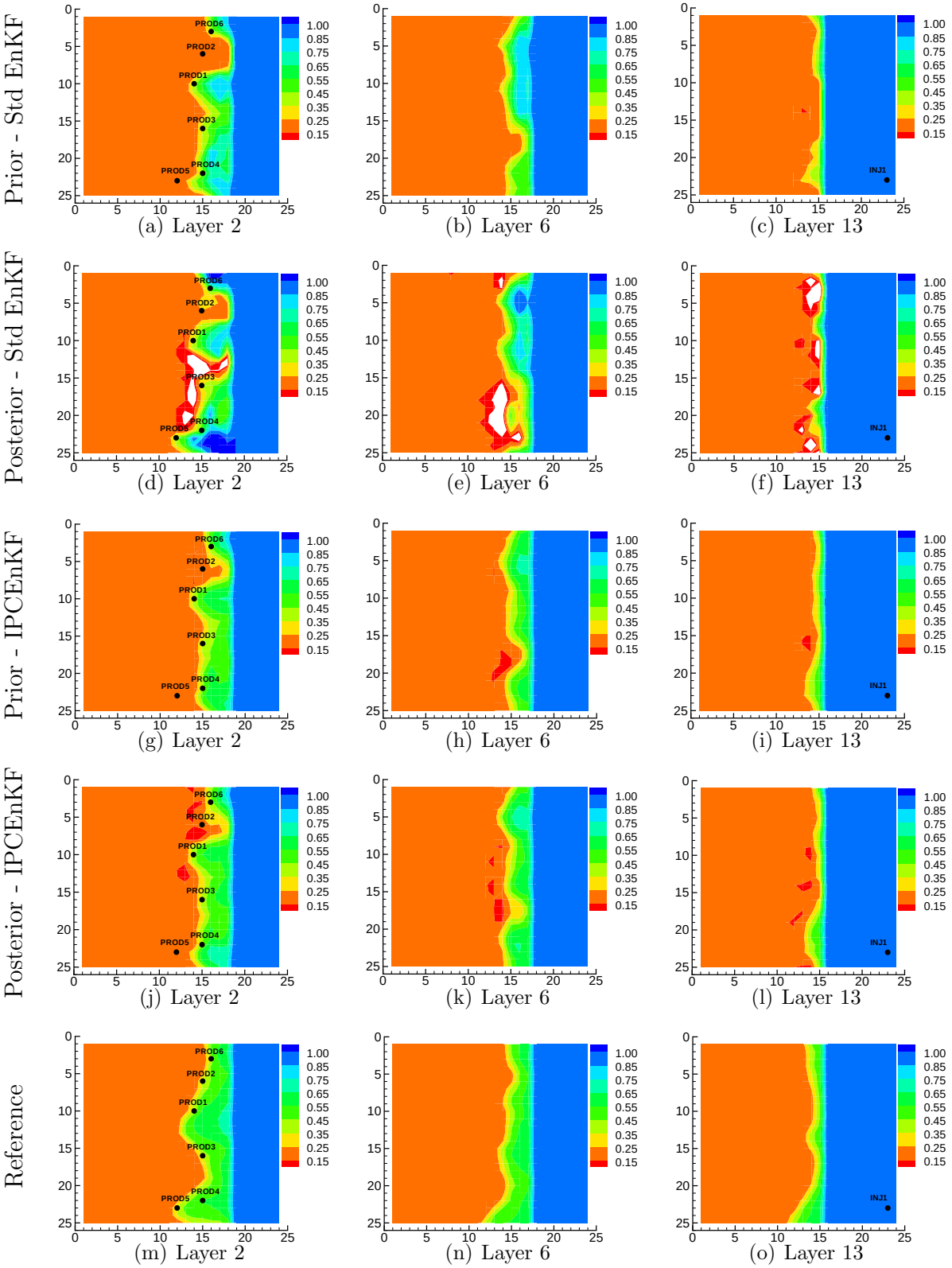



**Figure 5.20:** The prior and posterior water saturation fields for different model-layers of realization 51 at the second data assimilation time (day 740) from the standard EnKF and the IPCEnKF (Case 2) together with the corresponding reference saturation fields.

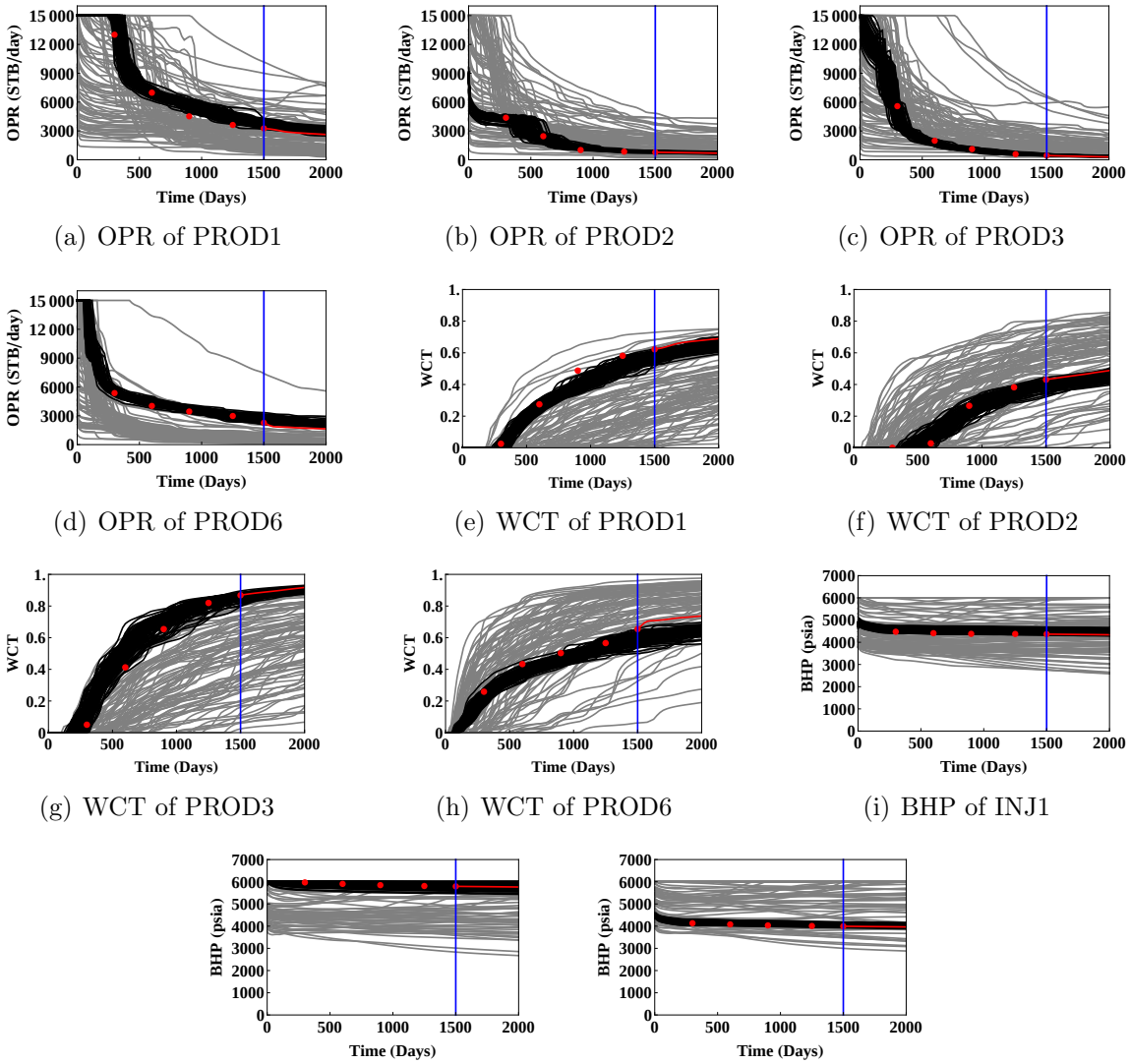
simulator.

Figs. 5.21 and 5.22 show the results of the production data for Case 1 obtained by rerunning the final updated ensembles from the standard EnKF and the IPCEnKF methods, respectively. The results include the oil production rate (OPR) and water cut (WCT) for different producers and bottom hole pressure (BHP) for three injectors. The gray and the black curves denote the response from the initial and the updated ensemble, respectively, the red dots are the true observations from the reference model, the red curve is the reference production data in the prediction period, and the vertical blue line indicates the end of the history matching period (day 1500). The updated ensemble from the standard EnKF has large variability in the history matching period and shows a bias in the prediction period. A large fraction of the updated realizations from the standard EnKF show higher oil production rates at the producer PROD1 (Fig. 5.21(a)) and PROD6 (Fig. 5.21(d)) and lower water cut at the producers PROD1 (Fig. 5.21(e)), PROD2 (Fig. 5.21(f)), and PROD6 (Fig. 5.21(h)) compared to the reference model. A large fraction of the updated realizations from the standard EnKF also show an early water breakthrough at the producer PROD2 (Fig. 5.21(f)) compared to the true observation. As we can see in Fig. 5.22, for Case 1, the proposed IPCEnKF method provided improved history matching results compared to the standard EnKF. The updated ensemble obtained from the IPCEnKF performs fairly well in the prediction period for different types of production data. The predictions of oil production rate and water cut at the producer PROD6, however, are biased compared to the reference model. The RMSE for different types of production data from the standard EnKF and the IPCEnKF for Case 1 are given in Table 5.2.

Figs. 5.23 and 5.24 show the production forecasts for Case 2 obtained by rerunning the final updated ensembles from the standard EnKF and the IPCEnKF, respectively. The updated ensemble from the standard EnKF show a strong bias and a poor match

to the observations of OPR (Fig. 5.23(a)) and WCT (Fig. 5.23(e)) at the producer PROD1. The data-match to the observations of OPR at the producers PROD2 (Fig. 5.23(b)), PROD3 (Fig. 5.23(c)), and PROD6 (Fig. 5.23(d)) are reasonable. The forecasts from the updated ensemble from the standard EnKF, however, are biased in the prediction period and predict higher oil production rates compared to the actual observations. Similarly, the updated ensemble from the standard EnKF does not show a good match to the observations of WCT at the producers PROD2 (Fig. 5.23(f)) and PROD6 (Fig. 5.23(h)) and the ensemble forecasts in the prediction period are biased compared to the reference model. As we can see in Fig. 5.23(j), the updated ensemble from the standard EnKF underpredicts the BHP at the injector INJ2 throughout the production period. The updated ensemble also shows a bias for the BHP of the injector INJ3 as most of the updated realizations overpredict the BHP at INJ3. From the results presented in Fig. 5.24, it is clear that, the IPCEnKF method was able to improve the history matching results compared to the standard EnKF. The data-match to the observations of OPR (Fig. 5.24(a)) and WCT (Fig. 5.24(e)) at the producer PROD1 and BHP (Fig. 5.24(j)) at the injector INJ2 has been improved using the IPCEnKF. Neither the standard EnKF nor the IPCEnKF was able to obtain a good match to the WCT observations at the producer PROD3 in the history-matching period, although the ensemble forecasts in the prediction period are better from the IPCEnKF than those obtained from the standard EnKF. The RMSE for different types of production data from the standard EnKF and the IPCEnKF for Case 2 are given in Table 5.3.

Similar to the CEnKF method presented in Chapter 4, the IPCEnKF method proposed here is an iterative process. The computational cost required during each iteration of the two methods, however, is small compared to the cost of running the forward model. We have demonstrated that the constrained solutions obtained from the two methods are plausible and often closer to the reference model. The resulting



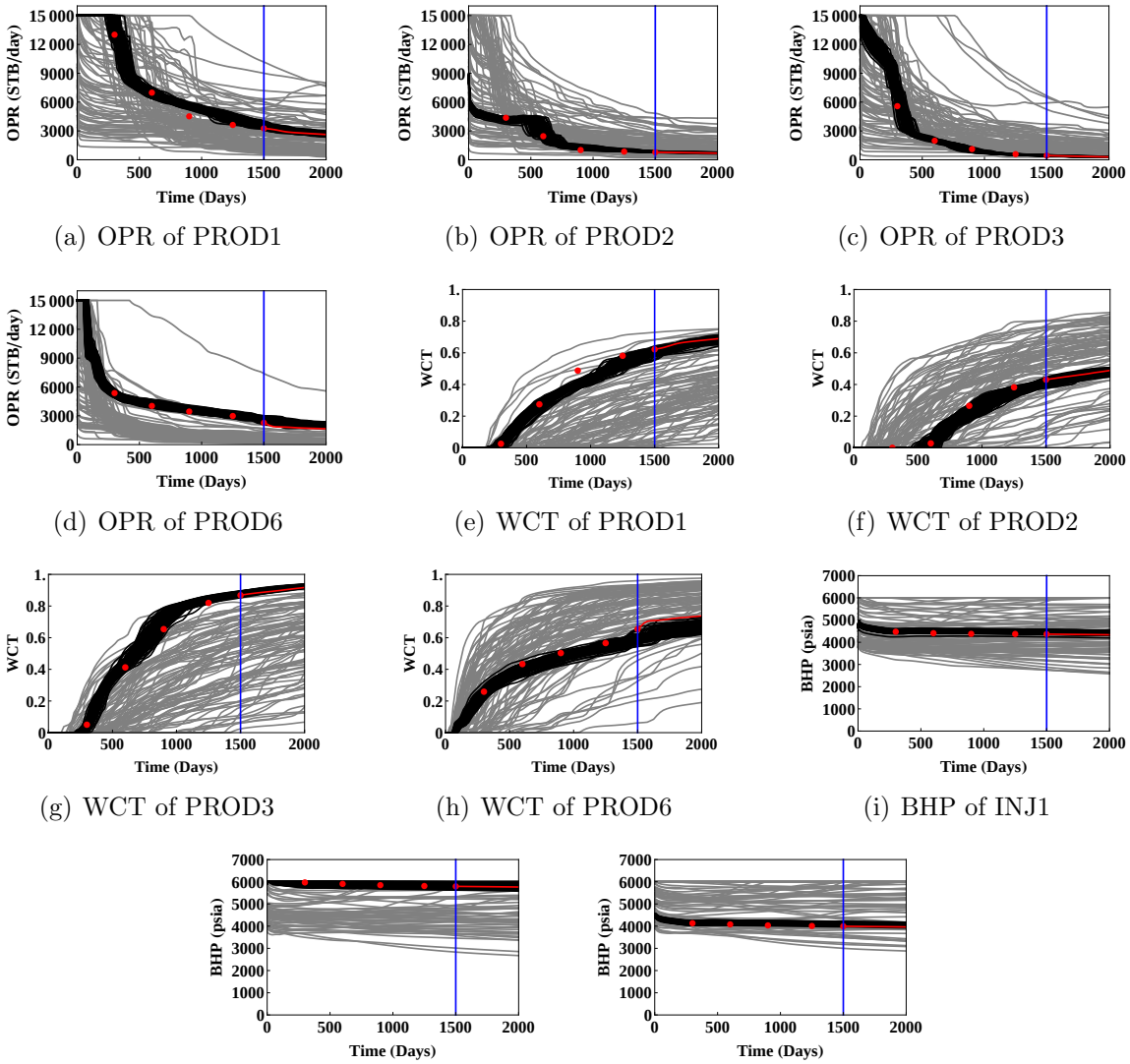

(a) OPR of PROD1 (b) OPR of PROD2 (c) OPR of PROD3

(d) OPR of PROD6 (e) WCT of PROD1 (f) WCT of PROD2

(g) WCT of PROD3 (h) WCT of PROD6 (i) BHP of INJ1

(j) BHP of INJ2 (k) BHP of INJ3

**Figure 5.21:** The production data for Case 1 predicted from the ensemble of porosity and permeability updated at day 1500 by the standard EnKF. The gray curves represent the initial ensemble, the black curves represent the updated ensemble, the red dots are the observations from the reference model, the red curve is the reference production data in the prediction period, and the vertical blue line indicates the end of the history matching period (day 1500).



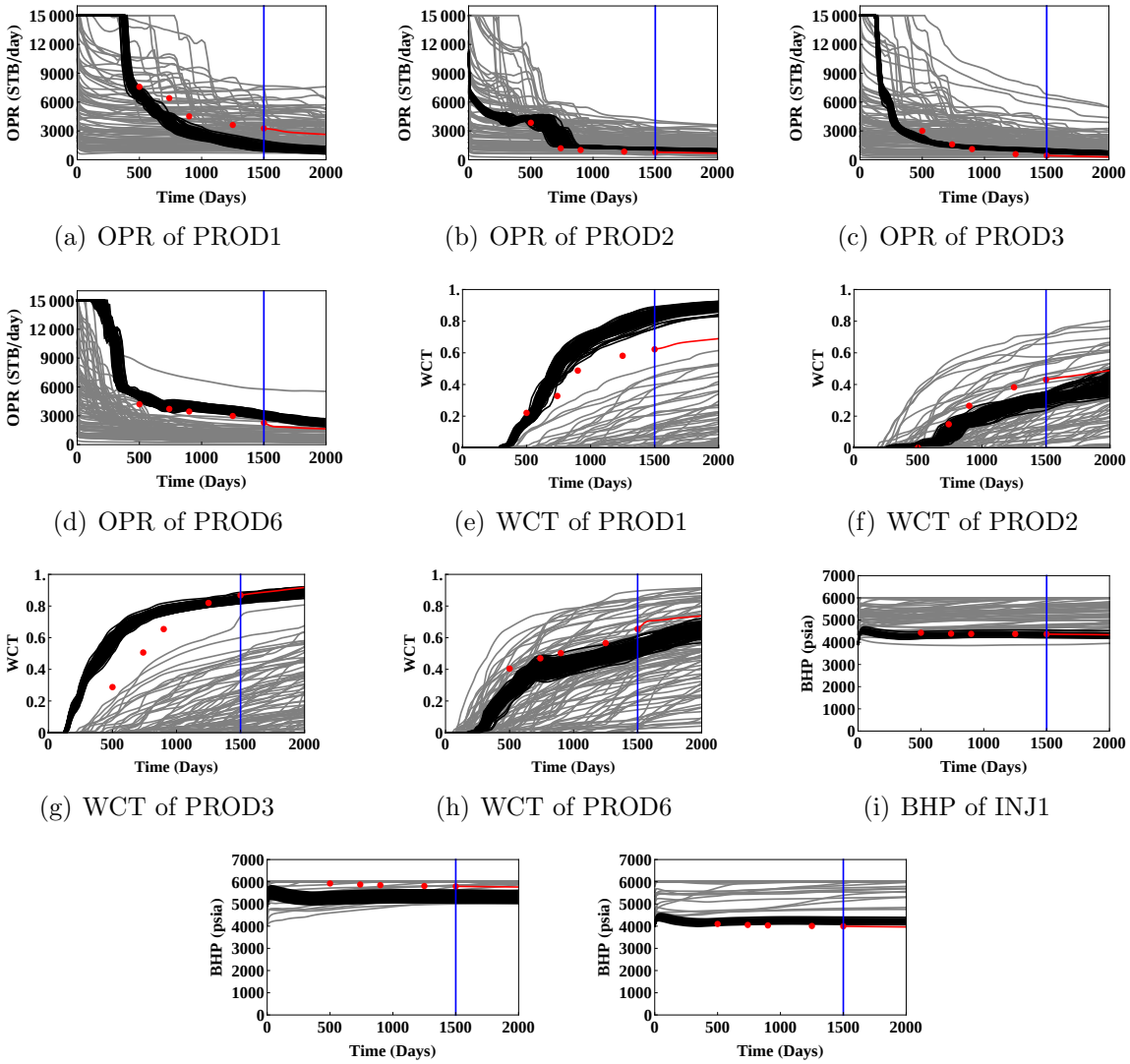

(a) OPR of PROD1 (b) OPR of PROD2 (c) OPR of PROD3

(d) OPR of PROD6 (e) WCT of PROD1 (f) WCT of PROD2

(g) WCT of PROD3 (h) WCT of PROD6 (i) BHP of INJ1

(j) BHP of INJ2 (k) BHP of INJ3

**Figure 5.22:** The production data for Case 1 predicted from the ensemble of porosity and permeability updated at day 1500 by the IPCEnKF. The gray curves represent the initial ensemble, the black curves represent the updated ensemble, the red dots are the observations from the reference model, the red curve is the reference production data in the prediction period, and the vertical blue line indicates the end of the history matching period (day 1500).



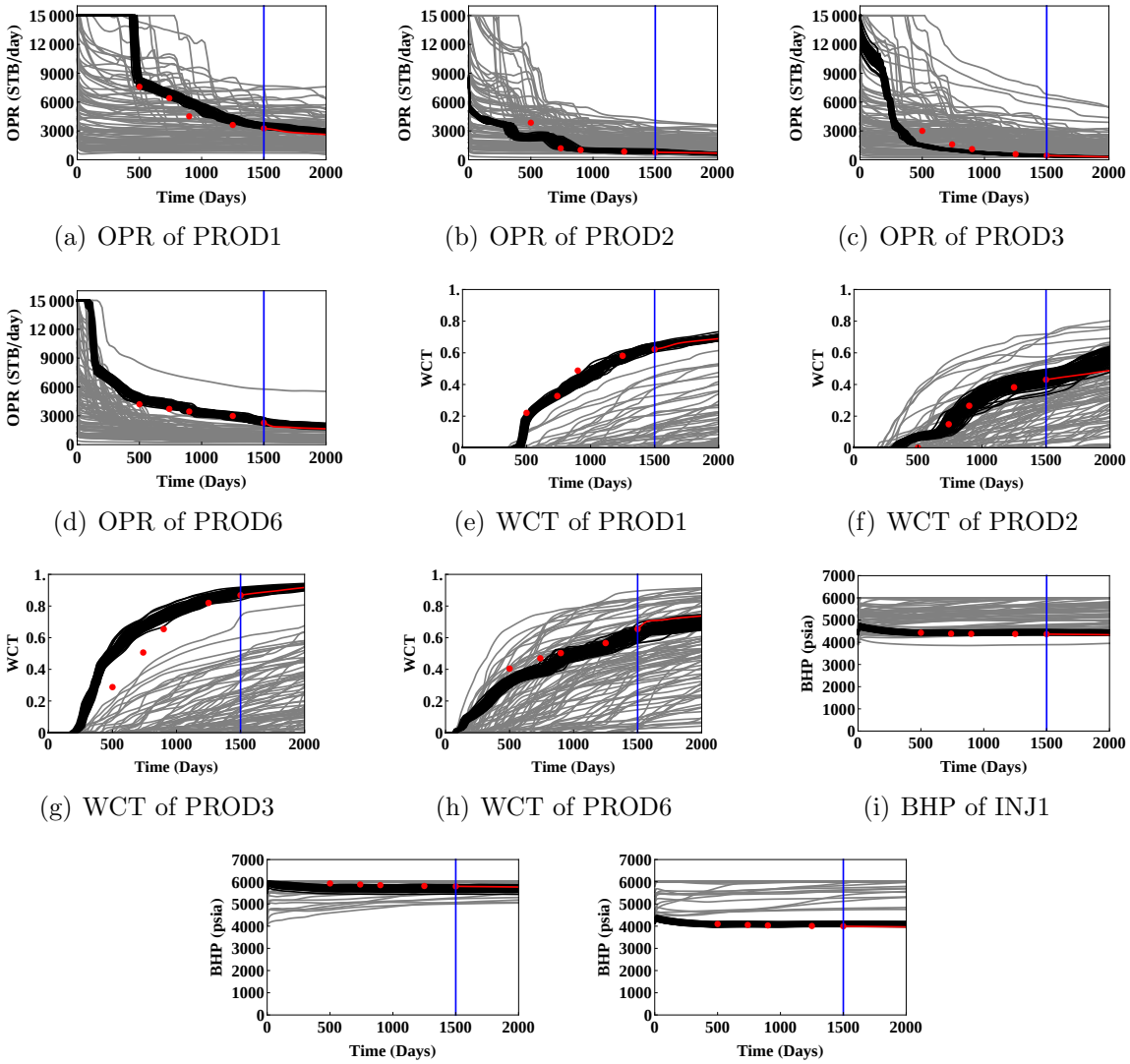

(a) OPR of PROD1 (b) OPR of PROD2 (c) OPR of PROD3

(d) OPR of PROD6 (e) WCT of PROD1 (f) WCT of PROD2

(g) WCT of PROD3 (h) WCT of PROD6 (i) BHP of INJ1

(j) BHP of INJ2 (k) BHP of INJ3

**Figure 5.23:** The production data for Case 2 predicted from the ensemble of porosity and permeability updated at day 1500 by the standard EnKF. The gray curves represent the initial ensemble, the black curves represent the updated ensemble, the red dots are the observations from the reference model, the red curve is the reference production data in the prediction period, and the vertical blue line indicates the end of the history matching period (day 1500).




(a) OPR of PROD1 (b) OPR of PROD2 (c) OPR of PROD3

(d) OPR of PROD6 (e) WCT of PROD1 (f) WCT of PROD2

(g) WCT of PROD3 (h) WCT of PROD6 (i) BHP of INJ1

(j) BHP of INJ2 (k) BHP of INJ3

**Figure 5.24:** The production data for Case 2 predicted from the ensemble of porosity and permeability updated at day 1500 by the IPCEnKF. The gray curves represent the initial ensemble, the black curves represent the updated ensemble, the red dots are the observations from the reference model, the red curve is the reference production data in the prediction period, and the vertical blue line indicates the end of the history matching period (day 1500).

**Table 5.2:** SPE9 model - RMSE values for different types of production data for Case 1

| Observation | Standard EnKF | IPCEnKF |
|---|---|---|
| OPR - PROD1 (STB/day) | 978.41 | 750.29 |
| OPR - PROD2 (STB/day) | 227.23 | 280.62 |
| OPR - PROD3 (STB/day) | 384.39 | 448.41 |
| OPR - PROD6 (STB/day) | 359.55 | 254.03 |
| WCT - PROD1 | 0.0553 | 0.0424 |
| WCT - PROD2 | 0.0381 | 0.0265 |
| WCT - PROD3 | 0.0685 | 0.0358 |
| WCT - PROD6 | 0.0515 | 0.0384 |
| BHP - INJ1 (psia) | 145.75 | 99.07 |
| BHP - INJ2 (psia) | 109.64 | 91.82 |
| BHP - INJ3 (psia) | 79.74 | 90.74 |

**Table 5.3:** SPE9 model - RMSE values for different types of production data for Case 2

| Observation | Standard EnKF | IPCEnKF |
|---|---|---|
| OPR - PROD1 (STB/day) | 1782.63 | 701.55 |
| OPR - PROD2 (STB/day) | 561.50 | 481.98 |
| OPR - PROD3 (STB/day) | 452.95 | 507.46 |
| OPR - PROD6 (STB/day) | 650.49 | 381.31 |
| WCT - PROD1 | 0.1599 | 0.0364 |
| WCT - PROD2 | 0.0869 | 0.0413 |
| WCT - PROD3 | 0.1356 | 0.1042 |
| WCT - PROD6 | 0.1023 | 0.0559 |
| BHP - INJ1 (psia) | 91.44 | 82.51 |
| BHP - INJ2 (psia) | 539.43 | 164.46 |
| BHP - INJ3 (psia) | 181.97 | 79.46 |

match to the historical production data obtained from the two methods is better than the results obtained from the standard implementation of the EnKF with truncation. The additional computational costs for the two methods are different and are affected by different components that make-up the solution. For the CEnKF, the additional computational cost arises from the identification of the active constraints for each ensemble member during each iteration of the process. Subsequently, a separate Kalman gain is computed and each ensemble member is updated separately during the iterative process. For the IPCEnKF, the additional computational cost during each iteration comes from approximating the simulated data, computing the gradient of the barrier function at the current estimate of the state variables, computing the search direction, and determining the step-size using the line search for the individual ensemble member. For both constrained data assimilation methods (CEnKF and IPCEnKF), the computational cost depends on the cutoff values chosen for the state variable of interest. Selecting cutoff values in the close vicinity of the true inequality constraints may increase the computational cost. It is possible to decrease the computational cost (but still higher than that of the standard EnKF) for the constrained data assimilation methods if the cutoff values are sufficiently far from the true inequality constraints. It, however, may not show significant improvements in the estimates of the state variables and the match to the historical production data compared to the standard EnKF. In the current study, we observed that, the total computational time for the CEnKF method is approximately 2 times the computational time required for the standard EnKF. The computational time for the IPCEnKF method is observed to be 2.5 to 2.8 times the time needed for the standard EnKF. Although the computational cost for both constrained data assimilation methods is higher than that of the standard EnKF, the results presented in this study clearly show the benefits of incorporating the constraints into the data assimilation framework. Both constrained data assimilation methods were able to improve the estimates of the state variables

(e.g. molar densities or phase saturations) and obtained good history matching results compared to the standard EnKF. In some cases (e.g. compositional model), the application of constraints instead of truncation is critical and the only real alternative is iteration that involves running the reservoir simulator during each iteration, which can be quite expensive if it has to be done very often.

The standard implementation of the EnKF with truncation often gives satisfactory results for history matching. In some situations, however, it may not be clear as which state variables are subject to truncation or what the truncation-limits are. As we have demonstrated with the compositional model in Chapter 3, the $CO_2$ molar density in some part of the reservoir model became negative while the molar densities of the rest of the components were increased after assimilating data. It was necessary to apply the constraints for the compositional model since neither truncation, nor transformation of state variables gave good results. In case of black-oil reservoir models, it will be necessary to apply the constraints on the phase saturations when all three phase saturations are important in some regions of the model. As we showed in this chapter, the updated phase saturations can show strong constraint violations after data assimilation when the ensemble forecasts are significantly different from the actual observations. Truncating the non-physical phase saturations can have negative effect during the subsequent forecast step and so, we should resolve to the constrained data assimilation approach. For real-field reservoir problems, the reservoir simulation model often contains few hundred thousands of gridblocks. If there are large regions where the phase saturations show strong constraint violations after data assimilation, the CEnKF method may not be able to give good solution if a few number of iterations are used. It might be expensive to use a large number of iterations of the CEnKF method and therefore, resolving to the IPCEnKF method might be more advantageous.

A data assimilation problem of a reservoir model with faults is another example

where the constrained data assimilation approach is expected to give good results compared to the standard EnKF with truncation. In a reservoir model, the faults or the flow-barriers are often simulated using the transmissibility multipliers in certain parts of the model. The transmissibility multipliers are often generated using a uniform distribution and have a lower and an upper limits. When data are assimilated using the standard EnKF, it is reasonable to assume that the updated transmissibility multipliers violate the constraints. The transmissibility multipliers out of the physical bounds will be truncated to their limiting values in the standard EnKF framework. The incorrect estimates of fault transmissibility multipliers can have negative influence on the flow behavior of the reservoir model leading to poor results. The constrained data assimilation approach (CEnKF or IPCEnKF), proposed in this dissertation, provides a logical and reasonable solution for handling the constraints on the fault transmissibility multipliers during data assimilation.

## 5.6 Summary

In this chapter, we proposed a new method, called the interior-point formulation for the constrained EnKF (IPCEnKF), for incorporating the inequality constraints into the sequential data assimilation framework using the EnKF. The proposed IPCEnKF is the ensemble-based formulation of the interior-point method for constrained data assimilation and combines the standard EnKF with the primal interior-point method. In the proposed method, the original objective function for data assimilation, subject to constraints, is reformulated by adding a barrier function which prevents the state variables from moving into the infeasible region. The original constrained optimization problem is thus converted into an unconstrained optimization problem. The solution of the resulting unconstrained optimization problem is obtained using the Gauss-Newton method in its classical form. The IPCEnKF method updates the state vector in an iterative manner; the method, however, does not require additional

model evaluations for constraint application.

The IPCEnKF method is applied to a synthetic 1D linear example in order to demonstrate its effectiveness when the non-linearity is not an issue. The second test problem shows the application of the IPCEnKF technique to a data assimilation problem of a black-oil reservoir example (SPE9 model) involving three-phase flow conditions for which the water and gas saturations are subject to the inequality constraints on both sides of the feasible region. The performance of the IPCEnKF was evaluated on two different cases. The results obtained from the IPCEnKF were compared with those obtained from the standard EnKF. The application of the IPCEnKF method to the SPE9 model demonstrated the importance of incorporating the constraints on the state variables into the data assimilation framework and showed the advantages of the constrained solution. When the initial ensemble was biased, the IPCEnKF gave a better match to the historical production data compared to the standard EnKF. One of the important components of the IPCEnKF method is the sensitivity matrix. Although we would like to compute the sensitivity during each iteration of the IPCEnKF, it will make the method impractical for real-field reservoir data assimilation problems. In the present study, the sensitivity computed from the prior state variables and forecast observations was used during the iterative process. The performance of the proposed IPCEnKF method can be improved at the expense of running the reservoir simulator and recomputing the sensitivity at a particular data assimilation step which shows a large data-mismatch between the ensemble predictions and the true observations.

The results from the reservoir fluid flow example show the usefulness of the proposed IPCEnKF method for enforcing constraints on the dynamic state variables such as phase saturations. Although it has not been demonstrated in this study, we believe that the IPCEnKF method can be useful for applying the constraints to the static model parameters such as the fault transmissibility multipliers. The IPCEnKF

method is an iterative process, however, the additional computational cost and data-storage requirement during each iteration of the IPCEnKF method is fairly small. As a result, the implementation of the IPCEnKF method on a real-field reservoir problem should not impose any computational or data-storage challenge.

# CHAPTER VI

# CONCLUSIONS

Long-term reservoir management objectives often include identifying production strategies that can maximize the future reservoir performance and requires an ensemble of reservoir simulation models for quantifying the uncertainty in the reservoir performance. With the increasing use of the closed-loop reservoir management practices, an ensemble of reservoir models is continually updated using the production data obtained from the field. In the recent years, the ensemble Kalman filter (EnKF) has established itself as one of the most successful techniques for sequential reservoir model updating. The EnKF represents the probability of the model parameters through an ensemble of model realizations and reduces the dimensionality of the inverse problem from the number of unknown parameters to the number of realizations. The standard implementation of the EnKF, however, suffers from a number of challenges when it is applied to a strongly nonlinear data assimilation problem. This dissertation addresses a specific issue of the violation of the inequality constraints on the plausible values of state variables. In this dissertation, we discussed this issue in the context of the EnKF applied to highly nonlinear problems in reservoir engineering with non-Gaussian state variables. The usefulness of the methods and the findings discussed in this dissertation, however, is not limited to reservoir engineering-related applications. This chapter summarizes the findings presented in this dissertation.

The standard implementation of the EnKF to a data assimilation problem of a compositional reservoir model showed that the analysis step of the EnKF resulted in non-physical values of $CO_2$ molar densities and the molar densities of the rest of the components were increased at some locations in the reservoir. A truncation

scheme that was used to truncate the state variables with non-physical values did not affect the increased molar densities of the rest of the components. The result was an inability to maintain fluid injection into the reservoir. An alternate approach of transformation of state variables is often used to ensure that the updated state variables honor the inequality constraints. In the case of the compositional model, however, we showed that, the transformation of variables approach not only destroyed the consistency of the fluid composition but it did not prevent the molar densities of other components from increasing after data assimilation.

A constrained version of the EnKF (CEnKF) that does not require additional model evaluations for constraint application is presented in this dissertation. The application of the CEnKF technique to a data assimilation problem of a compositional model resolved the issues of excess mass and pressure build-up that were observed during the application of the standard EnKF with truncation. In case of a data assimilation problem of a black-oil reservoir model, the CEnKF method presents a viable option for handling three-phase flow conditions such that the updated phase saturations are plausible. The applications of the CEnKF method to compositional and black-oil reservoir models clearly showed the benefits of the constrained solution and obtained good matches to the historical production data. We also showed that, in cases where mass conservation of a component is important, the CEnKF gives superior results to the standard truncation scheme.

The second method presented in this dissertation is the ensemble-based formulation of the interior-point method for the constrained data assimilation and combines the standard EnKF with the primal interior-point method. Although the proposed method is iterative, it does not require identifying the active constraints during each iteration and most importantly, it does not require additional model evaluations for constraint application. The effectiveness of the proposed method is demonstrated by its application to a data assimilation problem of a three-phase black-oil reservoir

model. By incorporating the inequality constraints on phase saturations into the data assimilation framework, we showed that the resulting estimates of phase saturations are plausible and better than those obtained from truncation. It is possible to further improve the performance of the proposed IPCEnKF method by recomputing the sensitivity at a particular data assimilation time which shows a large data-mismatch between the ensemble predictions and the true observations. In this investigation, the primal interior-point method was used and the barrier parameter was updated according to the decrease in the objective function during each iteration. We believe that, the primal-dual interior-point method, which provides a better solution for updating the barrier parameter, can help decrease the number of iterations. This approach, however, increases the problem-dimensions and therefore, the net effect on the overall computational cost can be an area of further study.

The usefulness of the constrained data assimilation methods presented in this dissertation is not limited to the dynamic state variables alone. Although we have not demonstrated in this study, the proposed methods can be used for applying constraints on the static model parameters such as fault transmissibility multipliers. In certain situations, it is of interest to have a particular parameter (e.g. total hydrocarbon volume) unchanged for all ensemble members throughout data assimilation process. The CEnKF method provides an appropriate solution under such condition.

Truncation remains the widely used approach for dealing with the non-physical updates to state variables in data assimilation using the EnKF. If the degree of constraint violations on the plausible values of state variables is not severe, truncating the state variables that violate constraints to the value at the boundary of the feasible region might be the most efficient solution for handling the inequality constraints. The transformation of variables approach can be an alternate solution for ensuring that the state variables honor the inequality constraints upon data assimilation. It is important, however, that the state variable, subject to inequality constraints, does not show

a strong correlation with other type of state variable. In some cases, as demonstrated with the compositional model, the application of constraints, instead of truncation or instead of transformation of variables, is critical. The CEnKF technique is suitable for data assimilation problems where large numbers of forecast state variables either lie on the boundary of the feasible region (e.g. phase saturations during waterflood) or are very close to it. Contrarily, the IPCEnKF method can be advantageous over the CEnKF technique for data assimilation problems where the forecast state variables are well within the interior of the feasible region. The only real alternative of assimilating data and simultaneously satisfying the inequality constraints is resolving to an iterative approach that requires running the reservoir simulator for each ensemble member during each iteration, which can be quite expensive if it has to be done very often.

# BIBLIOGRAPHY


Aanonsen, S. I., G. Nævdal, D. S. Oliver, A. C. Reynolds, and B. Vallès, Ensemble Kalman filter in reservoir engineering — a review, *SPE Journal*, **14**(3), 393–412, 2009.

Agbalaka, C. C. and D. S. Oliver, Application of the EnKF and localization to automatic history matching of facies distribution and production data, *Mathematical Geosciences*, **40**(4), 353–374, 2008.

Agbalaka, C. C. and D. S. Oliver, Automatic history matching of production and facies data with non-stationary proportions using EnKF, SPE 118916, in *Proceedings of the 2009 SPE Reservoir Simulation Symposium, The Woodlands, Texas*, 2009.

Anderson, J. L., An ensemble adjustment Kalman filter for data assimilation, *Monthly Weather Review*, **129**(12), 2884–2903, 2001.

Anderson, J. L., Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter, *Physica D: Nonlinear Phenomena*, **230**(1–2), 99–111, 2007.

Antoniou, A. and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*, Springer, New York, N.Y., 2007.

Baldick, R., *Applied Optimization: Formulation and Algorithms for Engineering Systems*, Cambridge University Press, Cambridge, UK, 2008.

Bell, B. M., J. V. Burke, and G. Pillonetto, An inequality constrained nonlinear Kalman-Bucy smoother by interior point likelihood maximization, *Automatica*, **45**(1), 25–33, 2009.

Bertino, L., G. Evensen, and H. Wackernagel, Sequential data assimilation techniques in oceanography, *International Statistical Review*, **71**(2), 223–241, 2003.

Bianco, A., A. Cominelli, L. Dovera, G. Nævdal, and B. Vallès, History matching and production forecast uncertainty by means of the ensemble Kalman filter: A real field application (SPE-107161), in *SPE Europec/EAGE Annual Conference and Exhibition. Society of Petroleum Engineers. London, UK*, 2007.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, Adaptive sampling with the ensemble transform Kalman filter. part I: Theoretical aspects, *Mon. Wea. Rev.*, **129**(3), 420–436, 2001.

Burgers, G., P. van Leeuwen, and G. Evensen, Analysis scheme in the ensemble Kalman filter, *Monthly Weather Review*, **126**(6), 1719–1724, 1998.

Chang, H., Y. Chen, and D. Zhang, Data assimilation of coupled fluid flow and geomechanics via ensemble Kalman filter (SPE 118963), *SPE Journal*, **accepted**, 2010.

Chen, Y., *Ensemble-based closed-loop production optimization*, Ph.D. dissertation, University of Oklahoma, Norman, Oklahoma, 2008.

Chen, Y. and D. S. Oliver, Cross-covariances and localization for EnKF in multiphase flow data assimilation, *Computational Geosciences*, **Online First**, 2009.

Chen, Y. and D. S. Oliver, Ensemble-based closed-loop optimization applied to Brugge Field (SPE 118926), *SPE Reservoir Evaluation & Engineering*, **13**(1), 56–71, 2010.

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

Chen, Y. and D. Zhang, Data assimilation for transient flow in geologic formations via ensemble Kalman filter, *Advances in Water Resources*, **29**(8), 1107–1122, 2006.

Cominelli, A., L. Dovera, S. Vimercati, and G. Nævdal, Benchmark study of ensemble Kalman filter methodology: history matching and uncertainty quantification for a deep-water oil reservoir (IPTC 13748), in *International Petroleum Technology Conference, 7–9 December 2009, Doha, Qatar*, 2009.

De Geeter, J., H. Van Brussel, and J. De Schutter, A smoothly constrained Kalman filter, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**, 1171–1177, 1997.

Dennis, J. E. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, 1983.

Deutsch, C. V. and A. G. Journel, *GSLIB: Geostatistical Software Library and User's Guide*, second edn., Oxford University Press, New York, 1998.

Ehrendorfer, M., A review of issues in ensemble-based Kalman filtering, *Meterologische Zeitschrift*, **16**(6), 795–818, 2007.

Evensen, G., Inverse methods and data assimilation in nonlinear ocean models, *Physica D*, **77**(1–3), 108–129, 1994a.

Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research*, **99**(C5), 10,143–10,162, 1994b.

Evensen, G., The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dynamics*, **53**, 343–367, 2003.

Evensen, G., Sampling strategies and square root analysis schemes for the EnKF, *Ocean Dynamics*, **54**(6), 539–560, 2004.

Evensen, G., *Data Assimilation: The Ensemble Kalman Filter*, Springer, 2006.

Evensen, G., J. Hove, H. C. Meisingset, E. Reiso, K. S. Seim, and Ø. Espelid, Using the EnKF for assisted history matching of a North Sea reservoir model (SPE 106184), in *Proceedings of the 2007 SPE Reservoir Simulation Symposium,*, 2007.

Evensen, G. and P. J. van Leeuwen, Assimilation of geosat altimeter data for the agulhas current using the ensemble Kalman filter with a quasigeostrophic model, *Monthly Weather Review*, **124**, 85–96, 1996.

Fiacco, A. V. and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley & Sons, New York, N.Y., 1968.

Floris, F. J. T., M. D. Bush, M. Cuypers, F. Roggero, and A.-R. Syversveen, Methods for quantifying the uncertainty of production forecasts: A comparative study, *Petroleum Geoscience*, **7**(SUPP), 87–96, 2001.

Frisch, K. R., The logarithmic potential method of convex programming, Tech. Rep. Technical Report, University Institute of Economics, Oslo, Norway, 1955.

Furrer, R. and T. Bengtsson, Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants, *J. Multivar. Anal.*, **98**(2), 227–255, 2007.

Gao, G., M. Zafari, and A. C. Reynolds, Quantifying uncertainty for the PUNQ-S3 problem in a Bayesian setting with RML and EnKF, *SPE Journal*, **11**(4), 506–515, 2006.

Gaspari, G. and S. E. Cohn, Construction of correlation functions in two and three dimensions, *Quarterly Journal of the Royal Meteorological Society*, **125**(554), 723–757, 1999.

Gerber, M. and F. Joos, Carbon sources and sinks from an ensemble Kalman filter ocean data assimilation, *Global Biogeochemical Cycles*, **24**, GB3004, 2010.

Gill, P. E., W. Murray, M. A. Saunders, J. Tomlin, and M. H. Wright, On projected Newton methods for linear programming and an equivalence to Karmarkar's projective method, *Mathematical Programming*, **36**, 183–209, 1986.

Gu, Y. and D. S. Oliver, History matching of the PUNQ-S3 reservoir model using the ensemble Kalman filter, *SPE Journal*, **10**(2), 51–65, 2005.

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

Gu, Y. and D. S. Oliver, An iterative ensemble Kalman filter for multiphase fluid flow data assimilation, *SPE Journal*, **12**(4), 438–446, 2007.

Gupta, N. and R. Hauser, Kalman filtering with equality and inequality state constraints, Tech. Rep. Report No. 07/18, Numerical Analysis Group, Oxford Univ. Computing Laboratory, Univ. Oxford, Oxford, UK, 2007.

Hacker, J. P., J. L. Anderson, and M. Pagowski, Improved vertical covariance estimates for ensemble-filter assimilation of near-surface observations, *Monthly Weather Review*, **135**(3), 1021–1036, 2007.

Hamill, T. M., C. Snyder, D. P. Baumhefner, Z. Toth, and S. L. Mullen, Ensemble forecasting in the short to medium range: Report from a workshop, *Bull. Amer. Meteor. Soc.*, **81**, 2653–2664, 2000.

Hamill, T. M., J. S. Whitaker, and C. Snyder, Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Monthly Weather Review*, **129**(11), 2776–2790, 2001.

Hansen, J. A., Accounting for model error in ensemble-based state estimation and forecasting, *Monthly Weather Review*, **130**(10), 2373–2391, 2002.

Haugen, V., G. Naevdal, L.-J. Natvik, G. Evensen, A. M. Berg, and K. M. Flornes, History matching using the ensemble Kalman filter on a North Sea field case, *SPE Journal*, **13**(4), 382–391, 2008.

Hayward, S., Constrained Kalman filter for least-squares estimation of time-varying beamforming weights, *Mathematics in Signal Processing IV*, pp. 113–125, 1998.

Houtekamer, P. L. and H. L. Mitchell, Data assimilation using an ensemble Kalman filter technique, *Monthly Weather Review*, **126**(3), 796–811, 1998.

Houtekamer, P. L. and H. L. Mitchell, A sequential ensemble Kalman filter for atmospheric data assimilation, *Monthly Weather Review*, **129**(1), 123–137, 2001.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations, *Monthly Weather Review*, **133**(3), 604–620, 2005.

Irving, A. and E. Robert, Optimisation of uncertain structural parameters using production and observation well data, in *SPE EUROPEC/EAGE Annual Conference and Exhibition, 14–17 June 2010, Barcelona, Spain*, 2010.

Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

Kalman, R. E., A new approach to linear filtering and prediction problems, *Transactions of the ASME, Journal of Basic Engineering*, **82**, 35–45, 1960.

Karmarkar, N., A new polynomial-time algorithm for linear programming, *Combinatorics*, **4**, 373–395, 1984.

Keppenne, C. L. and M. M. Rienecker, Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model, *Monthly Weather Review*, **130**(12), 2951–2965, 2002.

Ko, S. and R. R. Bitmead, State estimation for linear systems with state equality constraints, *Automatica*, **43**(8), 1363–1368, 2007.

Krymskaya, M. V., R. G. Hanea, and M. Verlaan, An iterative ensemble Kalman filter for reservoir engineering applications, *Computational Geosciences*, **13**(2), 235–244, 2009.

Lauvernet, C., J.-M. Brankart, F. Castruccio, G. Broquet, P. Brasseur, and J. Verron, A truncated gaussian filter for data assimilation with inequality constraints: Application to the hydrostatic stability condition in ocean models, *Ocean Modelling*, **27**(1-2), 1–17, 2009.

Li, G. and A. C. Reynolds, An iterative ensemble Kalman filter for data assimilation, 2009a.

Li, G. and A. C. Reynolds, Iterative ensemble Kalman filters for data assimilation, *SPE Journal*, **14**(3), 496–505, 2009b.

Li, R., A. C. Reynolds, and D. S. Oliver, History matching of three-phase flow production data, *SPE Journal*, **8**(4), 328–340, 2003.

Liu, N. and D. S. Oliver, Automatic history matching of geologic facies, *SPE Journal*, **9**(4), 188–195, 2004.

Liu, N. and D. S. Oliver, Critical evaluation of the ensemble Kalman filter on history matching of geologic facies, *SPE Reservoir Evaluation & Engineering*, **8**(6), 470–477, 2005a.

Liu, N. and D. S. Oliver, Ensemble Kalman filter for automatic history matching of geologic facies, *Journal of Petroleum Science and Engineering*, **47**(3–4), 147–161, 2005b.

Lødøen, O. P. and H. Omre, Scale-corrected ensemble Kalman filtering applied to production history conditioning in reservoir evaluation, *SPE Journal*, **13**(2), 177–194, 2008.

Lorenc, A. C., The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var, *Quarterly Journal of the Royal Meteorological Society*, **129**(595), 3183–3203, 2003.

Lorentzen, R. J., K. K. Fjelde, J. Frøyen, A. C. V. M. Lage, G. Nævdal, and E. H. Vefring, Underbalanced and low-head drilling operations: Real time interpretation of measured data and operational support (SPE–71384), in *SPE Annual Technical Conference and Exhibition, 30 September–3 October 2001, New Orleans, LA*, 2001a.

Lorentzen, R. J., K. K. Fjelde, J. Frøyen, A. C. V. M. Lage, G. Nævdal, and E. H. Vefring, Underbalanced drilling: Real time data interpretation and decision support, in *SPE/IADC Drilling Conference*, 2001b.

Lorentzen, R. J., G. Nævdal, B. Vàlles, A. M. Berg, and A.-A. Grimstad, Analysis of the ensemble Kalman filter for estimation of permeability and porosity in reservoir models, SPE–96375, in *SPE Annual Technical Conference and Exhibition*, 2005.

Mandel, J., M. Chen, L. P. Franca, C. Johns, A. Puhalskii, J. L. Coen, C. C. Douglas, R. Kremens, A. Vodacek, and W. Zhao, A note on dynamic data driven wildfire modeling, *Lecture Notes in Computer Science*, pp. 725–731, 2004.

Massicotte, D., R. Z. Morawski, and A. Barwicz, Incorporation of a positivity constraint into a Kalman filter-based algorithm for correction of spectrometric data, *IEEE Transactions on Instrumentation and Measurement*, **44**(1), 2–7, 1995.

Megiddo, N., Pathways to the optimal set in linear programming, Technical Report RJ 5295, IBM Almaden Research Center, San Jose, CA, 1986.

Mitchell, H. L. and P. L. Houtekamer, An adaptive ensemble Kalman filter, *Monthly Weather Review*, **128**(2), 416–433, 2000.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, Ensemble size, balance, and model-error representation in an ensemble Kalman filter, *Monthly Weather Review*, **130**(11), 2791–2808, 2002.

Moradkhani, H., S. Sorooshian, H. V. Gupta, and P. R. Mouser, Dual state-parameter estimation of hydrological models using ensemble Kalman filter, *Advances in water resources*, **28**(2), 301–305, 2005.

Nævdal, G., L. M. Johnsen, S. I. Aanonsen, and E. H. Vefring, Reservoir monitoring and continuous model updating using ensemble Kalman filter, *SPE Journal*, **10**(1), 66–74, 2005.

Nævdal, G., T. Mannseth, and E. H. Vefring, Near-well reservoir monitoring through ensemble Kalman filter: SPE 75235, in *Proceedings of SPE/DOE Improved Oil Recovery Symposium*, 2002.

Nocedal, J. and S. J. Wright, *Numerical Optimization*, second edn., Springer, 2006.

Oliver, D. S., A. C. Reynolds, and N. Liu, *Inverse Theory for Petroleum Reservoir Characterization and History Matching*, first edn., Cambridge University Press, Cambridge, 2008.

Oliver, D. S., Y. Zhang, H. A. Phale, and Y. Chen, Distributed parameters and state estimation in petroleum reservoirs, *Computers & Fluids*, **submitted**, 2010.

Pan, M. and E. F. Wood, Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter, *Journal of Hydrometeorology*, **7**, 534–547, 2006.

Peters, E., R. J. Arts, G. K. Brouwer, S. Cullick, R. J. Lorentzen, Y. Chen, K. N. B. Dunlop, F. C. Vossepoel, R. Xu, P. Sarma, A. H. Alhutali, and A. C. Reynolds, Results of the Brugge benchmark study for flooding optimization and history matching, *SPE Reservoir Evaluation & Engineering*, pp. 391–405, 2010.

Pham, D. T., Stochastic methods for sequential data assimilation in strongly nonlinear systems, *Monthly Weather Review*, **129**(5), 1194–1207, 2001.

Porrill, J., Optimal combination and constraints for geometrical sensor data, *International Journal of Robotics Research*, **7**(6), 66–77, 1988.

Quintana, V. H., G. L. Torres, and J. Medina-Palomo, Interior-point methods and their applications to power systems: A classification of publications and software codes, *IEEE Transactions on Power Systems*, **15**(1), 170–176, 2000.

Qureshi, A. G., M. P. Res, and B. C. Burnaby, Constrained Kalman filtering for image restoration, in *Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on*, pp. 1405–1408, 1989.

Rao, C. V., J. B. Rawlings, and D. Q. Mayne, Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations, *IEEE Transactions on Automatic Control*, **48**(2), 246–258, 2003.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, Hydrologic data assimilation with the ensemble Kalman filter, *Monthly Weather Review*, **130**(1), 103–114, 2002.

Rotea, M. and C. Lana, State estimation with probability constraints, *International Journal of Control*, **81**(6), 920–930, 2008.

Schlumberger, Eclipse technical and reference manual, 2006.

Seiler, A., G. Evensen, J. A. Skjervheim, J. Hove, and J. G. Vabo, Advanced reservoir management workflow using an EnKF based assisted history matching method, (SPE 118906), in *SPE Reservoir Simulation Symposium, 2–4 February, The Woodlands, Texas, USA*, 2009.

Simon, D. and T. L. Chia, Kalman filtering with state equality constraints, *IEEE Transactions on Aerospace and Electronic Systems*, **38**(1), 128–136, 2002.

Simon, E. and L. Bertino, Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment, *Ocean Science Discussions*, **6**(1), 617–652, 2009.

Skjervheim, J.-A., G. Evensen, S. I. Aanonsen, B. O. Ruud, and T. A. Johansen, Incorporating 4D seismic data in reservoir simulation models using ensemble Kalman filter, *SPE Journal*, **12**(3), 282–292, 2007.

Tarantola, A., *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, Amsterdam, The Netherlands, 1987.

Thacker, W. C., Data assimilation with inequality constraints, *Ocean Modelling*, **16**(3-4), 264–276, 2007.

Thacker, W. C. and R. B. Long, Fitting dynamics to data, *Journal of Geophysical Research*, **93**(1), 1227–1240, 1988.

Thulin, K., G. Li, S. I. Aanonsen, and A. C. Reynolds, Estimation of initial fluid contacts by assimilation of production data with EnKF, in *Proceedings of the 2007 SPE Annual Technical Conference and Exhibition*, 2007.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, Ensemble square-root filters, *Monthly Weather Review*, **131**, 1485–1490, 2003.

Ungarala, S., E. Dolence, and K. Li, Constrained extended Kalman filter for nonlinear state estimation, in *Proccedings of the 8th International IFAC Symposium on Dynamics and Control of Process Systems*, 2007.

Wang, Y., G. Li, and A. C. Reynolds, Estimation of depths of fluid contacts by history matching using iterative ensemble Kalman smoothers (SPE 119056), *SPE Journal*, **accepted**, 2010.

Whitaker, J. S. and T. M. Hamill, Ensemble data assimilation without perturbed observations, *Monthly Weather Review*, **130**(7), 1913–1924, 2002.

Wu, Z., A. C. Reynolds, and D. S. Oliver, Conditioning geostatistical models to two-phase production data, *SPE Journal*, **4**(2), 142–155, 1999.

Zafari, M., G. Li, and A. C. Reynolds, Iterative forms of the ensemble Kalman filter, in *Proceedings of the 10th European Conference on the Mathematics of Oil Recovery — Amsterdam*, p. A030, 2006.

Zhang, D., Z. Lu, and Y. Chen, Dynamic reservoir data assimilation with an efficient, dimension-reduced Kalman filter, *SPE Journal*, **12**(1), 108–117, 2007.

Zhang, Y. and D. S. Oliver, Evaluation and error analysis: Kalman gain regularization versus covariance regularization, *Computational Geosciences*, **submitted**, 2010a.

Zhang, Y. and D. S. Oliver, History matching using the ensemble Kalman filter with a multiscale parameterization: a field case study (SPE 118879), *SPE Reservoir Evaluation & Engineering*, **minor revision**, 2010b.

Zhang, Y. and D. S. Oliver, Improving the ensemble estimate of the Kalman gain by bootstrap sampling, *Mathematical Geosciences*, **accepted**, 2010c.

Zhao, Y., A. C. Reynolds, and G. Li, Generating facies maps by assimilating production data and seismic data with the ensemble Kalman filter, SPE-113990, in *Proceedings of the 2008 SPE Improved Oil Recovery Symposium, Tulsa, Oklahoma, April 21–23*, 2008.