Oil Reservoir Characterization Using Ensemble Data Assimilation

by

Behnam Jafarpour

B.Sc. Civil Engineering
The University of Tehran, 1999

M.Sc. Civil and Environmental Engineering
The University of Delaware, 2003

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2008

Signature of Author .....
Department of Civil and Environmental Engineering
October 5th, 2007

Certified by .....
Dennis McLaughlin
H.M. King Bhumibol Professor of Civil and Environmental Engineering
Thesis Supervisor

Certified by .....
Ruben Juanes
Assistant Professor of Civil and Environmental Engineering
Thesis Committee Chair

Accepted by .....
Daniele Veneziano
Chairman, Departmental Committee for Graduate Students

Oil Reservoir Characterization using Ensemble Data Assimilation

by

Behnam Jafarpour

Submitted to the Department of Civil and Environmental Engineering
on October 5th, 2007 in partial fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
the field of Civil and Environmental Engineering

Abstract

Increasing world energy demand combined with decreasing discoveries of new and accessible hydrocarbon reserves are necessitating optimal recovery from the world's current hydrocarbon resources. Advances in drilling and monitoring technologies have introduced intelligent oilfields that provide real-time measurements of reservoir conditions. These measurements can be used for more frequent reservoir model calibration and characterization that can lead to improved oil recovery though model-based closed-loop control and management.

This thesis proposes an efficient method for probabilistic characterization of reservoir states and properties. The proposed algorithm uses an ensemble data assimilation approach to provide stochastic characterization of reservoir attributes by conditioning individual prior ensemble members on dynamic production observations at wells. The conditioning is based on the second-order Kalman filter analysis and is performed recursively, which is suitable for real-time control applications. The prior sample mean and covariance are derived from nonlinear dynamic propagation of an initial ensemble of reservoir properties. Realistic generation of these initial reservoir properties is shown to be critical for successful performance of the filter. When properly designed and implemented, recursive ensemble filtering is concluded to be a practical and attractive alternative to classical iterative history matching algorithms.

A reduced representation of reservoir's states and parameters using discrete cosine transform is presented to improve the estimation problem and geological consistency of the results. The discrete cosine transform allows for efficient, flexible, and robust parameterization of reservoir properties and can be used to eliminate redundancy in reservoir description while preserving important geological features. This improves under-constrained inverse problems such as reservoir history matching in which the number of unknowns significantly exceeds available data. The proposed parameterization approach is general and can be applied with any inversion algorithm.

The suitability of the proposed estimation framework for hydrocarbon reservoir characterization is demonstrated through several waterflooding examples using synthetic reservoir models.

Thesis Supervisor: Dennis McLaughlin
Title: H.M. King Bhumibol Professor of Civil and Environmental Engineering,

Acknowledgements

I could not have even dreamed of being where I am without the help and support of many people from my past and present. Unfortunately, it is not possible to name all of these great people in here. However, this thesis will be incomplete without expressing my gratitude to those who have had direct and significant contribution to my PhD work.

I would like to thank my thesis supervisor, Professor Dennis McLaughlin, for his valuable advice and support during the past four years. I have greatly enjoyed working with him and learning from him. I wish to extend my sincere gratitude to my thesis committee members who have provided important input and guidance in carrying out this research. In particular, I am grateful to Professor Jan-Drik Jansen from Delft University of Technology for his kind effort to attend my thesis committee meetings despite the long distance. I would like to thank Professor Ruben Juanes and Professor Charles Harvey for their help and contributions to this thesis. I am indebted to Professor Louis Durlofsky from Stanford University for hosting me as a visiting student during my stay at Stanford and also his contributions to my research. I am deeply grateful to Professor Dara Entekhabi for being a great mentor and for providing invaluable professional advice.

I wish to thank Shell International Exploration and Production for sponsoring this project. I also would like to acknowledge Schlumberger for donating ECLIPSE reservoir simulator. I appreciate the assistance from the staff at Parsons Lab and CEE department, in particular, James Long, Shiela Frankel, Cynthia Stewart, Kris Kipp, and Jeanette Marchocki.

I have benefited from and greatly enjoyed my technical and social interactions with my fellow environmental data assimilation research group members. Specifically, I would like to thank Crystal Ng for proof-reading my writings. I also wish to thank my good friend Adel Ahanin who has been very supportive during some of my most difficult times. I have been privileged to have great friends who made my past four years in Boston a memorable and precious experience. I am thankful to all of them. I have been fortunate to meet my kind and supportive girlfriend, Maryam Modir-Shanechi, during my years at MIT. I would like to thank her for sharing her life with me in the past two years and for the beautiful moments we have spent together.

Above all, I am deeply grateful to my family for their sacrifice and patience; I share my accomplishments with all of them. My greatest respect and most heart-felt gratitude go to my mother, Effat Gholizadeh Zargar, who has been a tower of strength, hope, and selflessness. I would like to thank my father, Ahmad, for his encouragements and support. I wish to thank my brothers Aref, Parviz (whom I lost in the past summer and whose memories I will keep alive and cherish forever), Behrouz, and my sister Fatemeh who have all had significant contributions to my education and life in general. Today would have only been a dream if it was not for the dedication, support, friendship, and encouragement of my wonderful brother Behrouz. I will always be indebted to him.

Finally, I dedicate this thesis, as a token of appreciation, to my family for their long-time support and dedication and to the fond memories of my brother Parviz.

Contents

3. Permeability Parameterization with the Discrete Cosine Transform

4. History Matching with an Ensmble Kalman Filter and Discrete Cosine Parameterization

5. History Matching of the Three-Dimensional SPE-10 Reservoir Model using EnKF and DCT

6. Conclusions and Future Research

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Reservoir Characterization and Management

1.1.1 Introduction and Significance

The twenty first century is likely to witness a shift in world energy supply from heavily hydrocarbon-based to more sustainable and environmentally benign sources of energy. During this transition phase, the ever-increasing energy demand calls for efficient exploitation of our shrinking fossil fuel reserves until practical and reliable alternative sources are discovered. The need for optimized oil recovery is even more pressing considering descending trends in reserve discovery as well as challenges in accessing deeper oilfields [1.1].

Discovery of a new reservoir is followed by substantial effort to acquire information about its properties and characteristics that can affect its exploitation before development

strategies can be planned. In general, reservoir characterization refers to integration of all data sources (core, borehole, log, 3D seismic, production data, and 4D seismic) and modeling approaches (geological, geostatistical, forward flow simulation, inverse modeling, and uncertainty quantification) to provide the best possible description of a reservoir and its observed response to past development scenarios. Inherently, reservoir characterization crosses several disciplines, including geology and geophysics, inverse modeling, estimation, and optimization theories.

Recent advances in exploration and production technology such as intelligent well drilling and time-lapsed seismic surveys for continuous reservoir monitoring are providing extensive reservoir monitoring and control opportunities that can be used to improve reservoir management and development strategies [1.2-1.4]. Enhanced data acquisition systems are being deployed in oildfields to collect measurements of an array of variables at different scales, resolutions, and coverage. With the advent of the new drilling technologies and sensing capabilities in reservoir engineering comes a need for adapted scientific methodologies that can take advantage of the large quantity of the emerging monitoring data. Recently, researchers have strived to address this need by proposing automated reservoir characterization, i.e. *history matching* (many references provided throughout this thesis), *closed-loop control* [1.3, 1.5, 1.6], and integrated reservoir management approaches [1.7].

In the past decade, geostatistical reservoir modeling [1.8-1.11] and numerical flow simulation [1.12, 1.13] have benefited from unprecedented computational capabilities that paved the foundation for more ambitious and sophisticated reservoir flow modeling, production forecast, and uncertainty quantification. Accurate reservoir description and modeling, combined with the new drilling technology and data acquisition networks, can lead to better understanding of reservoir behavior and informed reservoir development planning which is hoped to improve hydrocarbon recovery.

To facilitate real time reservoir control and management, reservoir models should be tuned as soon as new measurements are acquired from installed sensors. This calls for innovative and efficient reservoir characterization algorithms to address real time control and management's data integration demands. Desirable features of such algorithms are robustness, efficiency, accuracy, uncertainty quantification, geological consistency, and real-time recursive data integration capabilities.

An important aspect of reservoir characterization is related to effective description of the reservoir model. In numerical reservoir simulation grid block description of the reservoir model is used to derive the dynamical states of the reservoir in space and time. Such grid-oriented description is usually inefficient for reservoir characterization (estimating reservoir properties from observed measurements) as geological continuity results in several neighboring grid blocks having identical properties. Therefore, a concise

parametric description of reservoir properties that reduces the number of independent unknowns (while maintaining the major features) is desirable. This process is commonly referred to as *parameterization*. Parameterization, when applied properly, can improve the computational cost of reservoir characterization and preserve geological continuity between adjacent locations.

This thesis presents an efficient estimation framework that is applicable to real time reservoir control and management. The proposed framework takes advantage of a new efficient parameterization algorithm for improved reservoir characterization. The main focus is on integration of dynamic production data into prior reservoir models using ensemble-based data assimilation techniques.

1.2 Reservoir Simulation and Mathematical Modeling

Reservoir engineers use a mathematical model to represent, analyze, and quantify a reservoir's flow displacement patterns and production behavior. Predictions of future performance can be used to prepare sound reservoir development and management strategies. Therefore, reservoir modeling plays a central role in planning future production scenarios and operational activities. In this section, a brief overview of the governing equations in reservoir simulation and multiphase flow modeling is presented. These equations will be referred to as the dynamic forward model in describing the data assimilation algorithm in later chapters.

1.2.1 Multiphase Flow Equations

The general form of the governing equations for the two phase (oil-water) immiscible flow in porous media is derived from mass and momentum (Darcy's Law) conservation principles as [1.12, 1.14]:

$$\nabla.\left(\frac{\lambda_o}{B_o}\mathbf{k}(\nabla P_o - \gamma_o \nabla Z)\right) = \frac{\partial}{\partial t}(\phi \frac{S_o}{B_o}) + q_o \qquad (1.1)$$

$$\nabla.\left(\frac{\lambda_w}{B_w}\mathbf{k}(\nabla P_w - \gamma_w \nabla Z)\right) = \frac{\partial}{\partial t}(\phi \frac{S_w}{B_w}) + q_w \qquad (1.2)$$

Here, λ_o and λ_w represent mobility of oil and water (measure of the ease with which a fluid can be moved in a specific medium), B_o and B_w are the formation volume factors (volume of fluid as function of pressure to its volume at standard atmospheric pressure), **k** is intrinsic permeability (a physical property of rocks that indicate their conductivity), P_o and P_w are oil and water pressures, γ_o and γ_w represent oil and water densities, Z is elevation, S_o and S_w are oil and water saturation (ratio of oil (water) volume to total volume of pore space in rocks), and finally q_o and q_w are sink and source (control) terms referring to injection and production rates per unit volume.

These two equations have four unknown states, namely P_o, P_w, S_o, S_w. For a given set of model input parameters, boundary conditions, initial conditions, and reservoir controls

(well rates/bottom-hole pressures) two additional (*constitutive*) equations are used to find a *unique* solution for these unknowns at any given time. The constitutive equations reflect capillary pressure (pressure difference between the two phases) relations and physical saturation constraint for a given control volume:

$$\begin{aligned} P_o - P_w &= P_c(S_w) \\ S_w + S_o &= 1 \end{aligned} \tag{1.3}$$

Forward integration of these equations provides model solutions in time that are used to predict flow behavior within the reservoir. A closed form solution to the above equations can only be achieved under rare circumstances when highly simplifying assumptions are used. In general, these equations need to be discretized (often through finite difference/element/volume approach or a combination of them) and solved to obtain reservoir states (saturation and pressure distributions) and production forecasts at desired times. However, forward modeling is only useful when accurate model inputs are available, and flow equations closely capture all the important physical processes present in the actual reservoir.

In practice, several sources of error exist that make the solution of the above system of equations uncertain. Among these are: inaccurate specification of initial and boundary conditions, model input parameters (such as relative and intrinsic permeabilities and porosity), fluid thermodynamic properties, as well as representativeness errors relating to boundary and source and sink (wells) locations. In addition, mathematical models often

include simplifying assumptions in explaining the underlying physical processes. It is, therefore, a common practice to calibrate reservoir models by incorporating observed past reservoir performance through inverse modeling (history matching). This is discussed in the following section.

While the estimation approach in this thesis uses the solution of the above equations, the forward model solution is not the focus of this research and, therefore, not discussed further (interested readers are referred to citations given above). A commercial reservoir simulator, ECLIPSE [1.13], is used to solve the coupled multiphase equations of (1.1-1.3).

1.3 Reservoir Inverse Modeling (History Matching)

1.3.1 Inverse Modeling Motivation

While the forward model is a powerful tool for predicting a reservoir's future performance, its utility depends strongly on the accuracy of input reservoir properties and specified model parameters. In practice, reservoir properties are only partially known, due to the limited accessibility and cost constraints associated with data acquisition. As a result, model calibration procedures must be performed to estimate poorly known reservoir properties. The process of tuning uncertain reservoir properties to produce observed production behavior is known as inverse modeling, also referred to as parameter

estimation or system identification in other contexts. In reservoir engineering literature, this process is named *history matching*.

History matching commonly refers to the estimation of reservoir model parameters such as permeability and porosity using historical observations of dynamic production and pressure data [1.15]. Specifically, direct or indirect observations of the state variables are used to adjust model parameters so that they produce model outputs that match observed measurements. In general, history matching of heterogeneous hydrocarbon reservoirs is an inherently underdetermined problem where the number of unknown parameters is much larger than the number of sparse available data. As a result, the solution to the estimation problem is known to be non-unique [1.15]. This means that several solutions can be found, with different geological characteristics, to match the observed data equally well.

1.3.2 Background on Inverse Modeling

This section presents a brief review of history matching approaches and their development in reservoir engineering literature.

1.3.2.1 Traditional (Manual) Reservoir History Matching

Traditionally, history matching has been done by manually adjusting sensitive reservoir model parameters until a reasonable data fit is obtained. This approach is still frequently used in practice and has its own merits when carried out by experienced reservoir engineers. In principle, since the history matching problem is naturally ill-posed, sound engineering judgment can play an important role in eliminating unrealistic solutions and save unnecessary computational effort. On the other hand, the number of unknown variables is typically so overwhelming that it is almost impossible to perform the history matching without resorting to computational algorithms.

1.3.2.2 Automatic Reservoir History Matching

In the past three decades the advances in computational technology has made automatic history matching a viable and powerful alternative to manual history matching. A commonly used form of automatic history matching problem minimizes an objective function that includes penalty terms for deviations from observed values, as well as a term for departure from the prior knowledge about parameters:

$$\arg\min_{\mathbf{m}} \mathbf{J}_1(\mathbf{m}) = (\mathbf{h}(\mathbf{m}) - \mathbf{d}_{obs})^{\mathrm{T}} \mathbf{C}_d^{-1} (\mathbf{h}(\mathbf{m}) - \mathbf{d}_{obs}) + (\mathbf{m} - \overline{\mathbf{m}})^{\mathrm{T}} \mathbf{C}_m^{-1} (\mathbf{m} - \overline{\mathbf{m}}) \qquad (1.4)$$

Where, $\mathbf{m}$ and $\overline{\mathbf{m}}$ are vectors of unknowns and prior parameter mean, respectively, with prior covariance specified as $\mathbf{C}_m$; $\mathbf{J(m)}$ is the minimization objective function; $\boldsymbol{h(m)}$ is the measurement operator and $\mathbf{d}_{obs}$ is the uncertain observation vector with covariance $\mathbf{C}_d$.

With linear model and Gaussian statistical assumptions [1.16-1.18], this objective function can be derived from Baysian probabilistic framework. The solution to this objective function, with the specified assumptions, yields the peak of the Gaussian *a posteriori* distribution. With these simplifying assumptions, the posterior mean, mode, and median overlap and the linear least square estimate (LLSE), the Best linear unbiased estimate (BLUE), and the maximum a posteriori (MAP) estimates all give the same solution. However, in general these estimators provide different point statistics of the *a posteriori* distribution.

A more general form of the history matching minimization problem includes other constraint terms such as the dynamic forward model, initial conditions, geological constraints, as well as inequality and boundedness constraints [1.18, 1.19]. Therefore, in a more general form, the history matching minimization problem can be formulated as:

$$\arg\min_{\mathbf{m}} \; \mathbf{J}_2(\mathbf{m}) = \tag{1.5}$$

$$(h(\mathbf{m}) - \mathbf{d}_{obs})^{\mathrm{T}} \mathbf{C}_d^{-1} (h(\mathbf{m}) - \mathbf{d}_{obs}) +$$
$$(\mathbf{m} - \overline{\mathbf{m}})^{\mathrm{T}} \mathbf{C}_m^{-1} (\mathbf{m} - \overline{\mathbf{m}}) + (\mathbf{x}_0 - \overline{\mathbf{x}}_0)^{\mathrm{T}} \mathbf{C}_{x0}^{-1} (\mathbf{x}_0 - \overline{\mathbf{x}}_0) +$$
$$\mathbf{g}^n(\mathbf{x}^{n+1}, \mathbf{x}^n, \mathbf{u}^n, \mathbf{m}, \mathbf{q})^T \mathbf{C}_q^{-1} \mathbf{g}^n(\mathbf{x}^{n+1}, \mathbf{x}^n, \mathbf{u}^n, \mathbf{m}, \mathbf{q})$$

s.t.

$\mathbf{m}_{\text{low}} < \mathbf{m} < \mathbf{m}_{\text{up}}$

$\mathbf{m} \in$ geologically consistent realization

Here, $\bar{\mathbf{x}}_0$, $\mathbf{x}_0$, and $\mathbf{C}_{x0}$ are the uncertain initial states, their corresponding prior mean and covariance, respectively; $\mathbf{g}$ defines the reservoir forward model that uses specified control variables $\mathbf{u}$ and modeling error $\mathbf{q}$ with covariance $\mathbf{C}_q$.

This problem formulation approach assumes that the forward model and initial conditions are not perfect (*soft* constraints). In petroleum engineering literature, however, it has been common to assume error-free forward model integration and perfect knowledge of initial conditions. Therefore, the corresponding terms in the objective functions can be dropped and used as *hard* equality constraints (this is equivalent to setting the corresponding covariances to zero).

In the past three decades, automatic history matching research has focused on solutions to some form of the above minimization problem [1.20-1.22]. Standard and modified gradient-based search algorithms such as Steepest Descent, Conjugate Gradient, Gauss-Newton, Newton, and Quasi-Newton methods have been applied to the above minimization problem. A brief review and comparison between these methods can be found in [1.21] and references therein. All of these methods require gradients of a specified objective function (usually in terms of well dynamic variables) with respect to unknown parameters. Numerical computation of these gradients for problems with large number of parameters is best done by solving a backward adjoint model (using variational calculus) [1.18, 1.22]. One major drawback of the adjoint method is the

complexity associated with the construction of the adjoint model (modification of the simulator source code), which presents a challenge for its incorporation in commercial reservoir simulators.

Non-gradient random search methods [1.23, 1.24] such as simulated annealing [1.25-1.27], and genetic algorithms [1.27, 1.28], have also been studied for minimizing the stated objective functions. While these heuristic global search algorithms are known to search the entire error space to find the absolute minimum, their computational cost increases exponentially with increasing size of unknown parameters. Hybrid search algorithms [1.29] can be devised to combine the merits of global and local search algorithms; that is, to avoid local minima while effectively reducing the objective function and converging to the absolute minimum independently of the starting point. Other techniques such as neural networks [1.30], gradual deformation [1.31, 1.32], probability perturbation [1.33], and randomized maximum likelihood [1.20, 1.34] methods have also been used, with varying degree of success, to solve the reservoir history matching problem and quantifying the associated uncertainty.

Regardless of the minimization method used, the history matching problem can yield multiple solutions (good fit to data) due to the underdetermined nature of the problem. Therefore, the dimensionality of the parameter space needs to be reduced using parameterization [1.15]. Aside from improving the illposedness of the history matching problem, parameterization can reduce the computational cost and impart geological

realism to the solution. This topic is further introduced in the next section and discussed in more detail in Chapter 3.

Another important and challenging aspect of the history matching problem is quantification of forecast uncertainty, which is often under-estimated [1.22]. Commonly, uncertainty analysis in reservoir forecasting involves simplifying and often unrealistic assumptions, including perfect knowledge of specific model parameters (e.g. relative permeability curves, porosity, reservoir geometry, initial reservoir conditions). A joint comparison study on a benchmark reservoir history matching and forecasting problem was conducted by several researchers to quantify production forecast uncertainty using several methods [1.35]. The results indicated that in a significant number of approaches, the reported ranges of production did not include true conditions [1.35].

In uncertainty characterization, Gaussian stochastic assumptions are often applied to petrophysical properties. This practice is motivated more by simplicity of mathematical treatment than geological realism. Because Gaussian spatial fields provide only second order characterization, they often fail to capture certain connected features (such as curvilinear channels) that are characteristic of oil reservoirs [1.22]. Training-image-based facies modeling approaches have recently been introduced by *Strebelle et al.* [1.11]. These use a training image (conceptual model of facies distribution) to generate pattern-based facies realizations. This approach appears to be more concept-oriented and the

simulated fields are easier to interpret geologically [1.11]. However, the method assumes availability of a reliable training image, which may be difficult to obtain in practice.

Reservoir characterization is becoming a data-rich field, where profusion of measurements from different sources and of different types and quality calls for more efficient, and effective estimation algorithms. A suitable data integration method to meet the aforementioned challenges should have the following properties: i) computational robustness and efficiency; ii) accuracy within measurement errors; iii) consistency with underlying geological model; iv) uncertainty quantification; v) integration of various data types (scale, accuracy); vi) model error and measurement noise incorporation; vii) easy implementation; and viii) recursive conditioning.

Ensemble methods provide a means to systematically include uncertainty in reservoir description and characterization. As a result, a more realistic (less biased) evaluation of a reservoirs properties and future performance is achieved. In ensemble data assimilation, states and parameters are estimated simultaneously. Consequently, the relationship between states and parameters is not constrained by forward model equations. This is the result of assuming imperfect forward model in this approach.

Sequential ensemble-based data assimilation techniques appear to offer several of the aforementioned capabilities. Hence, they are currently receiving considerable attention from researchers in the field. An overview of the literature in this area is presented next.

1.3.1.3 History Matching using Sequential Data Assimilation

Data assimilation aims at accurate re-analysis, estimation, and prediction of an unknown state by merging observed information with a model. It is mainly applied to areas that enjoy an abundance of data, such as oceanography and meteorology. Ensemble data assimilation techniques can be developed independently of the dynamic models and therefore are easily incorporated into the estimation algorithm. The underlying Markov model assumption in ensemble techniques yields recursive estimation algorithms that lend themselves to real time (filtering) applications. Hence, they only integrate the most recent data into the model without the need to reprocess past observations. Furthermore, ensemble data assimilation algorithms can accommodate model errors as well as uncertainty in observations. These properties of ensemble data assimilation algorithms make them a suitable choice for online data integration and real-time forecasting applications.

Data assimilation techniques can be classified into two main groups: *variational (adjoint-based) iterative methods* and *sequential filtering approaches*. In general, these two methods make different assumptions and have different implementations. However, both

of them can be formulated using the Bayesian framework [1.16-1.18] and can be shown to yield identical solutions in the case of linear state-space models and Gaussian statistics characterizing the uncertainties. Variational data assimilation has been used for reservoir history matching in the form of the problem statement in equation (1.5).

Alternatively, recursive filtering methods were developed for optimal state estimation with linear dynamical models. The *Kalman Filter* [1.36] is the most common filtering technique for linear Gaussian models. In essence, the Kalman filter state estimate is a weighted linear combination of the background (forecast) state and observations, where the weights depend on the uncertainties in model predictions and observations. For smaller observation errors (relative to the prediction errors) the analysis states are drawn closer toward the observation whereas for very uncertain observations model predictions are weighted more. Major limitations of Kalman filter are the linear dynamic assumption, the Gaussianity of the error statistics, and the computational cost of covariance propagation. The *Extended Kalman Filter* (EKF) is designed to address the nonlinearity issue, however it adds to the computational cost by introducing a linearization step of the forward model [1.37]. Furthermore, it does not provide a practical solution for highly nonlinear problems. Ensemble type filters were developed to remove linear error propagation constraint and avoid the computational cost associated with covariance propagation.

Evensen [1.38] proposed an ensemble version of the Kalman filter, the *Ensemble Kalman Filter* (EnKF), which can be used with any nonlinear state-space model. After its

introduction, several efforts have been made to improve the original EnKF algorithm. In a recent paper [1.39], Evensen provides a review of EnKF literature and its evolution, as well as an efficient implementation scheme using singular value decomposition. The EnKF has gained popularity in several large scale applications such as oceanography, metrology, and hydrology [1.40, 1.41]. However, its application to reservoir history matching has only recently been considered.

In the first use of EnKF in reservoir modeling, *Nævdal et al.* [1.42] updated near-well reservoir models by adjusting the permeability field. They found that early measurements were more important in tuning the permeability field than the later ones. *Lorentzen et al.* [1.43] successfully used EnKF with a two-phase flow model to tune model parameters in underbalanced drilling.

Brouwer et al. [1.5] applied EnKF in a controlled water flooding study of a two dimensional synthetic reservoir. They reported that main trends in the permeability field could be estimated using only a few days of observations, which they deemed sufficient for their control problem. However, as more observation times were included, the quality of the estimates deteriorated; the EnKF was believed to filter the additive noise in the later observations that had little incremental information when reservoir reached steady state.

Nævdal et al. [1.44] also applied EnKF for the tuning permeability field and estimating saturations and pressures in a two dimensional three phase reservoir model using observations of bottom hole pressure, watercut, and gas oil ratio. Despite the good estimation of the permeability field, they observed an unexplained deterioration of filter performance (permeability estimates) after a few update steps. In another study by *Gu et al.* [1.45], history matching of PUNQ-S3 [1.46] was studied using the EnKF. They concluded that the EnKF results were better than the gradient -based automatic history matching algorithms even with relatively small number of ensemble members (40 in their case).

Liu et al. [1.47] applied the EnKF to estimate geologic facies using a truncated pluri-Gaussian model to generate random facies realization. They concluded that the EnKF was highly efficient and outperformed gradient-based minimization methods when estimating facies boundaries. *Gao et al.* [1.47] compared EnKF with the Randomized Maximum Likelihood (RML) method for quantification of uncertainty in PUNQ-S3. They concluded that for the PUNQ-S3 problem, the RML and EnKF methods gave reliable assessments of the uncertainty in reservoir performance prediction.

In a recent study, *Wen et al.* [1.49] used EnKF for parameter estimation with synthetic reservoir models. They added a new step to the EnKF, which solved the forward flow equations with updated parameters, "confirming step", to ensure that static parameters and dynamic variables remain consistent and satisfy the dynamic model constraint.

Although EnKF was deemed promising in the studies cited above, many results remain inconclusive, with several showing filter divergence (deterioration of filter performance in time). The preliminary results obtained from these investigations have thus warranted further research for better understanding of EnKF application with reservoir modeling. The preliminary results of applying the EnKF for history matching are promising, yet better understanding of its properties and performance characteristics (such as robustness and divergence issues) with reservoir models are needed before it can be operational in practical settings.

In this thesis, application of the EnKF for history matching of oil reservoir is considered. One of the motivations of this thesis is to address the performance issues reported by some of the above authors who reported degradation of EnKF updates in time. The main objective, however, is to develop an efficient estimation framework that is suitable for characterization of large oil reservoirs.

1.4 Reservoir Parameterization

As mentioned in the previous sections, reservoir characterization is closely related to reservoir description. While grid-oriented description is suitable for forward modeling, a parametric description is often more appropriate in history matching and should be considered. The motivation behind parameterization is discussed next, followed by an overview of commonly used parameterization techniques.

1.4.1 Motivation for Parameterization

History matching solutions are well known to be non-unique; in other words, several solutions can be found that give satisfactory and essentially indistinguishable fits to observations [1.15]. This is due to the overwhelming number of unknown parameters compared to the sparse measurements. One way of improving this situation is "parameterization" of the high dimensional unknown properties. When reservoir properties are described using grid values, there is a high level of redundancy due to inherent correlations between neighboring blocks within a geological feature. While descritized grid values are required by flow simulators to solve forward model equations, the inverse problem is severely ill-posed when too many grid values are estimated. This can result in excessive computation, multiple local solutions, and potentially inconsistent estimation output. Parameterization aims at removing the redundancy in reservoir description before solving the inverse problem.

1.4.2 Parameterization Approaches

The simplest form of parameterization is traditional zonation [1.50], and a more recent adaptive multi-grid version of it [1.51]. This approach reduces the number of parameters by starting with a very coarse grid system and performing downscaling in areas of the reservoir where more refinement is needed. Another approach is to include additional information about the reservoir properties from their prior knowledge [1.15]. When strong structural correlations exist between individual grid blocks in a spatially

distributed property (such as permeability), most of this information can be captured in a lower dimensional subspace, often defined by spectral decomposition of the prior distribution of the parameters.

A classical subspace projection that is used in approximation analysis was developed by Karhunen [1.52] and Loeve [1.53]. In this approach, known as Karhunen-Loeve (KL) expansion, a discrete parameter set is projected onto the spectral space of its covariance matrix. This projection decorrelates the underlying parameter set (diagonalizes its covariance matrix), and separates the spectral content of the parameters. For highly correlated fields (such as permeability and porosity) a very small subspace, defined by the leading eigenvectors of the parameter covariance matrix, carries most of the important information contained in the original gridded description.

The KL expansion is an approximation technique that has been used in petroleum engineering to reduce the parameter space for Baysian estimation problems [1.15]. Recently, *Sarma et al.* [1.6] also used the same method to substantially reduce simple two-dimensional permeability fields before history matching. While KL expansion is an optimal (in a minimum mean squared error sense) [1.53, 1.54] approximation technique, it suffers from two major limitations. First, it requires correct specification of the prior covariance, which is often not known *a priori* for unknown parameters. Second, it

requires eigenvalue decomposition of the covariance matrix, which can be very expensive for realistic problems.

Wavelet-based parameterization approaches have also been used for history matching [1.55]. Wavelets (and filter banks) separate the spectral contents of a given image and distribute them among different frequency subbands [1.56, 1.57]. What differentiates the wavelet method from other spectral decomposition approaches is that they provide information in both space and frequency after the transformation. This time/space localization property of wavelets makes them a favorable choice in edge detection and other applications [1.58].

This work presents a novel parameterization approach that uses yet a different transform , the discrete cosine transform (DCT). While DCT is a popular image compression method, it has never been used before for parameterization of reservoir properties. Its performance and robustness in reservoir parameterization is compared with those of the KL transform in Chapter 3.

1.5 Thesis Objective and Outline

The objective of this thesis is to develop an ensemble-based recursive history matching framework that includes an effective reservoir description algorithm, an efficient state

and parameter estimation scheme that provides estimation uncertainty measures. This is achieved with the introduction of a novel parameterization technique, the discrete cosine transform (DCT), which removes redundancy in reservoir description to result in a parsimonious model representation. This new approach is computationally superior to previously used transform-based approaches and provides flexibility for including prior information. The full estimation approach is explained in detail in the body of the thesis, and its success is demonstrated through several examples.

The outline of the thesis is as follows: Chapter 2 discusses the ensemble Kalman filter for history matching and estimation of permeability field and explains its reported divergence behavior using synthetic waterflooding experiments. Chapter 3 describes the DCT and demonstrates its use as a powerful parameterization technique. The DCT-based parameterization is then compared with the classical Karhunen-Loeve transform (KLT) and is shown to have superior flexibility and robustness. Chapter 4 combines the ensemble Kalman filter with the proposed DCT parameterization technique to develop an efficient history matching framework and illustrates its potential use in two-dimensional waterflooding experiments for permeability estimation. In Chapter 5, the proposed history matching framework is extended to a three dimensional realistic reservoir model (upscaled SPE10). The three-dimensional DCT parameterization is described and used to represent all states and parameters before history matching. Permeability and porosity are included as uncertain reservoir properties in the history matching of the SPE10. Finally, Chapter 6 presents the conclusions of this thesis and possible future research directions.

References

1.1. Smil, Vaclav (2005). "Energy at the Crossroads: Global Perspectives and Uncertainties". MIT Press.

1.2. Beamer A., Bryant I., Denver L., Saeedi J., Verma V., Mead P., Morgan C. Rossi D. and Sharma S., (1998): "From pore to pipeline, field-scale solutions.", *Oilfield Review,* Summer, 2-19.

1.3. Jansen J.D., Brouwer D.R., Naevdal G., van Kruijsdijk C.P.J.W., (2005): "Closed-loop reservoir management", *First Break*, 23, 43-48.

1.4. Kretz V., Le Ravalec M., Roggero F., (2004) : "An Integrated Reservoir Characterization Study Matching Production Data and 4D seismic.", SPE-88033.

1.5. Brouwer D. R., Nævdal G., Jansen J.D., Vefring E.H., van Kruijsdijk C. P. J. W., (2004): "Improved reservoir management through optimal control and continuous model updating.", SPE 90149.

1.6. Sarma P., Durlofsky L.J., Khalid A., Chen W.H., (2006): "Efficient real-time reservoir management using adjoint-based optimal control and model updating," *Computational Geosciences*, vol. 10, pp. 3-36 .

1.7. Kalaydijan F., Bourbiaux B., (2002): "Integrated Reservoir Management: A Powerful Method to Add Value to Companies' Assets. A Modern View of the EOR Techniques. Oil & Gas Science and Technology" – Rev. IFP, Vol. 57 (2002), No. 3, pp. 251-258.

1.8. Deutsch, C.V. (2002): *Geostatistical reservoir modeling*. Oxford University Press, New York.

1.9. Dutsch C.V., Journel A.G. (2002): *GSLIB: Geostatistical software library.* Oxford University Press.

1.10. Caers J., Zhang T., (2004), "Multiple-point geostatistics: a quantitative vehicle for integration of geologic analogs into multiple reservoir models." In: Integration of outcrop and modern analog data in reservoir models, *AAPG memoir* 80, p. 383-394.

1.11. Strebelle S., Journel A., (2001): "Reservoir Modeling Using Multiple-Point Statistics": paper SPE 71324 presented at the 2001 *SPE Annual Technical Conference and Exhibition*, New Orleans, Sept.30-Oct. 3.

1.12. Aziz K., Settari A. (1979): *Petroleum Reservoir Simulation.* Applied Science Publishers LTD, London.

1.13. ECLIPSE reservoir simulator: "Manual and Technical Description," *Schlumberger GeoQuest*, Houston (2006).

1.14. Mattax, C.C., Dalton, L.R.: *Reservoir simulation SPE monograph*, Volume 13, 60-61.

1.15. Gavalas G.R., Shah P.C., Seinfeld J.H., (1976): "Reservoir history matching by Bayesian estimation.", *Soc. Petrol. Eng. J.*, 16(6):337–350.

1.16. Schweppe F.C., (1973): *Uncertain Dynamic Systems*, Prentice-Hall, Inc., Englewood Cliffs, NJ.

1.17. Tarantola A., (2004): *Inverse Problem Theory. Methods for Model Parameter Estimation,* Philadelphia, PA: SIAM.

1.18. Bennett A.F., (2002): *Inverse modeling of the ocean and atmosphere*, Cambridge University Press, Cambridge

1.19. Van Leeuwen, P. J., and G. Evensen, (1996): "Data assimilation and inverse methods in terms of a probabilistic formulation", *Mon. Weather Rev., 124*, 2898–2913.

1.20. Oliver, D. S., He, N., and Reynolds, A. C., (1996): "Conditioning permeability fields to pressure data", *in European conference for the mathematics of oil recovery*, V, p. 1–11.

1.21. Zhang F., Reynolds A.C., (2002): "Optimization algorithms for automatic history matching of production data.", *8th European Conference on the Mathematics of Oil Recovery*, Freiberg, Germany.

1.22. Oliver D.S., Reynolds A.C., Bi Z., Abacioglu Y., (2001): "Integration of production data into reservoir models.", *Petroleum Geoscience,* vol. 7, no. SUPP, pp. 65-73(9).

1.23. Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E., (1953): "Equation of state calculations by fast computing machines." *J. Chemical Physics* 21:1087- 1092, 1953.

1.24. Spall J.C., (2003): *Introduction to stochastic search and optimization: estimation, simulation and control*, John Wiley & Sons Inc., New Jersey.

1.25. Beckner B.L., Song X., (1995): "Field development planning using simulated annealing-optimal economic well scheduling and placement.", SPE Paper 30650.

1.26. Ouenes A., Brefort B., Meunier G., Dupere S., (1993): "A new algorithm for automatic history matching: Application of simulated annealing method (SAM) to reservoir inverse modeling.", SPE Paper 26297.

1.27. Sen M.K., Datta-Gupta A., Stoffa P.L., Lake L.W., Pope G.A., (1995): "Stochastic reservoir modeling using simulated annealing and genetic algorithms.", *SPE Formation Evaluation.*

1.28. Romero C.E., Carter J.N., Zimmerman R.W., Gringarten A.C., (2000): "Improved reservoir characterization through evolutionary computation", *in Proc. 2000 SPE Ann. Tech. Conf. Exh., Soc. Pet. Eng.*, Dallas, paper SPE 62942.

1.29. Chunduru R.K., Sen M.K., Stoffa P.L., (1997): ''Hybrid optimization methods for geo-physical inversion'' *Geophysics* 62, 1196–1207.

1.30. Caers J., Ma X., (2002): "Modeling Conditional Distribution of Facies from Seismic Using Neural Nets.", *Math. Geol.* 34, No.2, 143-167.

1.31. Reis, L.C., Hu, L.Y., and Eschard, R., (2000): "Production Data Integration Using a Gradual Deformation Approach: Application to an Oil Field (Offshore Brazil)," paper SPE 63064 presented at the *Annual Technical Conference and Exhibition*, Dallas, Texas.

1.32. Roggero F. and Hu L.Y., (1998): "Gradual deformation of continuous geostatistical models for history matching.", Paper SPE 49004 prepared for the *SPE ATCE*, New Orleans.

1.33. Hoffman B.T., and Caers J., (2003): "Geostatistical History Matching Using a Regional Probability Perturbation Method," paper SPE 84409 presented 78th Annual Fall Meeting of SPE in Denver, CO.

1.34. Kitanidis, P. K., 1995, Quasi-linear geostatistical theory for inversing:Water Resour. Res., v. 31, no. 10, p. 2411–2419.

1.35. Floris F.J.T., Bush M.D., Cuypers M., Roggero F., Syversveen A-R., (2000): "Methods for quantifying the uncertainty of production forecasts - a comparative study.", *Journal of Petroleum Geoscience.* p. 1-22.

1.36. Kalman R.E., (1960): "A New Approach to Linear Filtering and Prediction Problems", *Transactions of the ASME--Journal of Basic Engineering.* vol. 82, pp. 35-45.

1.37. Jazwinski A.H., (1970): *Stochastic processes and filtering theory.* Academic, San Diego, California.

1.38. Evensen G., (1994): "Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics." *J Geophys Res* 99(C5): 10 143–10 162.

1.39. Evensen G., (2003): "The ensemble Kalman filter: theoretical formulation and practical implementation.", Ocean Dyn. 53, 343– 367.

1.40. Houtekamer P.L., Mitchell H.L., (1998): "Data assimilation using an Ensemble Kalman Filter technique.", *Mon Weather Rev* 126: 796–811.

1.41. Reichle R.H., McLaughlin D.B., Entekhabi D., (2002): "Hydrologic data assimilation with the Ensemble Kalman Filter.", *Mon Weather Rev* 130: 103–114.

1.42. Nævdal G., Mannseth T., Vefring, E.H., (2002): "Near-well reservoir monitoring through ensemble Kalman filter.", SPE-75235.

1.43. Lorentzen R.J., Nævdal G., Lage A.C.V.M., (2003): "Tuning of parameters in a two-phase flow model using an ensemble Kalman filter", *International Journal of Multiphase Flow*, 29, 1283-1309.

1.44. Nævdal G., Johnsen L.M., Aanonsen S.I., Vefring E.H., (2003): "Reservoir monitoring and continuous model updating using ensemble Kalman filter." SPE-84372.

1.45. Gu Y., Oliver D.S., (2005): "History matching of the PUNQ-S3 reservoir model using the ensemble Kalman filter", SPE 89942.

1.46. PUNQ-S3 project website: http://www.nitg.tno.nl/punq/.

1.47. Liu N., Oliver D.S., (2005): "Ensemble Kalman filter for automatic history matching of geologic facies", *Journal of Petroleum Science and Engineering*, 47, 147-161.

1.48. Gao G., Zafari M., Reynolds A.C., (2005): "Quantifying Uncertainty for the PUNQ-S3 Problem in a Bayesian Setting With RML and EnKF," *proceedings the 2005 SPE Reservoir Simulation and Symposium*, Houston, Texas, USA.

1.49. Wen X.H., Chen W., (2005): "Real-time reservoir model updating using ensemble Kalman filter,", SPE-92991. Proceedings of the *SPE Reservoir Simulation Symposium*, pp. 1 – 14.

1.50. Jacquard P., Jain C., (1965): "Permeability distribution from field pressure data". *Soc. Pet. Eng. Journal* 281-294.

1.51. Aanonsen S.I., (2005): "Efficient history matching using multiscale technique.", SPE-92758.

1.52. Karhunen K., (1947): "Ueber lincare Methoden in der Wahrscheinlichts-keitsrechnung". *Ann. Acad Sci. Fenn. Ser.* A. I. 37, 3-79.

1.53. Loeve M.M., (1978): "Probability Theory". New York, NY: Springer-Verlag, 2nd Vols., Fourth Edition.

1.54. Jain A.K., (1989): *Fundamentals of Digital Image Processing*, Prentice Hall.

1.55. Sahni I., Horne R.N., (2005): "Multiresolution Wavelet Analysis for Improved Reservoir Description," , *SPEREE*, 8(1), 53-69.

1.56. Mallat, S.G., (1998): *A Wavelet Tour of Signal Processing,* Academic, San Diego.

1.57. Strang G., Nguyen T., (1995): *Wavelets and Filter Banks*", Wellesley-Cambridge Press, Wellesley.

1.58. Gonzalez R.C., Woods R.E., (2002): *Digital Image Processing*, 2nd ed., Prentice Hall, Upper Saddle River, NJ.

Chapter 2

Estimating Reservoir Permeabilities with the Ensemble Kalman Filter

2.1 Summary

Efficient management of smart oilfields requires a reservoir model that can provide reliable forecasts of future production as well as realistic measures of prediction uncertainty. Reliable forecasts depend on an accurate representation of reservoir geology, which is conveyed largely by the permeabilities used in the reservoir simulator. Since these permeabilities cannot be measured directly they must be inferred from measurements of related variables, using procedures such as history matching or Bayesian estimation. The ensemble Kalman filter is an attractive option for permeability estimation in real-time reservoir control applications. It is easy to implement, provides considerable flexibility for describing geological heterogeneity, and supplies valuable information about prediction uncertainty. In this chapter we investigate the procedure used to generate the random replicates of an ensemble Kalman filter that estimates reservoir permeabilities. We consider two synthetic water flooding problems based on

"true" permeability distributions characterized by conductive channels. The permeability ensembles are obtained from either a classical covariance-based random field generator or a multi-point geostatistical generator. If the ensemble replicates are derived from a covariance model that does not provide for high permeability channels or if the multi-point geostatistical training images do not properly describe the channel geometry the Kalman filter has difficulty identifying the correct permeability field. In fact, in both cases the permeability estimates tend to diverge from the true values as more measurements are included. However, if the filter ensemble replicates are generated by a training image that captures the dominant features of the true permeability field, the filter's estimates are much better. These results emphasize the importance of generating realistic permeability replicates when using ensemble methods to estimate reservoir properties. In fact, a realistic permeability ensemble appears to be essential for successful estimation performance. In practical applications where the true permeability distribution is highly uncertain the prior information used for ensemble generation should properly reflect the full range of possible geological conditions.

2.2 Introduction

Parameter estimation is one of the fundamental challenges encountered in real-world reservoir simulation and control applications. The estimation process is greatly complicated by the nature of geological heterogeneity, which is not amenable to simple parametric descriptions. The geological features that govern the flow of liquids in a petroleum reservoir often form preferred pathways that are geometrically complex and

difficult to identify from available data. The spatial configuration and properties of these features must be inferred from geophysical, pressure, and flow measurements that are only indirectly related to the parameters of interest. If the independent parameters are assigned to every cell (or pixel) of the simulator's computational grid the estimation procedure has sufficient flexibility to reproduce more or less arbitrary geometries but the estimation problem may be poorly posed, since many different parameter sets can give comparable fits to available measurements [2.1-2.3]. On the other hand, if the number of independent parameters is constrained in an effort to force the problem to be well-posed it may be difficult to adequately describe the true reservoir geometry. Methods for describing the spatial variability of reservoir properties must be both geologically realistic and efficient if the associated parameter estimation procedure is to give satisfactory results. This is true whether the procedure is based on manual adjustment of model inputs or on an automated optimization or history matching procedure.

The parameter estimation approach adopted in a particular application must consider the purpose of the reservoir simulation. Recently, there has been considerable interest in real-time reservoir control applications where reservoir simulators are used to guide operational activities such as water flooding [2.4, 2.5]. In such cases real-time measurements are used to continually update the simulator so that it can provide better predictions of the reservoir's response to well adjustments. This, in turn, enables the control algorithm to make the best possible decisions about present and future well settings. Figure 2.1 shows the basic components of this real-time approach to reservoir operations.

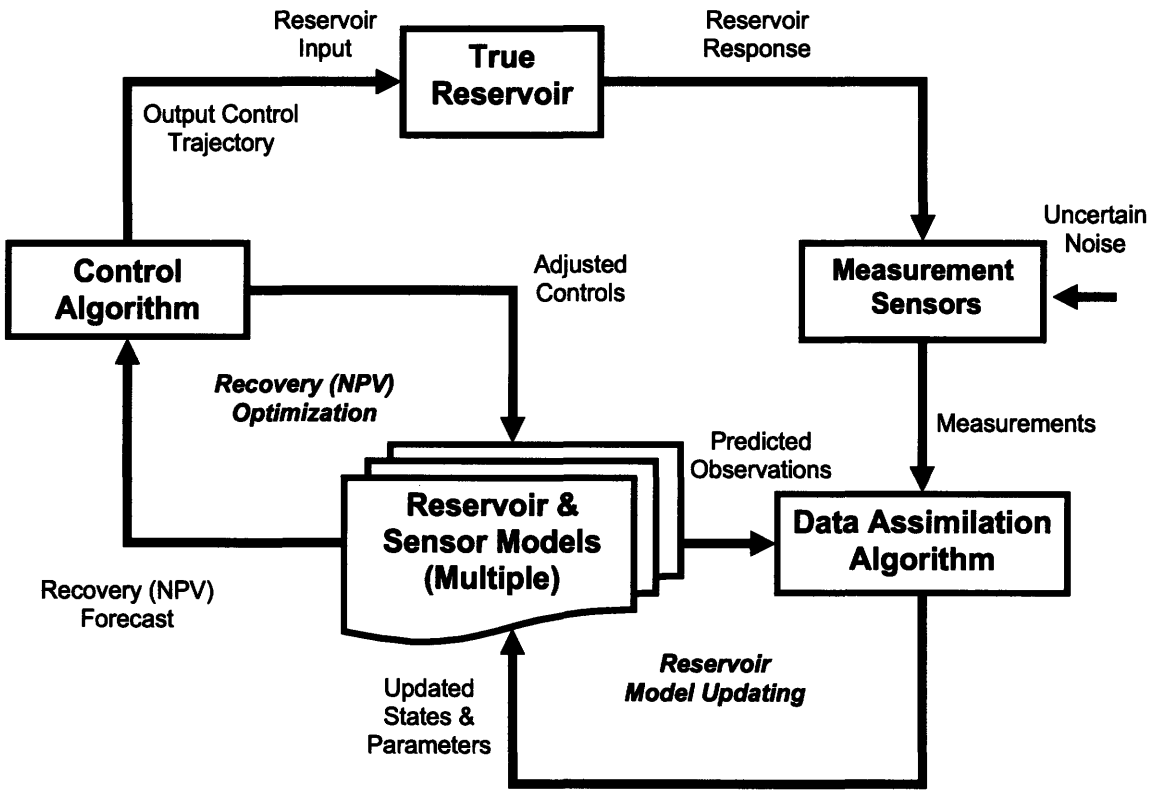


Figure 2.1 Reservoir closed-loop control body diagram adapted from [4]. Model updating (history matching) component is shown in the bottom right loop.

Real-time control requires that simulator parameters and related states be updated frequently, whenever new information becomes available. Automated estimation methods are usually preferred for such applications because of the relatively short turn-around times, the large amount of data to be processed, and the complexity of the estimation problem. In fact, automated real-time parameter and state estimation (or data assimilation) algorithms are now used routinely in many geophysical applications, most notably in weather forecasting [2.6, 2.7]. Sequential estimation methods such as Kalman

filtering [2.8] are particularly convenient for real-time control problems since they work only with the most recent measurements and estimates, rather than the entire measurement history. However, such methods make assumptions and impose computational demands that may be problematic in some applications.

One of the most popular sequential estimation techniques is the ensemble Kalman filter [2.9], which is especially well suited for nonlinear problems where model and measurement uncertainties enter in complex ways. In recent years, this method has been introduced to the petroleum engineering literature as a way to estimate uncertain reservoir properties and states [2.10-2.12]. The ensemble Kalman filter does not require an adjoint model, is straightforward to implement with commercial reservoir simulators, and is readily parallelized.

Ensemble Kalman filters propagate and update many independent realizations of the uncertain reservoir inputs and states. In non-compositional reservoir simulation applications the model states typically include pressure and saturation. Each realization of the state vector is generated by running the simulator with a particular set of inputs (e.g. permeabilities and other formation properties) drawn at random from specified populations. The input populations should be geologically plausible while including enough variability to properly account for uncertainty. At measurement times input and

propagated state replicates can be updated to account for new information. The update process is based on sample covariances derived from the ensemble.

Brouwer et al. [2.4] used an ensemble Kalman filter to estimate pixel-scale permeabilities from bottom hole pressure observations in a synthetic water flooding experiment. The permeability realizations were generated by randomly sampling a population of exponentially correlated isotropic Gauss-Markov fields with a specified mean, variance, and correlation scale. The true permeability field used to generate the synthetic bottom hole pressure measurements included a preferential pathway that was qualitatively different from any of the sample realizations provided to the filter.

The well controls (pressures and flow rates) used in the *Brouwer et al.* example were derived from a nonlinear programming algorithm that maximized the net present value of recovered oil, evaluated from the current time to the end of the water flood. The controls varied over time, in response to changing reservoir conditions. This dynamic well excitation tends to make the problem better posed by increasing the sensitivity of the measured states to the uncertain parameters.

The permeability estimates obtained from the ensemble Kalman filter in [2.4] were able to capture some aspects of the true permeability field at early time but they soon degraded. In particular, connected channels that were partially identified at early times gradually broke up into disconnected pixels of different permeability values. In this

revealing experiment the estimation error actually increased as more information was added. This is a counterintuitive result that could not happen if the system were linear and the estimator optimal. The authors speculated that measurement noise began to dominate the update process as the system approached steady-state conditions behind the advancing water-oil interface. An optimal estimator would handle this situation by decreasing its measurement weights over time.

The ensemble Kalman filter's ability to properly characterize uncertain reservoir properties depends greatly on the nature of the ensemble it uses. The individual replicates of this ensemble must be geologically realistic and variations across the ensemble must properly capture the dominant sources of uncertainty. If these requirements are not met, the filter's measurement updates may not be helpful and, in fact, may even be counterproductive. In this chapter we reexamine the performance of the ensemble Kalman filter for the water flooding problem. In particular, we consider the benefits of using multi-point geostatistics to generate unconditional permeability ensembles from geologically realistic training images [2.13, 2.14]. We test this ensemble generation approach on two variants of the *Brouwer et al.* synthetic water flooding experiment. Our first experiment uses the same "true" permeability field, the same pixel-based parameterization, and the same type of ensemble Kalman filtering algorithm as *Brouwer et al* [2.4]. Our second experiment also relies on a pixel-based parameterization and an ensemble Kalman filtering algorithm but it uses the "true" permeability field adopted by *Sarma et al.* [2.5] in their test of a nonlinear least-squares estimation procedure. These

two synthetic experiments suggest that the ensemble Kalman filter is able to provide reasonable estimates of geologically realistic permeability fields if the underlying ensemble is properly chosen.

2.3 The Ensemble Kalman Filter

We begin with a brief review of the ensemble Kalman filter, as applied to real-time permeability estimation. The Kalman filter can be viewed as a Bayesian estimator that approximates conditional probability densities of the time-dependent state vector x_t [2.8]. The sequential formulation of the filter distinguishes a forecast (or prior) density $p[x_t \mid y_{0:t-1}]$ conditioned on all measurements $y_{0:t-1}$ taken prior to time t and an updated (or posterior) density $p[x_t \mid y_{0:t}]$ conditioned on all measurements $y_{0:t}$ taken through t. In our application the state vector consists of the pressure and saturation at the nodes of the spatially discretized simulator computational grid (since capillary pressure is neglected the water and oil pressures are the same). In addition, uncertain model parameters are included in the state vector. This so-called state augmentation approach enables the parameters to be estimated together with the other system states. The measurement vector in our application consists of flow rates and bottom hole pressures measurements inside the wells.

The ensemble Kalman filter approximates the forecast and updated densities with relatively small ensembles of N random realizations, denoted by $x^j_{t|t-1}$ and $x^j_{t|t}$, respectively, where $j = 1,\ldots,N$ represents a particular replicate. The sequence of forecasts and updates is initialized with an ensemble $x^j_{0|0}$ drawn at random from a specified population of initial states. Subsequent forecasts are obtained from the simulator, which may be written as:

$$x^j_{t|t-1} = f_t(x^j_{t-1|t-1}, u_{t-1}, w^j_{t-1}) \quad ; \quad j = 1,\ldots,N \qquad (2.1)$$

where u_{t-1} is a vector of known (non-random) time-dependent boundary conditions and controls and w^j_{t-1} is a random vector that accounts for uncertain model errors. The function $f_t(\cdot,\cdot,\cdot)$ represents the reservoir simulator, which generates states at t from states and inputs at t-1. Time dependent states such as pressure and saturation will generally change over the forecast period while time-invariant states, such as permeabilities, will not.

The updated replicates at t are obtained from a version of the classical Kalman filter update [2.6, 2.8]:

$$x_{t|t}^{j} = x_{t|t-1}^{j} + Cov[x_{t|t-1}^{j}, y_{t|t-1}^{j}]Cov^{-1}[y_{t|t-1}^{j}, y_{t|t-1}^{j}][y_t - y_{t|t-1}^{j}] \quad ; \quad j = 1,\ldots,N \tag{2.2}$$

Here the notation $Cov[\cdot,\cdot]$ represents the sample covariance between the ensembles associated with the two arguments, y_t is the actual measurement at t, and $y_{t|t-1}^{j}$ represents a perturbed measurement prediction obtained from the following measurement equation:

$$y_{t|t-1}^{j} = g_t(x_{t|t-1}^{j}) + v_t^{j} \quad ; \quad j = 1,\ldots,N \tag{2.3}$$

where v_t^{j} is a vector of measurement errors, drawn at random from a specified population, and $g_t(\cdot,\cdot)$ relates the measurements at t to the states at t. Equations (2.1), (2.2), and (2.3), together with the initial state $x_{0|0}^{j}$, define the ensemble Kalman filter recursion for the problem of interest here.

The ensemble Kalman filter approach to coupled state/parameter estimation has several characteristics which deserve to be mentioned. First, the method offers the flexibility of generating the random realizations $x_{0|0}^{j}$, w_t^{j}, and v_t^{j} from any desired population. In the real-time reservoir control context, this means that permeability realizations included

in $x_{0|0}^{j}$ may be drawn from a population of physically realistic alternatives. These populations can, for example, be constructed from libraries of preferred channel configurations that share certain distinctive features.

Another attractive characteristic of the ensemble Kalman filter is its ability to generate non-Gaussian sample distributions of the states $x_{t|t-1}^{j}$ and $x_{t|t}^{j}$. There is no need to linearize or otherwise approximate the state transition function $f_t(\cdot,\cdot,\cdot)$ or to assume that the random states and inputs are Gaussian. However, the ensemble Kalman filter converges to the true conditional densities $p[x_t \mid y_{0:t-1}]$ and $p[x_t \mid y_{0:t}]$, only when all prior states and measurements are jointly Gaussian. This typically only occurs when the state and measurement equations are both linear. In most petroleum reservoir applications the joint Gaussian condition is not met and the filter's sample densities and moments are only approximations. In practice, the updated sample mean $E[x_{t|t}^{j}]$ is typically used as a point estimate for characterization and control purposes. This is the estimate that we will be examining when comparing different ensemble generation approaches.

It should be noted that the covariance inversion operation in (2) is computationally expensive and can be ill-conditioned for large problems. This can be avoided by using the pseudo inverse procedure based on singular value decomposition, as proposed in [2.6]. Also, several variants of the EnKF have been developed to improve the computational

efficiency of the algorithm and to address some of the issues in its implementation, mostly due to sampling errors (see [2.6-2.7] for a brief review).

2.4 Experimental Setup

We revisit the water flooding example originally studied in [2.4] and later examined in [2.5]. In this example a 450(m) × 450(m) × 10(m) synthetic reservoir is discretized into a two-dimensional 45×45×1 uniform grid block system, as shown in Figure 2.2. The simulations are performed with the commercially available ECLIPSE [2.15] reservoir simulator, which is set up for two phase (oil and water) black oil flow. The total simulation time is 1080 days, divided into 12 intervals of 90 days. Horizontal wells with 45 ports are used to inject water into the left side of the reservoir and to produce oil and water from the right side end.

The injection wells are operated with specified flow rates while the production wells are operated with specified bottom hole pressures. In [2.4] and [2.5] the well port settings at each simulation time were determined by an optimization algorithm designed to maximize the net present value of benefits obtained from oil recovery. Here we specify the port settings beforehand and focus on permeability estimation rather than optimal recovery.

A total of one pore volume of water is injected into the reservoir during the simulation. The injection and production wells are each divided into 3 different groups of 15 well ports and the simulation is divided into 6 time periods, with each period lasting 180 days. The ports are represented in Figure 2.2 by small colored circles on either side of the simulation domain. Each column of circles shows the well settings used in one of the 6 operating periods. The total amount of injected water is divided evenly over the 6 production periods and among the injection ports open in any given period. The injection strategy can be summarized as follows:

Injection ports:

- Periods A (0-180 days) and E (720-900 days) – Uniform distribution of 1/6 PV of water among all ports.
- Periods B (180-360 days), and D (540-720 days) – Uniform distribution of 1/6 PV of water among Groups 1 and 3 respectively.
- Periods C (360-540 days), and F (900-1080 days) – Uniform distribution of 1/6 PV of water among Group 2.

The production strategy is:

Production ports:

- Periods A and E – All production ports have a specified bottom-hole pressure of 2990 psi.

- Periods B and C – Production ports in Groups 1 and 3 have a specified bottom-hole pressure of 2990 psi while the pressure at other ports is kept at 3000 psi.
- Periods D and F – Production ports in Group 2 have a specified bottom-hole of 2990 psi while pressure at other ports was kept at 3000 psi.

In this study the only source of simulator uncertainty is the permeability, which is treated as a random field. Initial and boundary conditions are assumed to be known perfectly and dynamic model errors are assumed to be negligible. In situations where these assumptions may not hold additional error sources may be included in the ensemble filtering process. The initial reservoir pressure and connate water saturation are 3000 psi and 0.10, respectively, throughout the reservoir. The realizations used to construct the permeability ensemble are generated with the *snesim* algorithm of the Stanford Geostatistical Modeling Software (S-GeMS) [2.14]. This software relies on a multiple-point geostatistical method that uses a specified training image to determine the general structure of the realizations. In our experiments different training images were used to evaluate the impact of the ensemble on filter performance.

The ensemble Kalman filter uses two types of measurements in its updates of state and input replicates: 1) bottom hole pressure observations at each of the 45 ports in the injection wells and 2) oil and water flow rate measurements at each of the 45 ports in the production wells. In each experiment the "true" injection well bottom hole pressures and

production well flow rates are generated by running the simulator from a specified "true" permeability field. Uncorrelated zero mean random measurement errors are added to these "true" pressures and flow rates. The standard deviations of the random measurement errors are 20 psi and 20 sbpd for bottom hole pressures and flow rates, respectively.

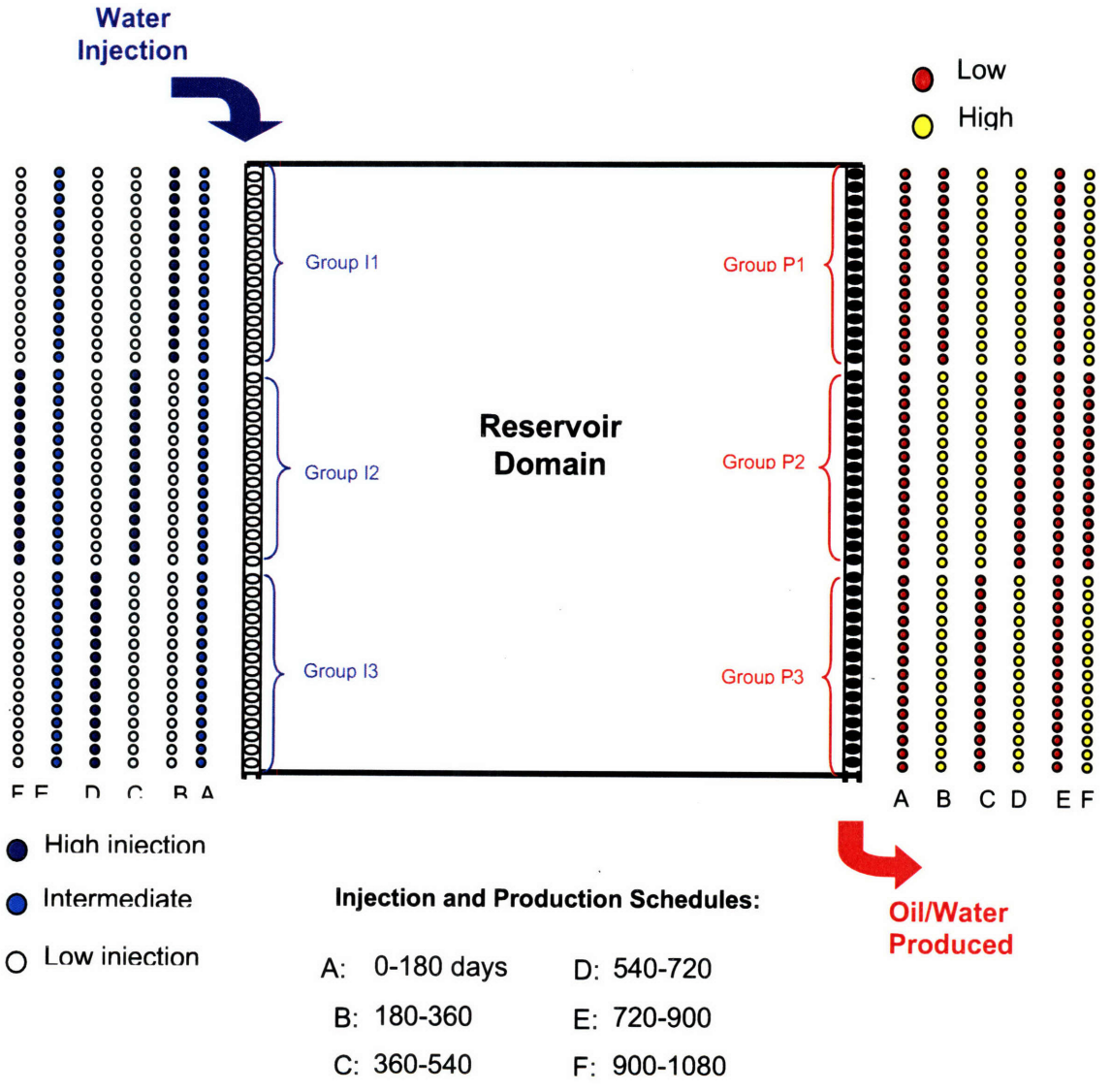


Figure 2.2 Reservoir configuration with injection/production schemes in a waterflooding experiment. Injection wells with dark blue, light blue, and cyan circles have high, intermediate, and low injection rates, respectively. The red and yellow circles in the production wells indicate low and high bottom hole pressures, respectively.

The ensemble filter estimates the natural log of permeability, which is then transformed to permeability for input to the simulator. *Jafarpour et al.* [2.16] discuss some of the limitations of log permeability transformations for parameter estimation applications. We use this transformation here to maintain consistency with [2.4] and [2.5]. In some cases the classical ensemble Kalman filter update can give unphysical saturation values outside the range [0, 1]. In order to avoid this problem, our Kalman filter works with the transformed saturation S^*, which is distributed over ($-\infty$, $+\infty$) and is computed as follows:

$$S^* = erf^{-1}(2 \cdot S - 1) \quad \rightarrow \quad S = \frac{1}{2} erf(S^* + 1) \tag{2.4}$$

where *erf* represents the error function. After the filter update S^* is transformed back to the saturation S for use in the reservoir simulator.

2.5 Experiments and Discussion

2.5.1 Experiment 1.

This experiment uses the "true" log permeability field adopted in the *Brouwer et al* synthetic water flooding study [2.4] and shown in Figure 2.3a. This "true" field includes a distinctive high permeability channel that cuts through the reservoir from the injection to production sides (see [2.4] for more details). The log permeability ensemble generated by *Brouwer et al.* [2.4] consists of 100 realizations of a Markov random field with a

Gaussian spatial correlation function. The correlation length is a normally distributed random variable with a mean of 20 grid blocks and a standard deviation of one grid block. The mean and standard deviation of the log normally distributed random permeability are 200 mD and 1.5 mD, respectively. The "true" field is qualitatively different in structure than the log permeability ensemble since it contains a channel that does not appear in any of the ensemble replicates.

In our version of the *Brouwer et al.* synthetic water flooding experiment we use two different permeability ensembles generated with multi-point geostatistics. These are based on two different training images, each consisting of two facies -- sandstone and shale. The sandstone and shale permeabilities are 10,000 mD and 500 mD, respectively. These are comparable in magnitude to the channel and mean background permeability in the *Brouwer et al.* field. The two log permeability training images and some typical replicates are shown in Figures 2.3b and 2.3c, respectively. Note that the training images cover a significantly larger region than the simulation domain.

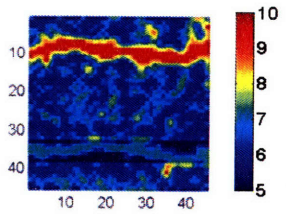
Ensemble Kalman filters rely on a limited number of samples (or replicates) drawn from a specified population or probability distribution. If the sample size is too small the resulting sample statistics may be inaccurate and the filter's performance may suffer [2.17, 2.18]. The dependence of estimation accuracy on ensemble size is application-specific. In our study we performed a sensitivity analysis to identify an ensemble size that yields robust permeability estimates with the least possible computational cost. This

study indicated that the filter results for our synthetic experiments converge only if the ensemble size is at least 300. Based on these results we selected an ensemble size of 300 for our study. This is significantly larger than the 100 replicate ensemble used in [2.4].

In Experiment 1-1 we used Training Image 1 (Figure 2.3b) to generate the ensemble members for estimation. This training image has wider channels than the true log permeability field used in [2.4]. It is the same image used in the synthetic water flooding experiment described in *Sarma et al.* [2.5]. Figure 2.4b shows the log permeability estimate (first row) and log permeability estimation error (the difference between "true" and ensemble mean log permeability, second row) obtained after each analysis step with an ensemble derived from Training Image 1. As seen in these figures, the ensemble estimate fails to capture the spatial structure of the high permeability channel. Also, the initial updates are better than the later ones, as was observed by Brouwer e al. [2.4].

Figure 2.4b also shows the standard deviation of the updated log permeability ensemble (third row) and the ensemble mean of the oil saturation (fourth row). The standard deviation is generally highest in the middle of the domain where the estimation pixels are furthest from the measurements. The filter's ensemble standard deviation significantly underestimates the actual estimation error, indicating that the filter is overconfident. This overconfidence tends to make the filter ignore information from the measurements. The

a) Experiment 1 "True" Log Permeability



b) Log Permeability Training Images Used in Experiments 1 and 2

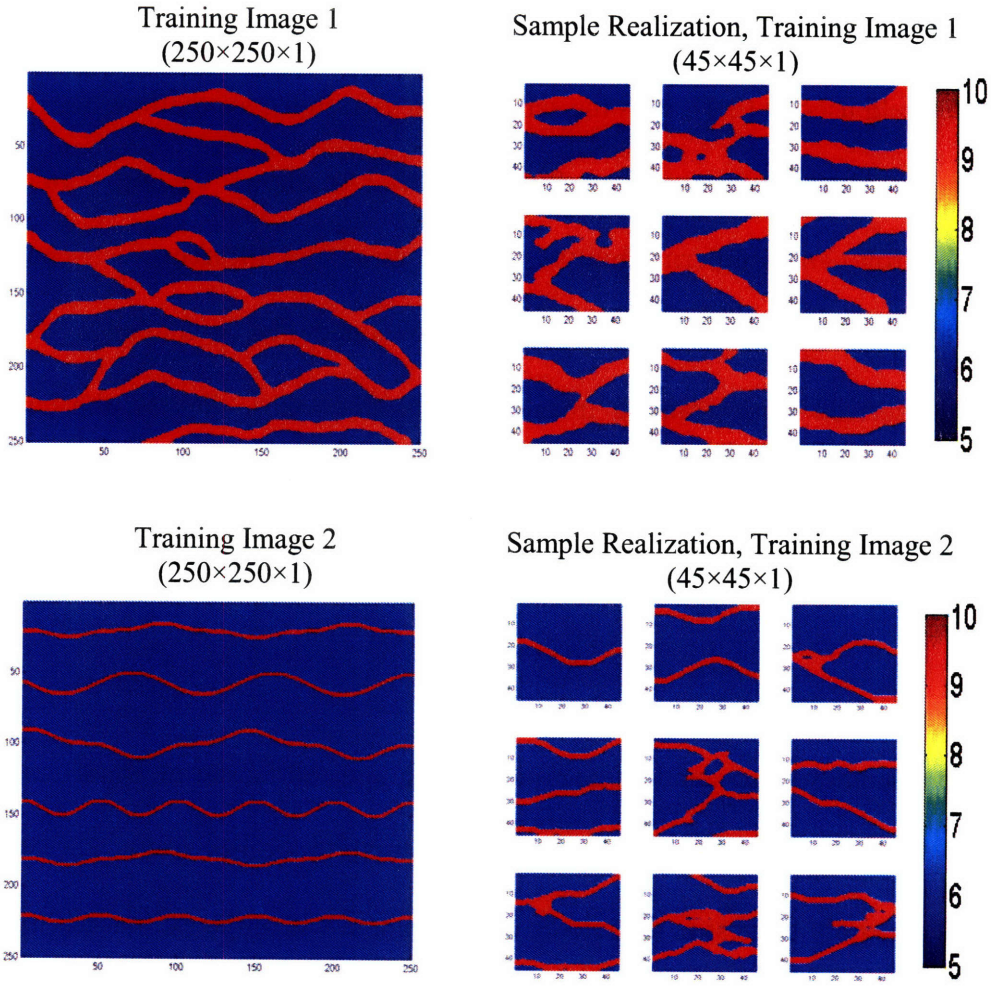


Figure 2.3 (a) True log-permeability for Experiment 1; (b) Log permeability training images used in Experiments 1 and 2. Training Image 1 has wide channels and is used in Experiments 1-1, 2-1, and 2-2. It is shown in the first row along with nine sample permeability realizations, generated by SGeMS using multiple point geostatistics. Training Image 2 has narrower channels and is used in Experiment 1-2. It is shown in the second row along with nine sample permeability realizations generated by SGeMS.

mean oil saturation differs most from the true saturation in the beginning of the simulation, where it misses the water front advancing through the high permeability channel. Near the end of the simulation the estimate and the true saturation values tend to converge, despite the relatively poor quality of the permeability estimate. This appears to reflect the fact that the water front moves slower at the two ends of the reservoir, where the permeability is underestimated while it moves faster in the middle where the permeability is overestimated. Overall, the average front speed is reasonably close to the true case.

The log permeability ensemble used in Experiment 1-1 is based on a training image characterized by high permeability channels that are consistently wider than the channel appearing in the "true" field. The average channel width in the true field is about 3-4 grid blocks (30-40 meters) while the Training Image 1 channels have an average channel width of about 10 grid blocks (100 meters). As a result, channels in the ensemble members generated from the Training Image 1 are consistently too wide and the sample covariances used to update these replicates do not adequately describe the spatial features of the log permeability field. It is difficult to predict the effects of such ensemble specification errors. In the particular example considered here our synthetic simulation indicates that these errors are sufficient to significantly degrade filter performance.

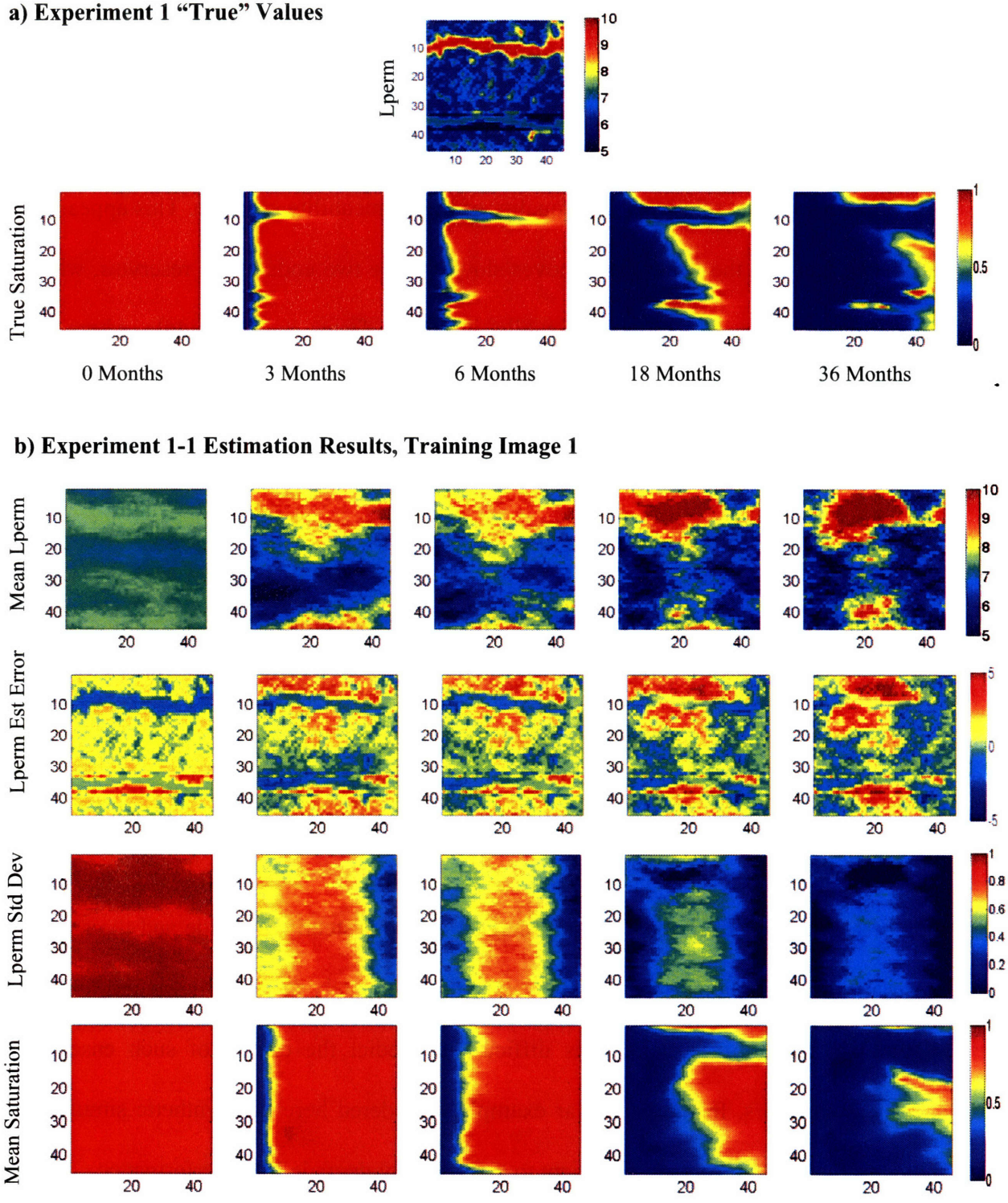


Figure 2.4 True log-permeability field in Experiment 1 (first row) and its corresponding saturation field (second row) for the injection/production scenarios shown in Figure 2.2; (b) Estimated mean (first row), error (second row), and standard deviation (third row) of the log-permeability field and estimated mean saturation field (fourth row) for Experiment 1-1 after the EnKF updates at 0, 3, 6, 18, and 36 months. The initial permeabilities were generated from Training Image 1 with wide channels (Figure 2.3b)

The importance of the training image and the resulting log permeability ensemble can be demonstrated if we run a new Experiment 1-2, which is the same as Experiment 1-1 except that we use Training Image 2 (Figure 2.3b). Training Image 2 has narrower channels and gives a better description of the true channel geometry. The results of Experiment 1-2 are shown in Figure 2.5. It is readily apparent that the filter is better able to capture the channel. However, the log permeability error plots indicate that the position of the estimated channel is displaced somewhat in the middle of the domain. This likely reflects the fact that different geometries in this region provide nearly the same fits to the measurements, which are taken at the edges of the simulation domain. So the filter has difficulty converging to the "true" channel position, even though it correctly infers that there is a channel in the general vicinity. The ensemble standard deviation is noticeably smaller in Experiment 1-2 than in Experiment 1-1. This indicates that the replicates are less dispersed around the mean in Experiment 1-2.

While experiment 1-2 demonstrates the dependence of EnKF performance on appropriate choice of the prior permeability replicates, it may not be realistic to assume channel properties that are similar to those present in the true permeability field. Therefore, a more interesting problem is to use a mixture of permeability replicates from training images 1 and 2. This is done in experiment 1-3, where the same number of ensemble replicates is used with each training image providing half of the ensemble.

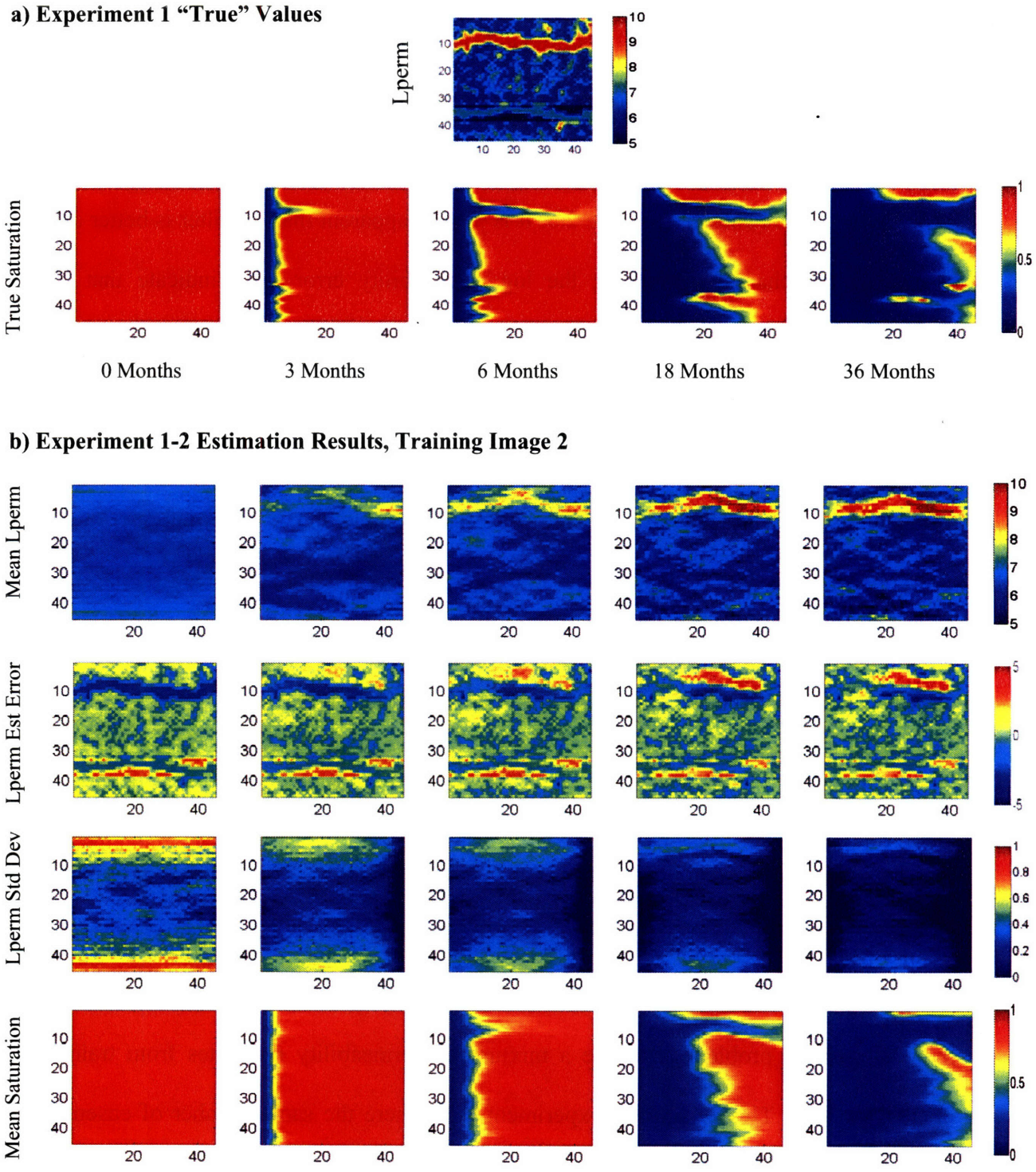
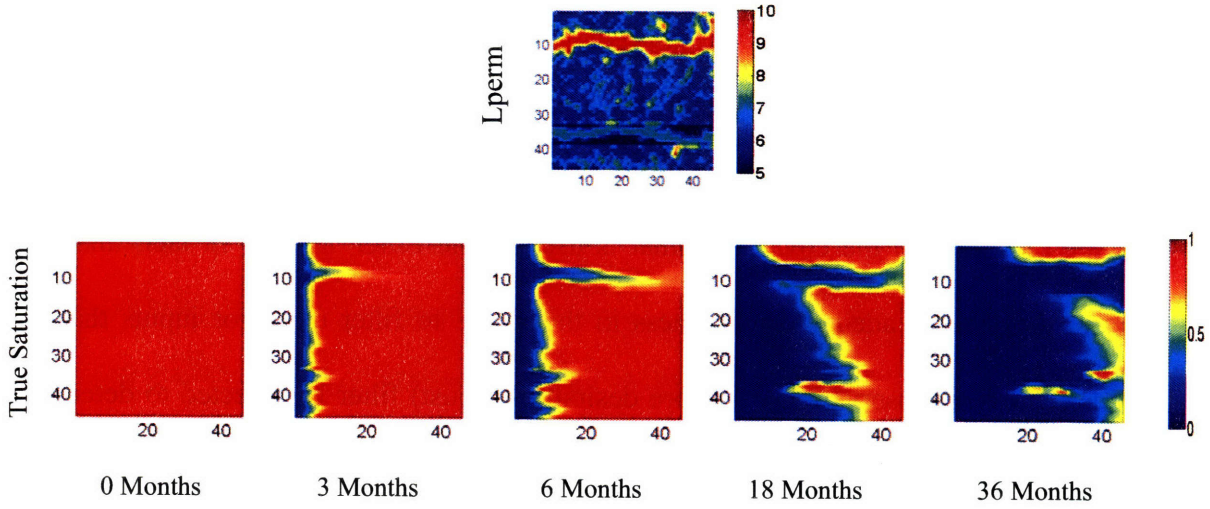


Figure 2.5 True log-permeability field in Experiment 1 (first row) and its corresponding saturation field (second row) for the injection/production scenarios shown in Figure 2.2; (b) Estimated mean (first row), error (second row), and standard deviation (third row) of the log-permeability field and estimated mean saturation field (fourth row) for Experiment 1-2 after the EnKF updates at 0, 3, 6, 18, and 36 months. The initial permeabilities were generated from Training Image 2 with narrow channels (Figure 2.3b)

Figure 2.6 shows the estimation results for experiment 1-3. The EnKF permeability estimates in this experiment have the same channel location as the true permeability; however, the channel width appears to be slightly overestimated. This is due to the effect of wide channels on the covariance that is used in the EnKF update. In fact, this may suggest the lack of adequate robustness in the EnKF to filter out prior inputs that are inconsistent with observations. A more desirable (robust) filter is expected to detect the correct channel width and location and filter out the incorrect samples.

Experiments 1-1, 1-2, and 1-3 suggest that the gradual degradation in the permeability estimates obtained by *Brouwer et al.* in [2.4] can be explained by ensemble selection. The replicates used to derive sample covariances in [2.4] and in our Experiment 1-1 do not adequately capture dominant features in the "true" log permeability field. This leads to incorrect updates that can cause the filter estimates to drift away from the true values, even as more measurements are added. When the replicates provide a better characterization of reality, as in Experiment 1-2 and 1-3, the ensemble Kalman filter performance is better.

a) Experiment 1 "True" Values



b) Experiment 1-3 Estimation Results, Training Image 1 and 2

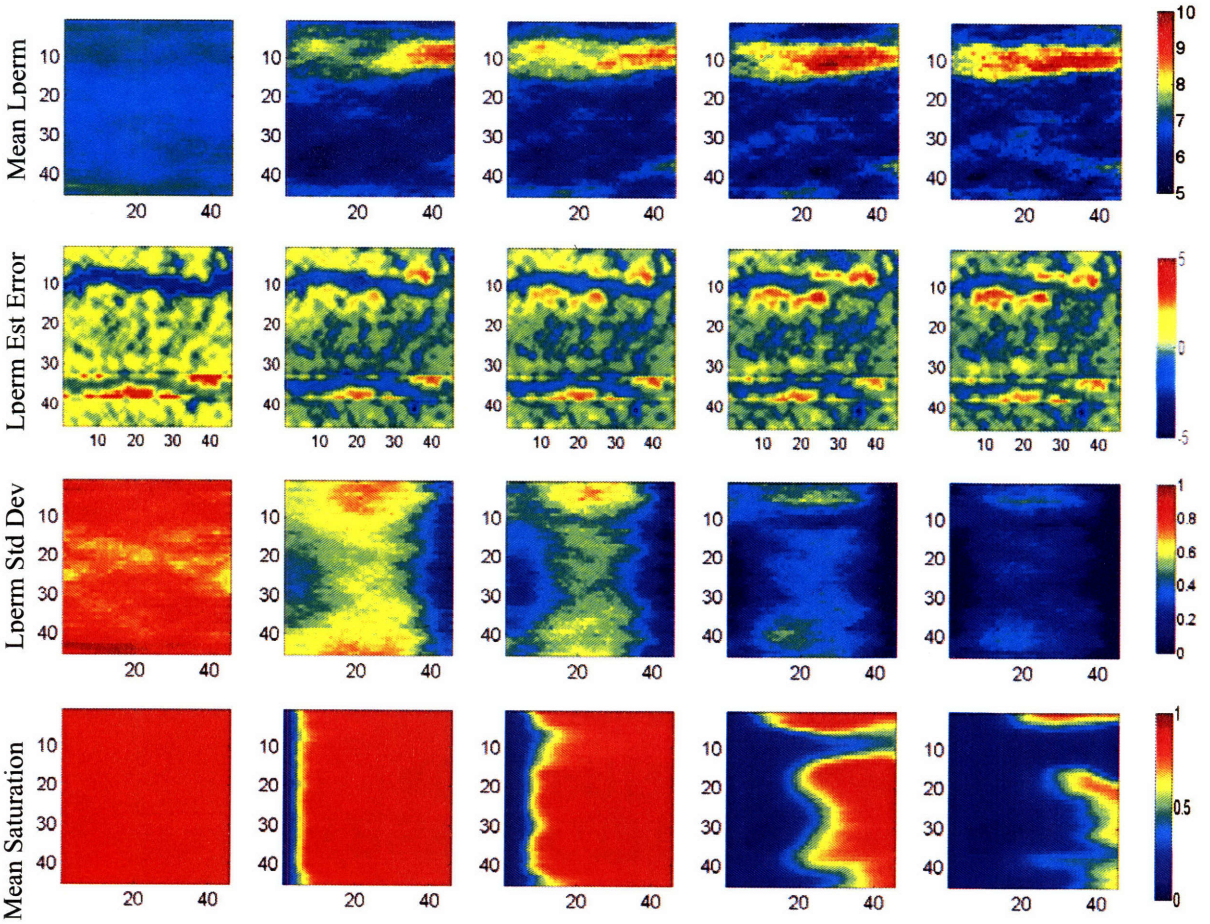


Figure 2.6 True log-permeability field in Experiment 1 (first row) and its corresponding saturation field (second row) for the injection/production scenarios shown in Figure 2.2; (b) Estimated mean (first row), error (second row), and standard deviation (third row) of the log-permeability field and estimated mean saturation field (fourth row) for Experiment 1-3 after the EnKF updates at 0, 3, 6, 18, and 36 months. The initial permeabilities were generated from Training Image 2 with narrow channels (Figure 2.3b)

Another contributing factor that could have caused the reported divergence of the EnKF in [2.4] may be associated with the size of the ensemble used. The sensitivity analysis in our study indicated that the ensemble size of 100 used in [2.4] does not provide a sufficiently accurate representation of the prior covariance used to derive updates. This further aggravates the covariances errors introduced by an inappropriate ensemble specification. Although larger ensemble sizes imply more computational effort, it may be argued that the increased effort is justified by the improved results. Also, the efficiency of the ensemble Kalman filter can be significantly improved, especially for larger problems, if parallel computation and more advanced solution procedures are adopted.

The improvement in the updated permeability enhances the predictive power of the reservoir model. This, in turn, can be used to improve decisions regarding well location or production strategies. Past studies have reported improved oil recovery when permeability estimation and optimal control approaches are combined [2.4, 2.5]. These studies have concluded that an approximate estimate of the permeability field (especially proper identification of low and high permeability features) gives significant improvement in the sweep efficiency. It is noteworthy that the degradation in log permeability estimates observed in *Brouwer et al.* [2.4] and in our Experiment 1-1 occurred at later update times after the most important control decisions had already been made. It would be useful to investigate to what extent these decisions will change when a more accurate estimate of the log permeability field is used to predict recovery.

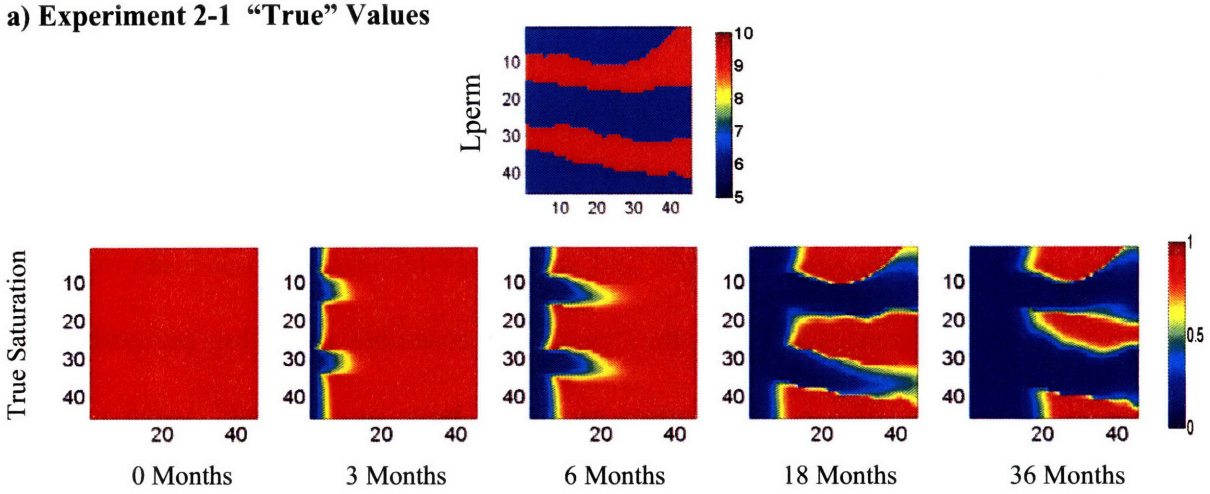
2.5.2 Experiment 2

In order to further investigate the connection between the log permeability ensemble and filter performance we consider another water flooding experiment that uses two "true" permeabilities drawn from an ensemble derived from Training Image 1 (Figure 2.3b).

Figures 2.7 and 2.8 summarize the results obtained for Experiments 2-1 and 2-2, which use the "true" permeability images shown in Figures 2.7a and 2.8a, respectively. These "true" images correspond to Realizations 9 and 22 of the Training Image 1 ensemble (this ensemble is drawn from the library used by *Sarma et al.* [2.5]). The channel geometries in these cases are somewhat more complex than in the Experiment 1 "true" image but the sandstone channels (10,000 mD) and background shale (500 mD) permeabilities are uniform rather than spatially variable. The estimation results summarized in Figures 2.7 and 2.8 confirm that the ensemble Kalman filter is able to identify the general structure of the log permeability field when the "true" permeability and ensemble are compatible. The filter captures most of the general features after the first two updates and there is no performance degradation with time. In fact, the quality of the estimates tends to improve until the end of the simulation (36 months) when the final solution is obtained. The log permeability estimates produced in Experiment 2-1 are somewhat better than in Experiment 2-2, probably because the image in Experiment 2-2 is more complex, especially in the center region that is further from the measurements. Experiments with other "true" log permeability images (not shown here) confirm that complex features located further from measurements are more difficult to identify.

In both "true" cases there are errors in the position of the estimated channel, similar to the errors observed in Experiment 1-2. These position offsets are evident in both the mean error and standard deviation plots, which mirror the geometry of the "true" channel. Channel position offsets result in high, and sometimes misleading, root-mean square errors since the differences between the "true" and ensemble mean log permeability values in the vicinity of a position offset can be as large as the "true" value in the channel. In this case, the root-mean squared errors on the edges of the channel can actually be higher if the channel is identified, but the estimated location is shifted slightly, than if the channel is missed altogether. The critical question here is "what error measure best reflects the quality of a permeability estimate in a channelized setting?". This is a topic that deserves further investigation.

a) Experiment 2-1 "True" Values



b) Experiment 2-1 Estimation Results, Training Image 1

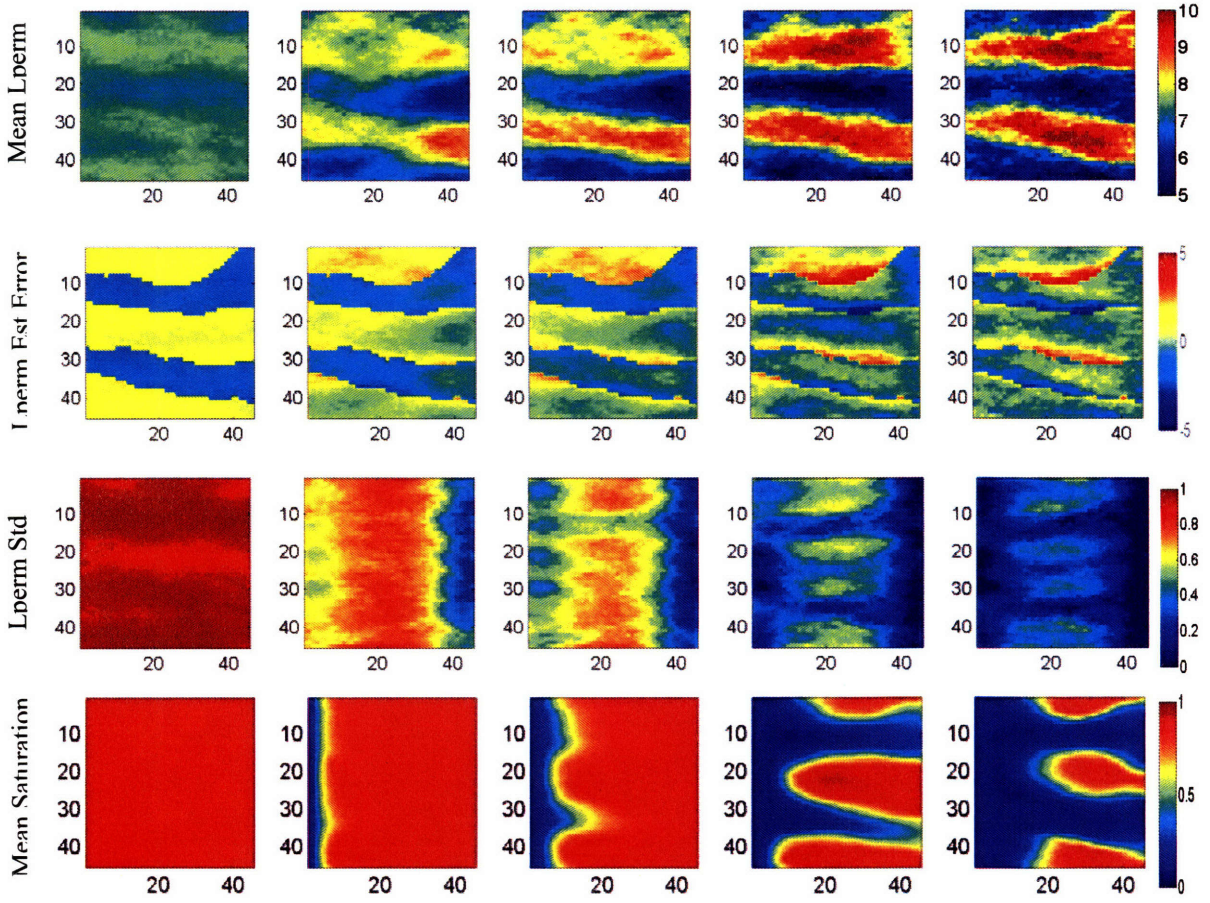
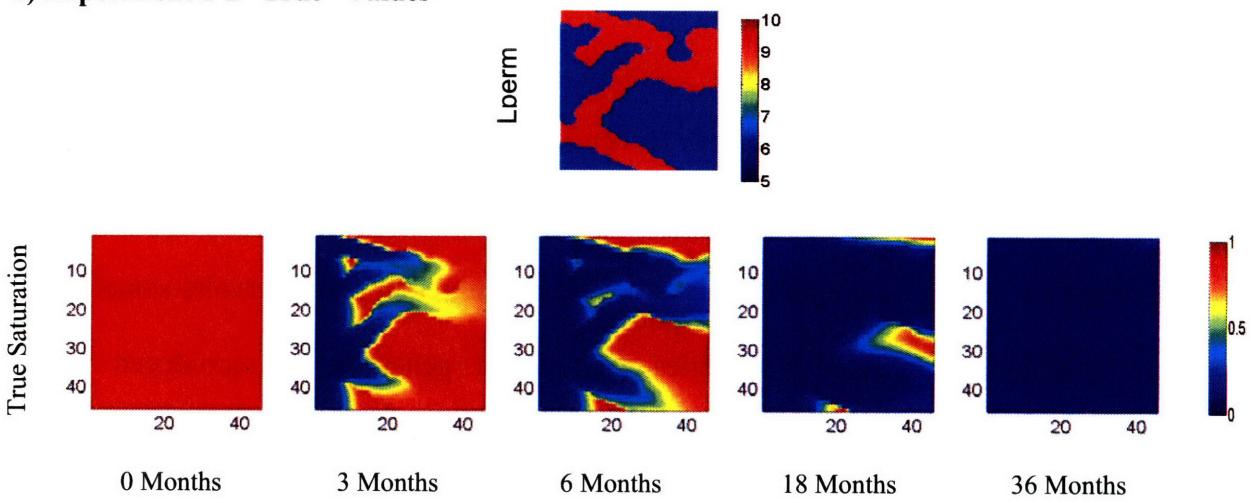


Figure 2.7 True log-permeability field in Experiment 2-1 (first row) and its corresponding saturation field (second row) for the injection/production scenarios shown in Figure 2.2; (b) The estimation mean (first row), error (second row), and standard deviation (third row) for the log-permeability field and the estimated mean saturation field (fourth row) for Experiment 2-1 after the EnKF updates at 0, 3, 6, 18, and 36 months. The initial permeabilities were generated from Training Image 1 with wide channels (Figure 2.3b)

a) Experiment 2-2 "True" Values



b) Experiment 2-2 Estimation Results, Training Image 1

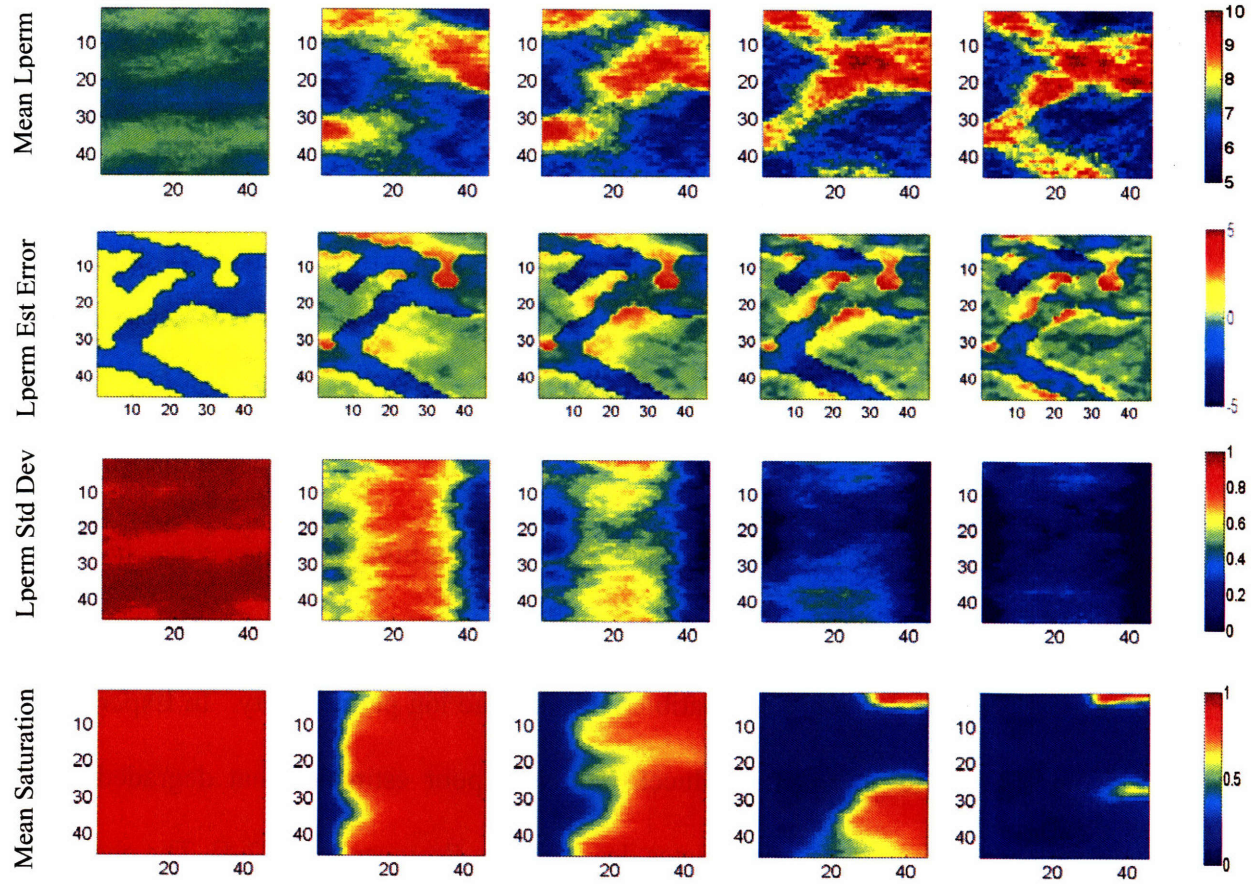


Figure 2.8 True log-permeability field in Experiment 2-2 (first row) and its corresponding saturation field (second row) for the injection/production scenarios shown in Figure 2.2; (b) The estimation mean (first row), error (second row), and standard deviation (third row) for the log-permeability field and the estimated mean saturation field (fourth row) for Experiment 2-2 after the EnKF updates at 0, 3, 6, 18, and 36 months. The initial permeabilities were generated from Training Image 1 with wide channels (Figure 2.3b).

2.6 Conclusions

The synthetic water flooding experiments conducted in this study support the view that proper definition of the permeability ensemble is essential to the success of the ensemble Kalman filter in reservoir applications. Geological features in real reservoirs sometimes have complex channelized geometries that cannot generally be reproduced with statistically homogeneous Markov random fields. Permeability generation methods based on training images and multi-point geostatistical methods seem to be better able to generate geologically credible realizations in channelized environments. The ensemble Kalman filter uses these realizations to derive sample covariances that provide approximate statistical descriptions of spatial variability. Our experiments indicate that the sample covariances, and the updates that depend on them, give better results when the underlying ensemble members have geometrical properties similar to the true permeability.

Experiment 1-1 shows that it is not enough to use training images that generate channels. The channels that appear in the ensemble replicates must have dimensions and other geometrical properties that are compatible with the true log permeability. In Experiment 1-1 a training image approach produced relatively poor estimates that degraded over time, in a manner similar to the estimates described by *Brouwer et al.* [2.4]. The channels in the ensemble replicates for this case were consistently too wide. In Experiment 1-2 the channels in the training image were narrowed to be closer in width to the channel included in the true permeability field. In this case, the prior information conveyed by the

ensemble was sufficiently accurate to give significantly better results. The dominant channel present in the "true" field was identified, although the position was slightly shifted.

The ensemble Kalman filter's ability to identify high permeability channels was confirmed in Experiment 2, where two different synthetic "true" channel configurations were considered. In both cases, the training image was compatible with the "true" image and the estimation results were encouraging. Features nearer the well measurements were generally estimated more accurately than features located further away.

The performance of the EnKF is also dependent upon the number of realizations used in the ensemble. In our example, a log permeability ensemble size of 100 was too small to give reliable results while an ensemble size of 300 seemed to be sufficient. The sample covariances that control the filter updating procedure can be expected to improve when the ensemble replicates are realistic and when the ensemble is large enough to provide an adequate characterization of uncertainty.

It is reasonable to ask whether we can expect an ensemble generated from a specified training image to include replicates that look like the unknown "true" log permeability field in a real application. Ensemble generation is definitely more challenging in a real application than it was in our Experiment 2, where the true permeability is one of the

replicates in the training image library. In more realistic situations it is important that the training image (or images) used for permeability estimation be derived from field data at the site of interest and that the image be sufficiently large, complex, and diverse to include all the features likely to occur at the site. At locations where the geology is highly uncertain the training image channels should vary in width, tortuosity, connectedness, and complexity, so that this uncertainty is reflected in the ensemble. Proper ensemble design is a critical part of the parameter estimation process. If the filter's ensemble reflects the likely range of true conditions the resulting estimates can be expected to be more accurate and robust. This is an important topic that deserves further investigation from both research and application perspectives.

References

2.1. Gavalas G.R., Shah P.C., Seinfeld J.H.: "Reservoir history matching by Bayesian estimation," *Soc. Petrol. Eng. J.* (1976) 16(6):337–350.

2.2. Oliver D.S., Reynolds A.C., Bi Z., Abacioglu Y.: "Integration of production data into reservoir models," *Petroleum Geosciences* (2001) vol. 7, no. SUPP, pp. 65-73(9)

2.3. Tarantola A., "*Inverse Problem Theory. Methods for Model Parameter Estimation,*" Philadelphia, PA: SIAM, 2004.

2.4. Brouwer D. R., Nævdal G., Jansen J.D., Vefring E.H., van Kruijsdijk C. P. J. W.: "Improved reservoir management through optimal control and continuous model updating," paper SPE 90149 presented at the SPE Annual Technical Conference and Exhibition, Houston, TX (2004).

2.5. Sarma P., Durlofsky L.J., Khalid A., Chen W.H.: "Efficient real-time reservoir management using adjoint-based optimal control and model updating," *Computational Geosciences*, (2006) vol. 10, pp. 3-36 .

2.6. Evensen G.: "The ensemble Kalman filter: theoretical formulation and practical implementation," *Ocean Dyn.* (2003) 53, 343– 367.

2.7. Whitaker J.S., Hamill T.M., "Ensemble data assimilation without perturbed observations," *Mon. Weather Rev.* (2002) 130: 1913–1924, 367.

2.8. Kalman R.E.: "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME*-Journal of Basic Engineering. (1960) vol. 82, pp. 35-45.

2.9. Evensen G.: "Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *J. Geophys. Res.* (1994) 99(C5): 10 143–10 162

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

2.11. Liu N., Oliver D.S.: "Ensemble Kalman filter for automatic history matching of geologic facies,": *Journal of Petroleum Science and Engineering* (2006) 47, 147-161.

2.12. Wen X.H., Chen W. "Real-time reservoir model updating using ensemble Kalman filter," SPE-92991. Proceedings of the 2005 SPE Reservoir Simulation Symposium, pp. 1–14.

2.13. Strebelle S., Journel A.: "Reservoir Modeling Using Multiple-Point Statistics," paper SPE-71324 presented at the 2001 SPE Annual Technical Conference and Exhibition, New Orleans, Sept. 30-Oct. 3.

2.14. Remy, N., 2004: S-GeMS, "A Geostatistical Earth Modeling Library and Software,": Ph.D. thesis, Stanford University.

2.15. ECLIPSE 100 (Black Oil): "Reference Manual and Technical Description," *Schlumberger GeoQuest*, Houston (2006).

2.16. Jafarpour B., McLaughlin D.: "Efficient Permeability Parameterization with the Discrete Cosine Transform," paper SPE-106453. Proceedings of SPE Reservoir Simulation Symposium, Febuary 2007.

2.17. Bishop CH, Etherton B.J., Majumdar S.J.: "Adaptive sampling with the ensemble transform Kalman filter,". part I. Theoretical aspects. *Mon. Weather Rev.* (2001) 129: 420–436.

2.18. Evensen G.: "Sampling Strategies and square root analysis schemes for the EnKF," *Ocean Dyn.* (2004) 54, 539– 560.

Chapter 3

Efficient Permeability Parameterization with the Discrete Cosine Transform

3.1 Summary

The inverse estimation (history matching) of permeability fields is commonly performed by replacing the original set of unknown spatially discretized permeabilities with a smaller (lower dimensional) group of unknowns that captures the most important features of the field. This makes the inverse problem better posed by reducing redundancy. The Karhunen-Loeve Transform (KLT) is a classical option for deriving low dimensional parameterizations for history matching applications. The KLT can provide an accurate characterization of complex permeability fields but it can be computationally demanding. In many respects this approach provides a benchmark that can be used to evaluate the performance of more computationally efficient alternatives. The KLT requires knowledge of the permeability covariance function and can give poor results when this matrix does not adequately describe the actual permeability field. By contrast, the Discrete Cosine Transform (DCT) provides a robust parameterization alternative that does not require

specification of covariances or other statistics. It is computationally efficient and in many cases is almost as accurate as the KLT. The DCT is able to accommodate prior information, if desired. Here we describe the DCT approach and compare its performance to the KLT for a set of geologically relevant examples.

3.2 Introduction

Reservoir characterization is generally based on localized borehole and outcrop observations that are interpolated to give regional descriptions of uncertain flow properties such as permeability. The interpolation process introduces uncertainty in the permeability field that translates directly into uncertainty about reservoir behavior. Incorporation of measurements of the dynamic variables during the production phase, i.e. history matching, provides a way to reduce permeability uncertainty. History matching identifies the permeability values that provide the best match, in terms of a specified performance measure, to observations of dynamic production variables such as bottom-hole pressure and fluid rates. This process can increase the accuracy and usefulness of model predictions if the estimated permeabilities provide a reasonable description of the true field.

It is generally accepted that history matching methods work best when they incorporate geologically realistic facies information. Realistic facies representations should account for depositional continuity and connectivity since these properties have a significant effect on fluid flow within the reservoir [3.1].

When the permeability field is characterized by finely discretized block values the history matching problem can be ill-posed and result in non-unique solutions [3.2, 3.3]. Ill-posed problems can produce reservoir models that honor observed measurements but provide incorrect predictions. Moreover, if estimated block permeabilities are not constrained to preserve facies connectivity, they may yield geologically inconsistent and unrealistic permeability fields. In order to deal with ill-posedness and to respect geological facies it is desirable to adopt a parametric description of permeability that is low-dimensional while also able to preserve important geological features and their connectivity.

Several parameterization approaches with varying complexity have been proposed and implemented for reservoir history matching problems. A simple zonation approach is used by [3.4] in which an aggregate of block properties are assembled and assigned a single value. Adaptive versions of this approach have been adopted to perform the history matching in multiple steps with increasing resolution [3.5, 3.6]. Other multi-resolution techniques have also been proposed for parameterization and history matching at different scales [3.7-3.9].

A particularly powerful parameterization approach suitable for history matching is the Karhunen-Loeve Transform (KLT), named after Karhunen [3.11] and Loeve [3.12]. This approach represents the permeability in any given block with a linear expansion (or

transform) composed of the weighted eigenvectors of a specified block permeability covariance matrix. This matrix can, in turn, be derived from a specified continuous permeability covariance function. In practice, the covariances used to derive the KLT basis functions are often derived from permeability measurements. When this is done the KLT is data-dependent (i.e. its characterization of permeability depends on correlation properties of a particular set of observations). This can be a benefit if the data in question are representative but can be a liability if they are not.

If the KLT weighting coefficients are properly selected, any given set of blocked permeabilities can be reproduced. When the actual permeability is uncertain the KLT coefficients can be treated as independent random variables with variances equal to the eigenvalues of the specified covariance matrix. These eigenvalues can be ranked from largest to smallest in an eigenvalue (or energy) spectrum that represents the relative impact of each term in the expansion. If the energy spectrum decreases sufficiently fast with rank, it is possible to capture much of the information in the permeability field with a truncated expansion that contains fewer terms than the number of permeability blocks. In this sense the KLT coefficients remaining after truncation provide a compressed representation of the original blocked field.

It can be shown that for a given/known image and a correct covariance matrix the KLT provides optimal compression, in the mean squared error (MSE) sense, among all linear

transforms [3.11, 3.14]. However, the KLT requires decomposition of large covariance matrices and can be computationally expensive.

The discrete cosine transform (DCT) is a computationally efficient alternative to the KLT. This approach uses a set of predefined basis functions that do not depend on the permeability covariance and do not need to be estimated from data. As a result, the DCT tends to be more robust with respect to errors in prior specification.

In the following sections we discuss in more detail the basis for the KLT and DCT parameterization methods and then illustrate how these methods may be applied to permeability estimation. Our emphasis is on the accuracy and robustness of the two methods. These methods can be used with a number of different history matching techniques, including sequential estimators such as Kalman filters and batch estimators such as nonlinear least-squares iterative search methods.

3.3 Using mathematical transforms for permeability parameterization

3.3.1 Transform-based image compression.

Linear transforms are operators that convert a function of an independent variable (such as position) to a related function of another independent variable (e.g. wave number). Transformed functions are often easier to work with than the original function [3.15]. For example, many data compression techniques rely on convenient properties of transformed data vectors [3.15, 3.16].

A general unitary (orthonormal) transformation of a one dimensional sequence $\{u(n),\ 0 \le n \le N\text{-}1\}$ can be expressed as convolution of $u(n)$ with a specified function $a(k,n)$ [3.16]:

$$v(k) = < u(n), a(k,n) > = \sum_{n=0}^{N-1} a(k,n).u(n) \quad \text{for} \quad 0 \le k \le N-1 \qquad (3.1)$$

The original sequence can be reconstructed by applying the inverse transform $a^*(k,n)$ to the transform coefficients:

$$u(n) = < a^*(k,n), v(k) > = \sum_{k=0}^{N-1} v(k).a^*(k,n) \qquad \text{for} \quad 0 \le n \le N-1 \qquad (3.2)$$

where $\mathbf{a}^*_k \equiv \{a^*(k,n),\ 0 \le n \le N\text{-}1\}^T$ is a set of basis vectors. Each term of the basis function expansion defines a particular mode of the original function. In matrix form, the transform and its inverse can be represented as ($\mathbf{A}^{-1} = \mathbf{A}^{*T}$):

$$\mathbf{v} = \mathbf{A}\mathbf{u} \Rightarrow \mathbf{u} = \mathbf{A}^{\mathrm{T}}\mathbf{v} \tag{3.3}$$

Extension of these equations to higher dimensions is straightforward [3.16]. Unitary transformations have several desirable properties including: signal energy conservation, energy compaction, decorrelation, entropy (information) preservation [3.16].

It is often possible to construct a good approximation to $u(n)$ with a truncated version of the inverse transform. In particular, suppose that we compute and retain only the first $K << N$ basis function expansion terms (or modes) of $u(n)$:

$$v(k) = \sum_{n=0}^{N} a(k,n).u(n) \qquad \text{for} \qquad 0 \le k \le K-1 \tag{3.4}$$

$$u(n) \approx \sum_{k=0}^{K-1} v(k).a^{*}(k,n) \qquad \text{for} \qquad 0 \le n \le N-1 \tag{3.5}$$

If the terms omitted from the $u(n)$ expansion have small coefficients, they make a small contribution to the signal variability and energy. In this case, the retained basis function coefficients provide an efficient low-dimensional (compressed) approximate representation of the original signal. The truncated expansion removes the redundancy present in signals (such as permeability) that exhibit significant correlation over space [3.16].

3.3.2 Compression of known vs. uncertain datasets.

In image processing, a transform is applied to a single known image of arbitrary structure. The truncated basis coefficient vector *v(k)* provides a compressed version of *u(n)* that requires less transmission time and storage than the original image [3.15, 3.16]. In this case it is only necessary to find a fixed basis that provides a good compression ratio while producing a satisfactory approximation of the particular image to be compressed [3.14-3.17]. This approach is appropriate in reservoir applications when we seek to produce an efficient compressed representation of a known permeability field.

The situation is somewhat different if we wish to find a single set of basis functions that can be used to compress any image contained in a specified "training set" composed of many images that share certain common features. This is the situation that arises in history matching, where we do not know the permeability field in advance, although we may believe that this field (or image) has the same general features as those included in the training set. In this case, it is beneficial to incorporate relevant prior knowledge about these common features when selecting basis functions for a parameterization. Bases derived from a training set can provide more efficient parameterizations for history matching problems but are not generally robust enough to handle images that differ from those included in the training set. That is, there is a trade-off between robustness and optimality when selecting a parameterization approach for history matching.

There are some important differences between the way truncation is performed when compressing individual images vs. when deriving efficient parameterizations for history matching. In the first case the most important modes (and basis functions) are those with the largest $v(k)$ coefficients. When the image is not known in advance (as is the case in history matching problems), it is not possible to compute these $v(k)$ coefficients. In this case we can only identify the modes (and basis functions) that are most important for the training set as a whole. These will generally be different than the modes selected for any given image. The examples described in the following sections explore this difference in more detail.

3.4 Application of the KLT and DCT Parameterization Methods

3.4.1 Permeability model.

In this section we use some simple two-dimensional examples to illustrate how the KLT and DCT methods can provide low-dimensional geologically realistic permeability field parameterizations suitable for history matching. The geologic features common to all members of the training set are defined by the channelized training image shown in Figure 3.1a. This training image has 250 × 250 × 1 pixels and comprises two facies types: low permeability background shale and high permeability sandstone channels. The background shale permeability was assumed to be 500md while the embedded high-permeability sandstone channels had a permeability value of 10,000md.

The 5000 permeability realizations included in the training set were generated from the training image of Figure 3.1a with the multiple-point geostatistical algorithm *snesim* [3.1]. Each realization is discretized over a grid of 64 × 64 × 1 (640m × 640m × 10m) grid block. Figure 3.1b shows nine realizations from the training set. The varying shape and geometry of the

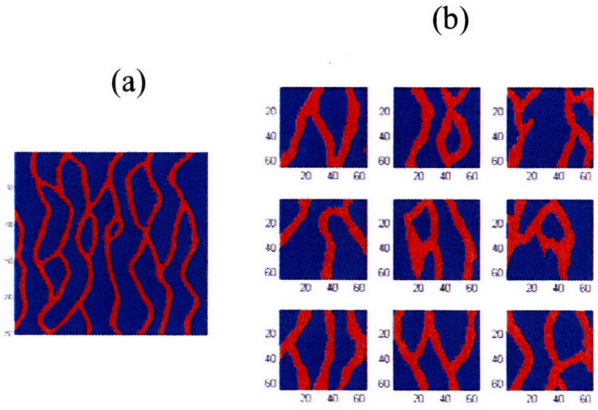


Figure 3.1 Training image (a) and nine sample realizations (b) generated from it using pattern-based multiple point geostatistics. Red channels represent fluvial depositions with high permeability embedded in background shale formation (blue).

channels in these realizations are the major sources of uncertainty in the unknown permeability distribution. The highly structured nature of the training set images reflects a high level of redundancy, suggesting that the field could be represented much more efficiently if the blocked values were transformed.

It is possible to have negative permeability values in the approximated permeability field when truncated KLT or DCT are used (this applies to history matching too). While the

logarithmic permeability transform is a convenient way of avoiding this, it can result in asymmetrical exponential error growth when the log permeability is transformed back to permeability. To overcome this, we used the inverse error function transform of permeability field with a rescaling. Appendix A gives a brief account of the potential pitfall in using the logarithm of the permeability field and describes a few alternatives.

In what follows, we use the term "leading modes" to refer to the modes that provide a reasonable approximation of an *ensemble* of images. The term "significant modes" is used to refer to the modes corresponding to largest transformation coefficients for a *single known* image, i.e. $v(k)$ in equation (3.4).

3.4.2 The Karhunen-Loeve Transform (KLT).

The basis functions of the spatially discretized KLT are the eigen-vectors of the permeability covariance matrix. In our application this covariance is derived from the 5000 blocked permeability realizations in the training set. In the KLT approach the covariance matrix provides the prior information used to concisely describe the features common to the members of the training set. In general, the KLT is a second order characterization that ignores the information contained in the higher moments of the permeability field. Nevertheless, it is widely used, especially when the permeability field is Gaussian and the first two statistical moments provide a complete characterization.

For a zero-mean and finite variance random vector **u** of dimension N the basis vectors ϕ_k are defined by the eigenvalue equation [3.16]:

$$\mathbf{C}\phi_k = \lambda_k \phi_k \quad \text{for} \quad 0 \le \mathrm{k} \le N-1 \tag{3.6}$$

Here, λ_k are the eigen-values corresponding to the eigenvectors ϕ_k. The KLT of **u** can then be written as:

$$\mathbf{v} = \mathbf{\Phi}^{*T}\mathbf{u} \tag{3.7}$$

and the inverse transform is given by:

$$\mathbf{u} = \mathbf{\Phi}\mathbf{v} = \sum_{k=0}^{N-1} v(k)\phi_k \tag{3.8}$$

where ϕ_k is the k^{th} column of $\mathbf{\Phi}$. If **u** is a random vector of uncertain permeabilities, the KLT basis function coefficients are independent random variables with variances equal to the corresponding eigenvalues. The largest eigenvalues are associated with the most variable (highest energy) coefficients.

$$\mathbf{\Phi}^{*T}\mathbf{C}\mathbf{\Phi} = \Lambda = diag\{\lambda_k\} \tag{3.9}$$

When all of the N eigenfunctions (or modes) are retained the KLT provides a lossless (error-free) representation of the original permeability field. A more concise approximation is obtained when the expansion is truncated. The eigen-spectrum of the covariance matrix is a plot of the ordered eigenvalues (the elements of $\boldsymbol{\lambda}$, ordered from the largest to smallest value) vs. the rank. The spectrum obtained from the first (leading) 1000 modes of the training set covariance matrix for our example is shown in Figure 3.2a. Figure 3.2b shows the fraction of explained variance (energy) for the same leading 1000 modes and Figure 3.2c shows the eigenvectors for the leading 64 modes (the elements of each eigenvector are associated with the appropriate blocks in the spatially discretized image).

The steep slope of the eigen-spectrum (usually observed in correlated fields) suggests that most of the variability (energy) in the permeability field can be described by a few leading eigenvectors. It should, however, be noted that the leading modes of the covariance do not correspond to the largest coefficients (significant modes) of any single realization (or image). The spectrum for any given realization is constructed from the known single image coefficient vector $\mathbf{v}$ rather than the eigenvalue vector $\boldsymbol{\lambda}$. That is, the ranked eigenvalues in the spectrum of Figure 3.2 identify the eigenvectors that capture the most important features over all realizations in the training set but not necessarily the most important features of any given realization.

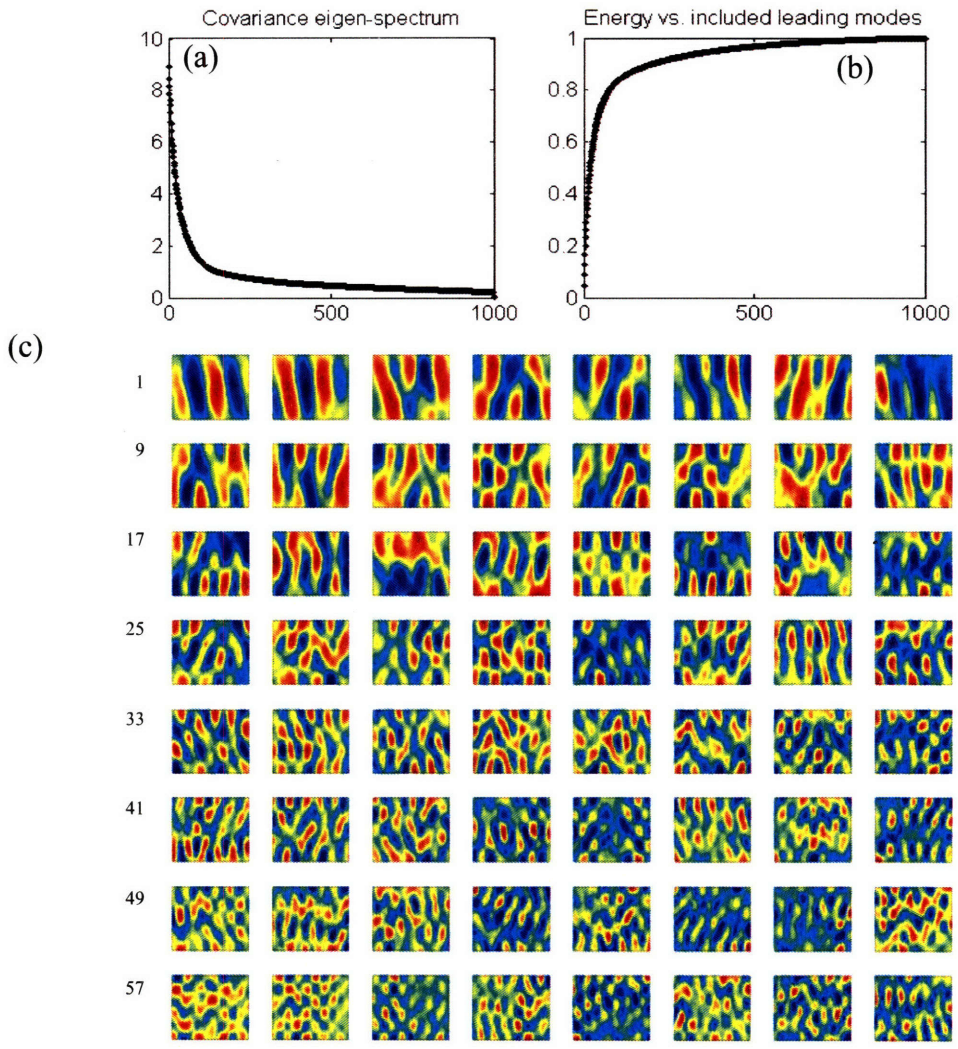


Figure 3.2 (a) Eigen-spectrum of the permeability covariance matrix generated from 5000 permeability realizations. (b) Variance compaction in the first 1000 leading modes of the covariance matrix. (c) The first 64 leading modes of the covariance matrix.

The difference between truncating the spectrum for the entire training set vs. the spectrum for an individual replicate is illustrated in Figure 3.3. In this figure, five sample realizations are shown with their KL expansion coefficients corresponding to the first 200 leading modes of the covariance matrix (second row). Notice that these coefficients do not follow the same order as the leading modes. Two approximations are presented in rows four and

six. Rows three and five show the modes that were included in their expansion, respectively. The first approximation (rows three and four) are obtained by selecting the largest 55 expansion coefficients and their corresponding eigenvectors. Note the difference between the significant coefficients of individual samples that determine their corresponding significant modes. If these optimal image-specific modes are selected, the corresponding approximation has minimum root mean square error (RMSE) of all linear transforms for a fixed number of modes [3.10, 3.11, 3.14]. Of course, this approximation is only possible if the sample permeabilities are known a priori and a correct covariance matrix is available.

When the permeabilities are not known in advance the significant modes for individual realizations (images) can not be predetermined. However, an approximation can be obtained (rows five and six) by using the first 55 leading modes of the covariance matrix and solving for the corresponding coefficients in history matching. In this case, some of the coefficients that are significant for the individual images are truncated and the approximation is suboptimal. It is noteworthy that the mean of the expansion coefficients over the 5000 realizations is similar to the eigen-spectrum shown in Figure 3.2.

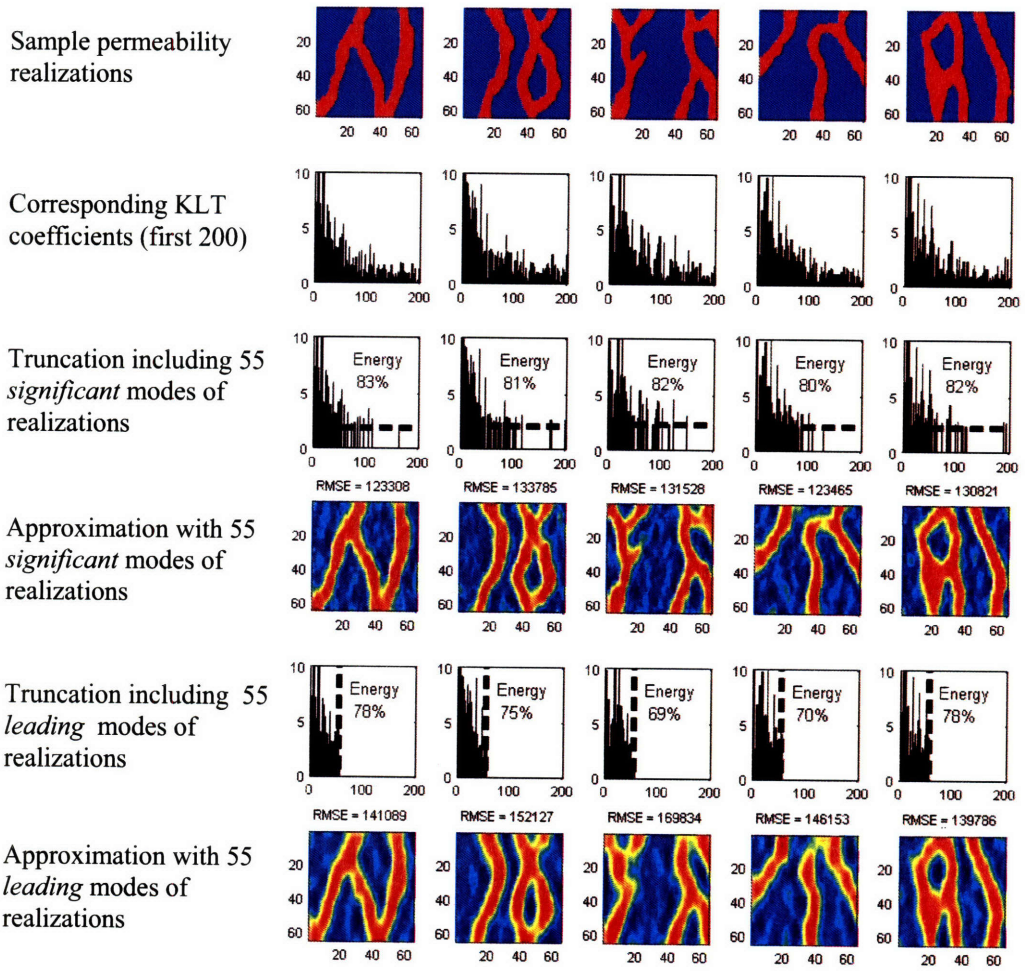


Figure 3.3 Permeability parameterization using the Karhunen-Loeve transform for a priori known and unknown permeabilities generated from the same training image.

The KLT has at least three major drawbacks that have restricted its application to data compression problems [3.15, 3.16] and make it problematic for history matching problems. First, it requires an expensive singular value decomposition (SVD) operation, which has a computational complexity of order $O(N^3)$ [3.16] and becomes prohibitive for large problems. Second, construction of its basis vectors requires correct specification of a covariance matrix derived from a particular training image. This renders the KLT

unsuitable for compression of arbitrary fields that are not well characterized by a training image. Finally, the KLT provides only a second order (covariance-based) characterization that ignores all higher moments.

3.4.3 The Discrete Cosine Transform (DCT).

The KLT's dependence on a specified covariance or training set and its high computational cost have discouraged widespread use for data compression. The DCT provides an attractive alternative that is faster and requires fewer assumptions [3.19]. DCT bases have been shown to asymptotically converge to KLT bases for first order stationary Markov processes [3.16]. The DCT has proven effective in many pattern recognition and image compression applications [3.15-3.17].

The transformation kernels used in the DCT are real cosine functions. Using a Fast Fourier Transform (FFT) [3.20], the DCT can be computed in $O(N \log_2 N)$ operations [3.16, 3.21, 3.22]. This is much more computationally efficient than the KLT, which requires a singular value decomposition of order $O(N^3)$ [3.16]. A comparison between DCT and other Fourier-related transforms is given in [3.23]. The details of the development, derivation, and properties of DCT and its relation to KLT can be found in [3.17].

A variety of predetermined similar transforms such as DFT, Walsh-Hadamard, and Haar transforms have also been applied for image compression. However, DCT has been shown to be superior to all of these for compression purposes [3.15-3.17]. Since the DCT basis vectors are prespecified and data-independent they only need to be computed and stored once. The orthogonality of the DCT basis functions facilitates computation of the inverse transform. Since the transform is separable it can process a multi-dimensional signal one dimension at a time [3.15-3.17]. Finally, large signals to be transformed with the DCT can be segmented to avoid large matrix manipulations. These attractive compression properties have promoted the use of DCT in JPEC and MPEG compression standards [3.15-3.17].

The discrete one dimensional DCT of a signal *u(n)* of length *N* has the following form [3.16]:

$$v(k) = \alpha(k)\sum_{n=0}^{N-1} u(n).\cos\left[\frac{\pi(2n+1)k}{2N}\right] \qquad 0 \le k \le N-1 \tag{3.10}$$

where $\alpha(k)$ is defined below:

$$\alpha(k) \equiv \begin{cases} \sqrt{\frac{2}{N}} & k = 0 \\ \sqrt{\frac{1}{N}} & 1 \le k \le N-1 \end{cases} \tag{3.11}$$

The inverse DCT is:

$$u(n) = \sum_{k=0}^{N-1} \alpha(k) v(k) . \cos\left[\frac{\pi(2n+1)k}{2N} \right] \quad 0 \le n \le N-1 \qquad (3.12)$$

Extension of the above equations to higher dimensions is given in [3.17]. However, the separability property of DCT can be exploited to achieve computational savings by applying the one dimensional transform in each direction [3.15-3.17]. In image compression, it is common to apply a two dimensional DCT of size 8 to separate 8 × 8 image segments [3.15-3.17]. Figure 3.4a shows the bases used for a 64-by-64 image without segmentation. The basis functions are arranged according to their orientation and level of detail (frequency content) in a descending order from upper left to lower right. Depending on the desired level of details in the approximation, more high frequency components (lower right modes) are included.

Figure 3.4b shows the original image (first column), all of the DCT coefficients for this image, using the same ordering convention as in Figure 3.4a (second column), the DCT coefficient spectrum ordered from largest to smallest coefficients (third column), and the RMSE between the true and approximate (truncated) images as a function of number of retained modes. Figure 3.4c illustrates how the DCT can be used to compress a known permeability realization. It shows the largest 21, 55, 105, 465 coefficients and their corresponding approximations. As expected, the DCT approximation is slightly less accurate than the KLT (the RMSE for KLT with 55 modes for this realization is 123308

(row four of the first column in Figure 3.3) while the RMSE for DCT with 55 modes is 130156 (row two of second column in Figure 3.4c). Notice the concentration of the large coefficients on the top left corner in the first row of Figure 3.4c. This clustering of coefficients generally corresponds to the modes with large scale variations in the horizontal, vertical, and diagonal directions. For our image, which has vertically dominant features, the vertical modes of DCT are more significant. This has important implications for the selection of retained modes, as discussed below.

In history matching applications the permeability field and its DCT coefficients are unknown, so it is not possible to identify the retained modes by ordering the coefficients. In the absence of any prior information, a reasonable orientation insensitive alternative is to retain modes associated with coefficients inside a diagonally symmetric triangle in the top left corner as shown in the third row of Figure 3.5. This *zonal* selection of coefficients provides a robust but suboptimal approximation that can be quite useful when there are no directional preferences. The power of this approach is better appreciated by noting that there is no need to use a training set or covariance to characterize the set of possible images, as is required with the KLT.

The triangular screening method for truncating the DCT can readily incorporate qualitative knowledge about the orientation of the channels. For example, if the dominant features of the field are expected to be vertically oriented more coefficients should be selected from the left side of DCT coefficients array. If the dominant features are horizontally-oriented

more coefficients should be selected from the top of the DCT coefficient array. When quantitative prior information is available (such as an image library) proper coefficients can be selected more systematically.

The fifth row of Figure 3.5 shows a different approach that uses an image library to train the screening method. This approach selects the 55 modes with largest (in magnitude) coefficient means, averaged over all the images in the training set (i.e. leading modes of the set). This trained screening approach is similar to the KLT in that it gives the best representation of all permeabilities in an average sense. It provides similar performance to the KLT with outputs that are generally smooth and preserve local connectivity due to cosine bases.

3.4.4 Sensitivity to prior specification.

This section uses an example to investigate the sensitivity of the KLT, untrained DCT, and trained DCT parameterization methods to errors in prior information. The first row in Figure 3.6 shows three reference permeabilities A, B, and C (from left to right) that are to be approximated using these methods. Permeability field A comes from the training image shown in Figure 3.1 and has mainly vertical channels. Permeability fields B and C do not belong to the training set derived from this vertically dominant image. The second through fourth rows show, respectively, the truncated approximate permeability fields obtained from the KLT, untrained DCT, and trained DCT transforms. The columns in these rows show the results obtained with 55, 210, and 465 leading modes, for each of the three

reference permeabilities. In all cases the leading modes of the KLT and trained DCT modes are obtained from the vertically oriented training image.

In the first case (left), when the correct training image is used to represent the reference image (A), the KLT and trained DCT results are better than the untrained DCT results. This reflects the benefits of the additional prior information provided by the training set. However, in the second and third cases (middle and right), when the training image does not reflect the dominant vertical properties of the true permeability field, the KLT and trained DCT representations are poor. As the deviation form the correct training image increases (from B to C), the results become worse. The trained bases require many more modes to give a reasonable approximation and therefore lose their compression power. It is noteworthy that perfect reconstruction of the original permeability field is possible if there is no truncation, even if the training image is misleading. Sensitivity to the training image is a result of using truncation (trained by training image) to compress the image.

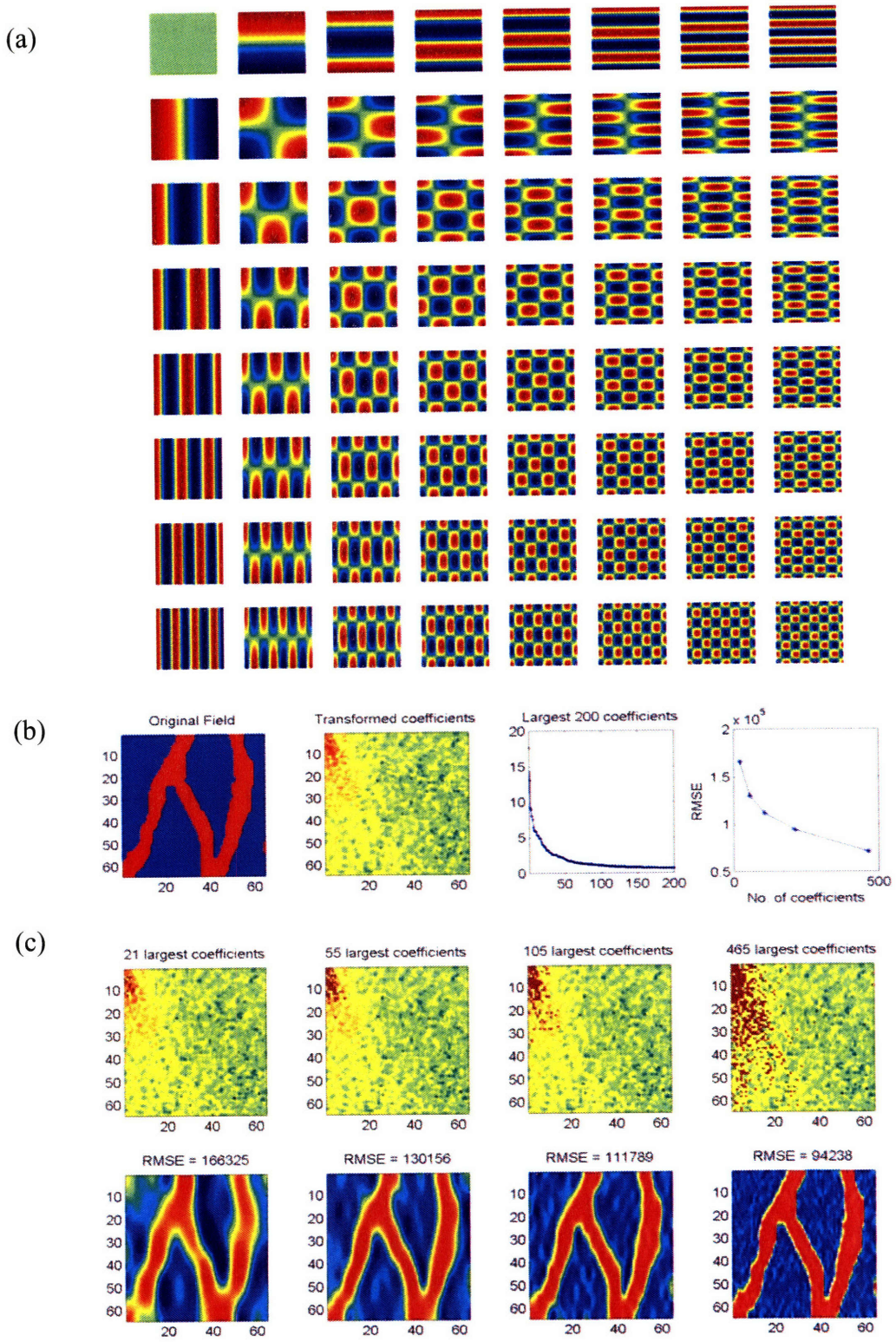


Figure 3.4 (a) The discrete cosine transform bases (modes) for 8 × 8 image representation. (b) A sample realization (first column) with its DCT transform (second column), sorted DCT coefficients (third column) and the RMSE of approximation with increasing number of modes (last column). (c) The retained significant coefficients (first row) and the corresponding reconstruction (second row) for increasing number of modes.

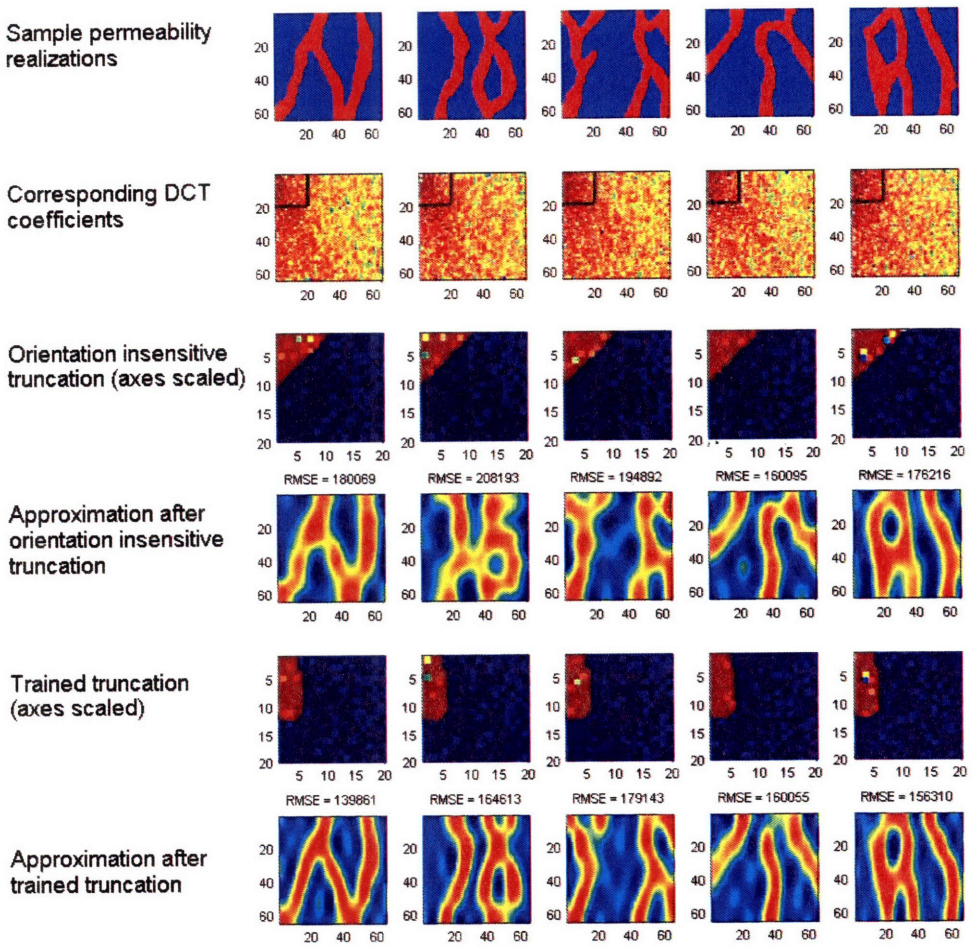


Figure 3.5 Permeability parameterization using the DCT with untrained and trained bases.

The performance of the untrained DCT is not affected by the orientation of the channels in the true permeability field. This is due to the orientation-insensitive screening method used to select the retained DCT coefficients. As can be seen from Figure 3.6 the untrained DCT approximations are not as good as the KLT and the trained DCT approximations when the correct training image is used. The untrained DCT effectively sacrifices accuracy for

robustness. Robustness is very advantageous when the training image is unavailable or uncertain. This situation is often the case in practice where a training image or image library must be constructed from sparse measurements.

3.5 Conclusions

The analyses and examples described in this chapter indicate that the discrete cosine transform is an attractive option for parameterizing permeability distributions. Although the origins of the method lie in image compression, it is well-suited for parameterization of ill-posed estimation (or history matching) problems. The coefficients of an appropriately truncated DCT are the parameters to be estimated in the history matching procedure and the complete blocked permeability field can be reconstructed from the inverse DCT. The DCT parameterization approach is robust and computationally efficient and has the flexibility of either including or not including prior information, through coefficient screening.

The DCT provides accuracy comparable to the Karhunen-Loeve transform, which is known to give the smallest possible MSE for a given number of retained modes and a correctly specified covariance. DCT approximations (compressed images) look very much like their KLT counterparts when both are trained from the same image library. However, the DCT is much more computationally efficient and much better able to perform adequately when prior information is unavailable or highly uncertain. In such cases, it is better to use the

untrained version of the DCT since the trained DCT does not do much better than the KLT if it is given an incorrect prior. Overall, the DCT appears to offer the best parameterization option for reservoir history matching applications, where flexibility, robustness, computational efficiency, and the ability to properly capture facies structure and connectivity are all important.

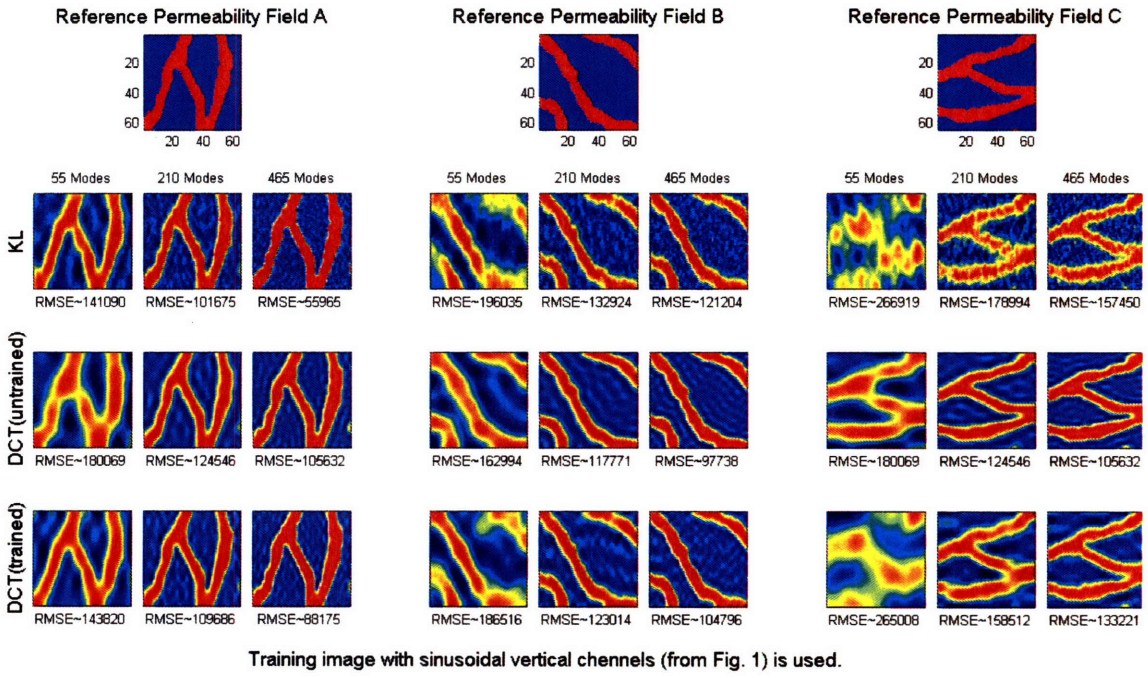


Figure 3.6 Sensitivity of the transforms to errors in the training image specification. Trained screening approaches (i.e. KLT and trained DCT) show strong sensitivity to deviations from the training image and provide a poor representation of permeabilities that do not belong to the training image.

References

3.1. Strebelle, S. and Journel, A.G.: " Reservoir Modeling Using Multiple-Point Statistics," paper SPE 71324 presented at the 2001 SPE Annual Technical Conference and Exhibition, New Orleans, 30 September–3 October.

3.2. Gavalas, G. R., Shah, P. C. and Seinfeld, J. H.: "Reservoir History Matching by Bayesian Estimation". *SPEJ* (Dec. 1976).

3.3. McLaughlin D., Townley L.R.: "A Reassessment of the Groundwater Inverse Problem". *Water Resour. Res.*, 32(5):1131-1161, 1996.

3.4. Jacquard P., Jain C.: "Permeability Distribution From Field Pressure Data". *Soc. Pet. Eng. Journal* (December 1965) 281-294.

3.5. Hans O. J. : "A rapid method for obtaining a two-dimensional reservoir description from well pressure response data". *Soc. Petrol. Eng. J.*, 6(12):315–327, 1966.

3.6. Aanonsen S.I.: "Efficient History Matching Using a Multiscale Technique". proc. 2005 SPE Reservoir Simulation Symposium, Houston, 31 Jan – 2 Feb 2005, Paper *SPE* 92758.

3.7. Grimstad A. A., Mannseth T., Nævdal G., Urkedal H.: "Adaptive multiscale permeability estimation". *Computational Geosciences* 7 (2003), 1-25

3.8. Sahni, I., and Horne R.N.: "Multiresolution Wavelet Analysis for Improved Reservoir Description," *SPEREE* (February 2005), 8(1), 53-69.

3.9. Shah P.C., Gavalas G.R., Seinfeld J.H.: "Error Analysis in History Matching: The Optimal Level of Parametrization". *Soc. Petroleum EnginemJ.,* June 1978, 219-228.

3.10. Reynolds A. C., He N., Chu L., Oliver D. S.: "Reparameterization techniques for generating reservoir descriptions conditioned to variograms and well-test pressure data". *Soc. Petrol. Eng. J.,* 1(4), 413--426, 1996.

3.11. Karhunen K.: "Ueber lincare Methoden in der Wahrscheinlichts-keitsrechnung". *Ann. Acad Sci. Fenn. Ser.* A. I. 37 (1947) 3-79.

3.12. Loeve M. M.: "Probability Theory". New York, NY: Springer-Verlag, 1978. 2 Vols., Fourth Edition.

3.13. Andrews H.C.: "Multidimensional Rotations in Feature Selection". *Trans. Comput.* ,20(9):1045-1051, Sept. 1971.

3.14. Pearl J., Andrews H., Pratt W.: "Performance Measures for Transform Data Coding". *Communications, IEEE Transactions on (legacy, pre – 1988)*, vol.20, no.3pp. 411-415, Jun 1972.

3.15. Gonzalez, R. C., Woods, R. E.: "*Digital Image Processing*". (2002).2nd ed., Prentice Hall, Upper Saddle River, NJ.

3.16. Jain A. K.: *"Fundamentals of Digital Image Processing"*, Prentice Hall, 1989.

3.17. Rao K. R., Yip P.: "Discrete Cosine Transform"*: Algorithms, Advantages, Applications* (Academic Press, Boston, 1990).

3.18. Kruger H., Mannseth T.: "Extension of the parameterization choices in adaptive multiscale estimation". *Inverse Problems in Science and Engineering* 13 (2005), 469-484.

3.19. Ahmed N., Natarajan T., Rao K. R.: "Discrete Cosine Transform". *IEEE Trans. Computers*, 90-93, Jan 1974.

3.20. Brigham E. O.: "The Fast Fourier Transform and Its Applications", 1988, Englewood Cliffs, NJ: Prentice-Hall, Inc., 448 pp.

3.21. Feig E., Winograd S.: "Fast algorithms for the discrete cosine transform". *IEEE Transactions on Signal Processing* 40 (9), 2174-2193 (1992).

3.22. Narasimha M.J., Peterson A.M. : "On the computation of discrete cosine transform". *IEEE Trans. Comm.* (1978) COM-26, 934-936.

3.23. Ersoy O. K.: "A comparative review of real and complex Fourier-related transforms". *IEEE Proceedings,,* Vol. 82, No. 3, pp. 429-447, March 1994.

3.24. Freeze, R.: "Astochastic-conceptual analysis of one-dimensional groundwater flow in non-uniform homogeneous media". *Water Resources Research*, (1975) 11(5), 725–741.

3.25. Hoeksema R., Kitanidis P.: "Analysis of the spatial structure of selected aquifers". *Water Resources Research*, (1985) 21(4), 563–572.

Appendix 3A. The Log Transform Pitfall

In subsurface hydrology and petroleum engineering literature the logarithm of the permeability is generally observed to follow a normal distribution [3.24, 3.25]. Field observations tend to support this assumption. Moreover, history matching procedures that adjust log permeabilities cannot produce unphysical negative estimates. While the logarithmic permeability transform is convenient in some respects, it can result in asymmetrical exponential error growth when the log permeability is transformed back to permeability. Reporting parameter estimation results on a logarithmic scale can be misleading since this can mask the magnitude of error in the original parameter.

The error amplification properties of the logarithmic transformation are particularly problematic when this transformation is used in combination with a truncated parameterization or image compression techniques such as the DCT or KLT. While the approximate log-permeability obtained after compression may seem quite satisfactory, the associated original permeability obtained by exponentiating it may be a poor approximation. Since the permeability rather than log permeability is used in the reservoir simulator, the truncated log permeability parameterization can have an adverse impact on the accuracy of the resulting pressure and saturation solutions.

It is possible to avoid the disadvantages of the log permeability transform while still insuring that the permeabilities produced by a history matching procedure are positive. One

approach is to use alternative transforms that do not exhibit exponential error growth. The other is to constrain the permeability values within their physical bounds in the original domain while applying the transform (constrained optimization). This is achieved by finding the transform coefficients that result in positive permeability values while minimizing a performance measure (L_2 or L_∞ norms of errors for instance).

Figure 3.A-1 shows a simple example that compares the performance of different methods under truncated KLT with the first 40 leading modes. The first column shows the original permeability field (a_1) and its KLT approximation (a_2). Negative permeability values were obtained after the approximation (a_2) that motivated the use of alternative approaches. Three different transformations, i.e. inverse error function (b_1-b_2), logarithm (c_1-c_2), and tangent (d_1-d_2), were applied separately prior to KLT transformation. Two constrained optimization results that minimized the L_2 and L_∞ norms of the deviations of the approximation from the true permeability field were also included (e_1-e_2 and f_1-f_2, respectively).

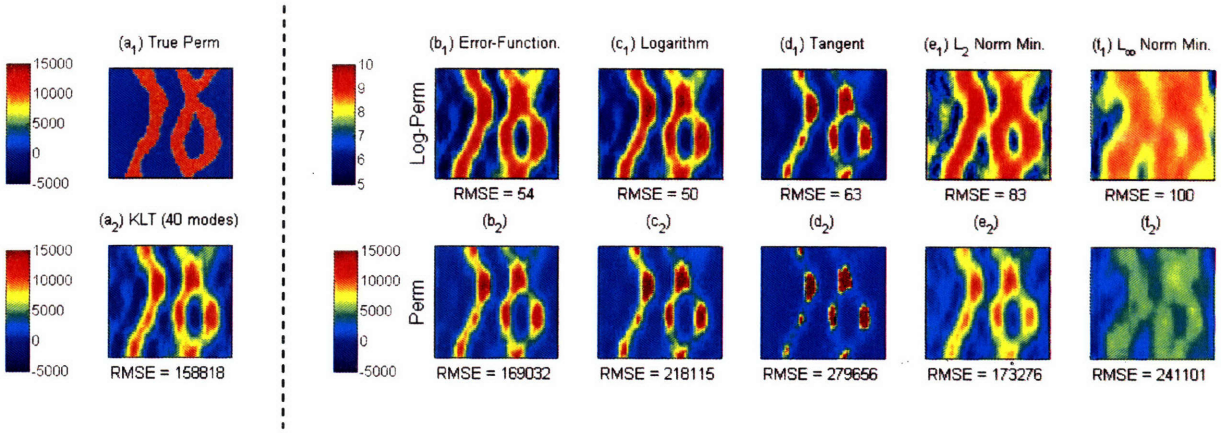


Figure 3A-1. Comparison between parameterization of the logarithm (c_1-c_2), the inverse error-function (b_1-b_2), and the tangent (d_1-d_2) of permeability field as well as the constrained minimization of the L_2 (e_1-e_2) and L_∞ (f_1-f_2) norms of the permeabilities. The first column shows the true (a_1) and KLT approximated (a_2) permeability fields (negative values of permeability are obtained as a result of the truncation used for approximation). The first row (b_1 through f_1), shows the logarithm of the estimated permeabilities in the second row (b_2 through f_2).

Examination of these results indicated that even though the log-permeability has the lowest RMSE in the logarithm space (as expected due to effectiveness of KLT in that space), its error is nonlinearly magnified after the exponential inverse transformation. This is due to larger error growth for overestimation than underestimations. The inverse error function transformation does not have this property and provides better results in the original permeability space. Note that each of these transforms provides effective KLT approximation (similar to a_2) in their respective spaces, i.e. logarithm, inverse error-function, and tangent. It is the nonlinear inverse transformation to the original permeability space that degrades the approximation quality.

The constrained minimizations of L_∞ norm was formulated as a convex linear programming problem and resulted in an overall underestimated approximation. Minimization of the L_2 norm was performed using constrained nonlinear least square minimization and provided an approximation RMSE that was only slightly larger than the direct KLT shown in the first column. However, the resulting RMSE is greater than in (b_2) suggesting that a local solution was reached. The computational effort in the constrained minimization approach can be considerable for large problems.

Based on the results from these experiments, the inverse error function transformation was selected in this chapter.

Chapter 4

History Matching with an Ensemble Kalman Filter and Discrete Cosine Parameterization

4.1 Summary

History matching of large hydrocarbon reservoirs is challenging due to several reasons including: 1) Scarcity of available measurements relative to the number of unknowns, leading to an ill-posed inverse problem; 2) Computational effort required for large reservoir problems; 3) The need to insure that solutions are geologically realistic. All of these problems can be helped by using algorithms that rely on efficient and parsimonious descriptions (or parameterizations) of reservoir properties. This chapter demonstrates the use of a novel parameterization approach, the discrete cosine transform, for history matching with a recently introduced sequential estimation technique, i.e. the ensemble Kalman filter. The proposed approach in this chapter exploits the structure of the estimation and parameterization algorithms to introduce reservoir states (pressures and saturation) as well as parameter (e.g. intrinsic permeability) reduction. The introduced methodology eliminates redundancy in posing the estimation problem and results in

additional computational savings The application and generality of this approach is demonstrated using two waterflooding experiments characterized by different types of geological variability.

4.2 Introduction

Predictive modeling of the production behavior of hydrocarbon reservoirs is of paramount importance for reservoir engineers. Accurate reservoir models can provide reliable production forecasts for guiding operational activities and development plans. Effective description of geological formations and their flow-related properties (such as permeability) is a significant part of any modeling exercise.

Several data acquisition techniques operating at various spatial scales provide information about the distribution of geological facies and rock properties within a petroleum reservoir. These include core characterizations, well logging, and seismic surveys. However, data acquisition costs limit the extent of such activities and the amount of data that can be collected in practice. Data limitations make it necessary to interpolate the available data to infer information about unobserved regions. Any interpolation scheme has to make assumptions (either implicit or explicit) about the structure of geological formations in the reservoirs. These assumptions can introduce a significant amount of uncertainty [4.1, 4.2].

To reduce the uncertainty in reservoir descriptions, reservoir model parameters are usually adjusted to improve the match between simulated and observed production data, a process known as history matching [4.3, 4.4]. Numerical reservoir simulators rely on discretized computational grids that require petrophysical properties such as intrinsic permeability and porosity to be defined for each grid block [4.5]. However, since geological media are connected layers of rocks with similar physical properties a strong correlation (or redundancy) exists between property values at neighboring grid blocks. When a history matching procedure independently adjusts property values in every grid block many different combinations of block values may yield comparable matches to a given set of production data. If this occurs the history matching problem is ill-posed (i.e. it is not possible to obtain a unique solution).

The problem of ill-posedness can be mitigated if the spatially variable grid block-based property values are replaced by a smaller number of parameters that capture all (or most) of the variability. This process is called parameterization (or sometimes reparameterization). Parameterization often improves ill-posedness because it constrains the set of possible grid block values. A parameterization approach should be selected so that it meets the needs of the reservoir simulator while retaining geological realism. An efficient and realistic parameterization can yield improved grid-based property estimates (by dealing with ill-posedness) and can decrease computational effort (by reducing the number of variables to be estimated).

Many parameterization approaches have been proposed in the petroleum engineering literature. These approaches obtain a reduced set of parameters by using methods that range from simple zonation [4.6] to more complex mathematical transforms [4.7-4.10]. Mathematical transforms that are based on spectral decomposition have the advantage of efficiently separating the large and small scale spatial features of a geological property such as permeability. If the finer details are omitted an approximate description of the original field is achieved with fewer parameters.

In Chapter 3 and [4.8], the discrete cosine transform (DCT) was proposed for parameterization of permeability field (and other spatially variable quantities) and compared its performance and flexibility with the Karhunen-Loeve transform (KLT). These two linear transformations use different basis functions. In Chapter 3, some of the important advantages of the DCT were discussed and demonstrated through examples.

In this chapter, an efficient and accurate history-matching algorithm is obtained by combining DCT parameterization with the ensemble Kalman filter (EnKF) [4.11], a recursive state estimation technique. In this algorithm, the DCT parameterization is applied to pressure and saturation as well as permeability and the Kalman filter state vector includes DCT transformed values of all three variables. This important extension of the parameterization concept is especially useful for Kalman filtering because it greatly reduces the size of the sample covariance matrices used to derive state estimates.

This chapter begins by discussing how the Ensemble Kalman filter can be applied to the reservoir history matching problem. Then two example problems that may be used to test the EnKF with different parameterization options are discussed. First the implementation and performance of an EnKF history matching algorithm that uses a conventional grid block-based description of spatial variability is discussed. Next, an overview of DCT parameterization and an illustration of its advantages with a modified EnKF history matching algorithm are presented. This is followed by a conclusion and a brief discussion of implementation and efficiency issues.

4.3 History Matching with the Ensemble Kalman Filter

The ensemble Kalman filter was first introduced by Evensen [4.11] in 1994 as a way to extend the classical Kalman filter to nonlinear problems [4.12]. A detailed discussion of the method and its evolution in the past decade as well as an efficient implementation of it can be found in [4.13]. Recent applications of the EnKF to history matching for petroleum engineering include [4.14-4.16].

The Kalman filter is designed to estimate the state vector x_t of a dynamic system from noisy measurements that are related to the states through a specified measurement equation. In petroleum reservoir applications the state vector typically includes pressure and oil saturation (in the examples of interest here capillary pressure is negligible so oil

and water pressures are approximately equal). The measurements in our water flooding examples are well bottom hole pressures and flow rates for each phase.

Reservoir history matching is designed to estimate geological properties such as intrinsic permeability, which are needed for predictive simulations. This exercise is often posed as a parameter estimation problem, rather than a state estimation problem. If a parameter estimation perspective is adopted the desired estimates are typically obtained by minimizing a performance measure that quantifies the mismatch between measured and predicted states. However, it is also possible to formulate the history matching problem as a state estimation problem. In this case the properties of interest are simply added to the state vector and estimated together with the other states (this process is often called state augmentation). A state estimation approach to history matching is particularly useful in real-time control applications, where new decisions are made as conditions change and new data become available.

The ensemble Kalman filter can be viewed as a recursive Bayesian procedure that approximates conditional probability densities of the state. These densities are the forecast (or prior) density $p[x_t \mid y_{0:t-1}]$, conditioned on all measurements $y_{0:t-1}$ taken prior to time t, and the updated (or posterior) density $p[x_t \mid y_{0:t}]$, conditioned on all measurements through time t.

Various properties of these conditional densities, such as the mean (the minimum error variance value) or the mode (most likely value) may be adopted as state estimates. A history matching procedure based on a state-augmented ensemble Kalman filter provides parameter and state estimates as well as distributions of possible values around these estimates. When a conventional grid block-based approach is adopted the discrete states are composed of the values of the state variables at the grid blocks. When a DCT parameterization is used the discrete states are the coefficients of the DCT basis vectors.

The ensemble Kalman filter approximates the forecast and updated densities with relatively small sets (or ensembles) of N random realizations (N = 300 in our experiments), denoted by $x^j_{t|t-1}$ and $x^j_{t|t}$, respectively, where $j = 1,\ldots,N$ denotes a particular replicate. An initial ensemble $x^j_{0|0}$ drawn at random from a specified population of initial states (pressures, saturations, and permeabilities) is used to initialize the sequence of forecasts and updates. Subsequent forecasts are derived from the augmented state equation, which may be written as:

$$x^j_{t|t-1} = f_t(x^j_{t-1|t-1}, u_{t-1}, w^j_{t-1}) \quad ; \quad j = 1,\ldots,N \qquad (4.1)$$

where u_{t-1} is a vector of known (non-random) time-dependent boundary conditions and controls and w^j_{t-1} is a random vector that accounts for uncertain model errors. The elements of the vector function $f_t(\cdot,\cdot,\cdot)$ corresponding to pressure and saturation

represent solutions obtained from the reservoir simulator ECLIPSE [4.17] by propagating $x^j_{t-1|t-1}$ and u_{t-1} forward from t-1 to t. The elements of $f_t(\cdot,\cdot,\cdot)$ corresponding to augmented parameters are selected to insure that $x^j_{t|t-1} = x^j_{t-1|t-1}$ when $x^j_{t|t-1}$ is an augmented parameter (e.g. permeability). This requirement follows because the parameter estimate should not change between updates.

A version of the classical Kalman filter update [4.12] is used to update the forecast replicates obtained from Equation (4.1):

$$x^j_{t|t} = x^j_{t|t-1} + Cov[x^j_{t|t-1}, y^j_{t|t-1}]Cov^{-1}[y^j_{t|t-1}, y^j_{t|t-1}][y_t - y^j_{t|t-1}] \quad ; \quad j = 1,\ldots,N \tag{4.2}$$

This update reflects information obtained from the measurement y_t taken at time t. The notation $Cov[\cdot,\cdot]$ represents the sample covariance between the ensembles associated with the two arguments. The perturbed measurement prediction $y^j_{t|t-1}$ is obtained from the following measurement equation:

$$y^j_{t|t-1} = g_t(x^j_{t|t-1}) + v^j_t \quad ; \quad j = 1,\ldots,N \tag{4.3}$$

where v_t^j is a vector of measurement errors, drawn at random from a specified population, and $g_t(\cdot,\cdot)$ is the measurement operator that relates the measurements at t to the states at t. Equations (4.1), (4.2), and (4.3), together with the initial state $x_{0|0}^j$, define the ensemble Kalman filter recursion for the problem of interest here.

It is useful to note some of the distinctive attributes of the ensemble version of the Kalman filter. The ensemble Kalman filter can produce non-Gaussian sample distributions of the states $x_{t|t-1}^j$ and $x_{t|t}^j$. There is no need to linearize or otherwise approximate the state transition function $f_t(\cdot,\cdot,\cdot)$ or to assume that the random states and inputs are Gaussian. However, the ensemble Kalman filter replicates converge to the true conditional densities $p[x_t \mid y_{0:t-1}]$ and $p[x_t \mid y_{0:t}]$ only when all states and measurements are jointly Gaussian. If this condition is not met, the filter's sample densities and moments are only approximations.

The covariance inversion operation in (2) is computationally expensive and can be ill-conditioned for large problems. This can be avoided by applying a pseudo inverse procedure based on a singular value decomposition, as proposed in [4.13]. However, for large problems even the singular value decomposition approach can be very slow. In general, more efficient implementations of the update step may be necessary for large

problems. In this chapter, efficiency was achieved by using the smaller state dimension provided by a DCT parameterization.

4.4 Formulation of the Example Problems

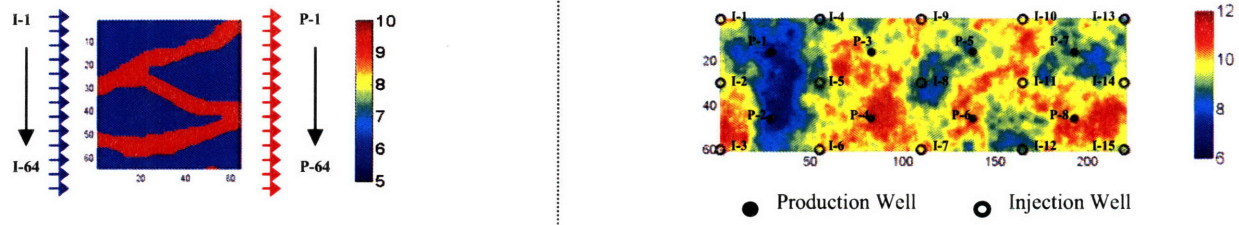
4.4.1 Reservoir Descriptions

Two water flooding examples are used to demonstrate the effectiveness of the approach presented in this chapter. Figure 4.1a shows the reservoir configurations for these two experiments.

Reservoir A has a line drive with 64 injectors (on the left) and 64 producers (on the right) while Reservoir B is produced using a pattern drive as shown in Figure 4.1a. General information about the simulations for each reservoir is provided in Tables 1. The simulations are performed with the ECLIPSE reservoir simulator, which is set up for two phase (oil and water) black oil flow. The total simulation time is 1080 days, divided into 12 intervals of 90 days.

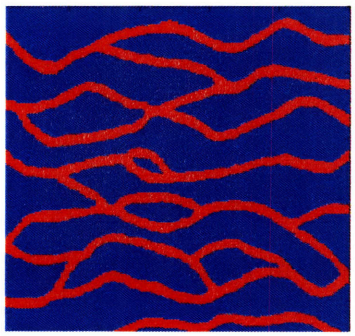
Reservoir A **Reservoir B**

(a) True Reservoirs Configuration

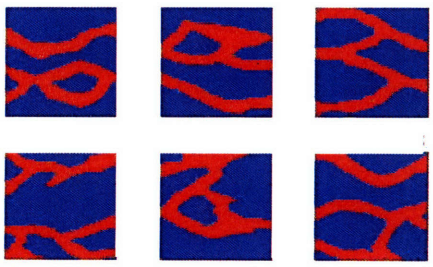


(b) Prior Permeability Models

Log-permeability training image for Reservoir A



Initial log-permeability realizations for Reservoir A



Initial log-permeability realizations for reservoir B

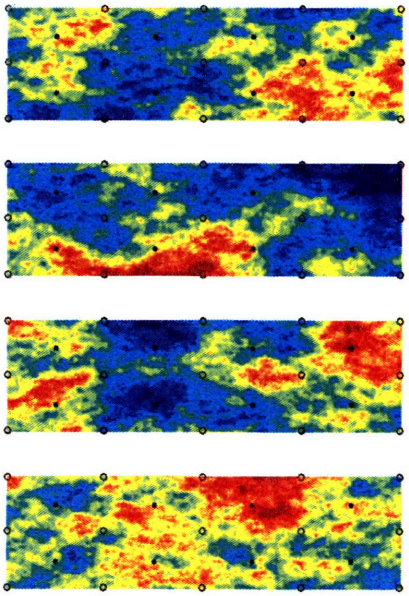


Figure 4.1 Reservoirs configuration with the "true" permeability fields (a), and prior permeability models (b) for Experiments A1, A2, B1, and B2.

The injection wells are operated with specified flow rates while the production wells are operated under bottom hole pressure control. The well settings are specified without any intent to maximize oil recovery. Rather, the dynamic well controls are used to excite the reservoir and increase the information content of its response. Each injection/production scenario lasts for 90 days before it changes to a new set of well settings. Details about production scenarios in each experiment are summarized in Table 2.

In this study the only source of uncertainty considered is the permeability, which is treated as a spatially variable random field. Initial and boundary conditions are assumed to be known perfectly and dynamic model errors are assumed to be negligible. In situations where these assumptions may not hold additional error sources may be included in the ensemble filtering process.

In its update step, the ensemble Kalman filter uses measurements of bottom hole pressure observations in injection wells and oil and water flow rate measurements in production wells. In each experiment the "true" injection bottom-hole pressures and volumetric production quantities are generated by running the simulator with a specified "true" permeability field. Uncorrelated

zero-mean random measurement errors are added to these "true" pressures and flow rates. The standard deviations of the random measurement errors for each experiment are listed in Table 1.

The classical ensemble Kalman filter update can potentially give unphysical saturation values outside the range [0,1]. In order to avoid this problem, our Kalman filter updates the following inverse error function transform of saturation, which is distributed over (-∞, +∞):

$$x^* = erf^{-1}(2 \cdot x - 1) \qquad (4.4)$$

After each filter update the saturation is computed by applying the inverse transformation to obtain updated saturations for future simulation runs.

4.4.2 Reservoir Permeabilities

Reasonable probabilistic models for the permeability field need to be constructed before a successful ensemble Kalman filter implementation can be obtained [4.18]. In an ensemble implementation a probabilistic model is simply a procedure for generating realistic unconditional permeability replicates. In practice, the generation procedure should use as many sources of information as possible and should be based on sound geological and geostatistical principles

Table 4.1 Simulation and Assimilation Information for Reservoirs A and B

Simulation/Assimilation Information		
Simulation Information	**Reservoir A**	**Reservoir B**
Simulation	*2 Phase Oil-Water*	*2 Phase Oil-Water*
Simulation Time (days)	*12x90 = 1080*	*12x90 = 1080*
Grid Blocks	*64x64x1*	*60x220x1*
Grid Block Dimensions (m)	*10x10x10*	*10x10x10*
Geometry	*2D-Areal*	*2D-Areal*
Porosity	*0.2 (constant)*	*0.2 (constant)*
Connate Water Saturation	*0.1 (uniform)*	*0.1 (uniform)*
Initial Pressure	*3000 psi (uniform)*	*3000 psi (uniform)*
Total Injected Water	*1 PV*	*1 PV*
No. Injection Wells	*64*	*15*
No. Production Wells	*64*	*8*
Injection Well Control	*Water Rate*	*Water Rate*
Production Well Control	*BHP*	*BHP*
Observation Information		
Observation Intervals	*90 days*	*90 days*
Injection Wells	*bhp*	*bhp*
Noise Standard Deviation	*20 psi*	*100 psi*
Production Wells	*Oil and Water Rates*	*Oil and Water Rates*
Noise Standard Deviation	*20 sbpd*	*100 sbpd*

Table 4.2 Production scenarios for Reservoirs A and B

Production Scenarios								
Reservoir A	*IG1*	*IG2*	*IG3*	*IG4*	*PG1*	*PG2*	*PG3*	*PG3*
Production Period	***Injection Volume (PV)***				***Production BHP(psi)***			
A) 1-180 days	1/6 PV	1/6 PV	1/6 PV	1/6 PV	2950	2950	2950	2950
B) 181-360 days	1/6 PV	0	0	1/6 PV	2950	3000	3000	2950
C) 361-540 days	0	1/6 PV	1/6 PV	0	3000	2950	2950	3000
D) 541-720 days	0	1/6 PV	1/6 PV	0	3000	2950	2950	3000
E) 721-900 days	1/6 PV	0	0	1/6 PV	2950	3000	3000	2950
F) 901-1080 days	1/6 PV	1/6 PV	1/6 PV	1/6 PV	2950	2950	2950	2950

IG1: Wells I-1 to I-16 IG2: Wells I-17 to I-32 IG3: Wells I-33 to I-48 IG4: Wells I-49 to I-64
PG1: Wells P-1 to P-16 PG2: Wells P-17 to P-32 PG3: Wells P-33 to P-48 PG4: Wells P-49 to P-64

Reservoir B	*IG1*	*IG2*	*IG3*	*PG1*	*PG2*
Production Period	***Injection Volume (PV)***			***Production BHP(psi)***	
A) 1-180 days	1/6 PV	1/6 PV	1/6 PV	2500	2500
B) 181-360 days	0	1/6 PV	0	2500	3000
C) 361-540 days	1/6 PV	0	0	2500	3000
E) 721-900 days	1/6 PV	1/6 PV	1/6 PV	2500	2500
F) 901-1080 days	0	1/6 PV	0	3000	2500

IG1: Injection wells I-1, I-4, I-7, I-10, and I-13 all on the top row in Figure 1
IG2: Injection wells I-2, I-5, I-8, I-11, and I-14 all in the middle in Figure 1
IG3: Injection wells I-3, I-6, I-9, I-12, and I-15 all on the bottom row in Figure 1

PG1: Production wells P-1, P-3, P-5, and P-7 in Figure 1
PG2: Production wells P-1, P-4, P-6, and P-8 in Figure 1

In our synthetic experiments unconditional permeability realizations were generated for two structurally different formations, identified as Reservoirs A and B. These correspond to a predominantly channelized environment (fluvial deposition) and a prograding near-shore basin, respectively. Reservoir A is the subject of our Experiments A1 and A2 while Reservoir B is considered in Experiments B1 and B2. The "true" permeability fields shown in Figure 4.1a were used to generate noisy measurements and to generate "true" saturation and pressure values. The set of true permeability, pressure, and saturation states is used to assess history matching performance. The true permeability field for Reservoir A is a random realization drawn from the training image in Figure 4.1. The "true" permeability field shown for Reservoir B corresponds to the top layer of the SPE10 comparative model [4.19]. In the original SPE10 model, the contrast between the highest and lowest permeability values is about six orders of magnitude with the lowest permeability values in the order of 10^{-3} mD. The SPE10 permeability values were rescaled to the range 10^{2}–10^{5} mD to avoid numerical stability issues and speed the convergence of the linear solver in the reservoir simulator.

Two different geostatistical modeling methods were used to generate prior permeability replicates that are structurally similar to the two true fields. For Reservoir A, the geologic features common to all members of the training set are defined by the channelized training image shown in Figure 4.1b. This training image has 250 × 250 × 1 grid blocks comprising two facies types: low permeability background shale and high permeability sandstone channels. The background shale permeability is assumed to be 500mD while

the embedded high-permeability sandstone channels have a permeability value of 10,000mD.

The permeability realizations shown in Figure 4.1b are generated using the multiple-point geostatistical algorithm *SNESIM* [4.20-4.21]. Each realization is discretized over a 64 × 64 × 1 (640m × 640m × 10m) grid block system. The varying shape and geometry of the channel facies in these realizations are the major source of uncertainty. The highly structured nature of the facies distribution in these realizations suggests a distinctive preferential flow displacement pattern. In addition, a high level of correlation (redundancy) is observed in description of these facies, suggesting that the field could be represented much more efficiently if instead of the block values a more appropriate description was adopted (this will be discussed in later sections).

Figure 4.1b also shows sample realizations for Reservoir B , which lies in a different type of geological formation. The Reservoir B samples are from a prior permeability ensemble generated with the two-point geostatistical algorithm *SGSIM* [4.2, 4.21]. Each realization is discretized over a grid of 60 × 220 × 1 blocks, with each block size of 10ft × 10ft × 10ft. The Reservoir B field is based on an exponential variogram model with assumed correlation lengths of 100 and 50 grid blocks in the x and y directions, respectively. This correlation length ratio is not based on the true SPE10 model. In fact, the correlation length in the top layer of SPE10 model appears to be higher in the *y* direction than in the

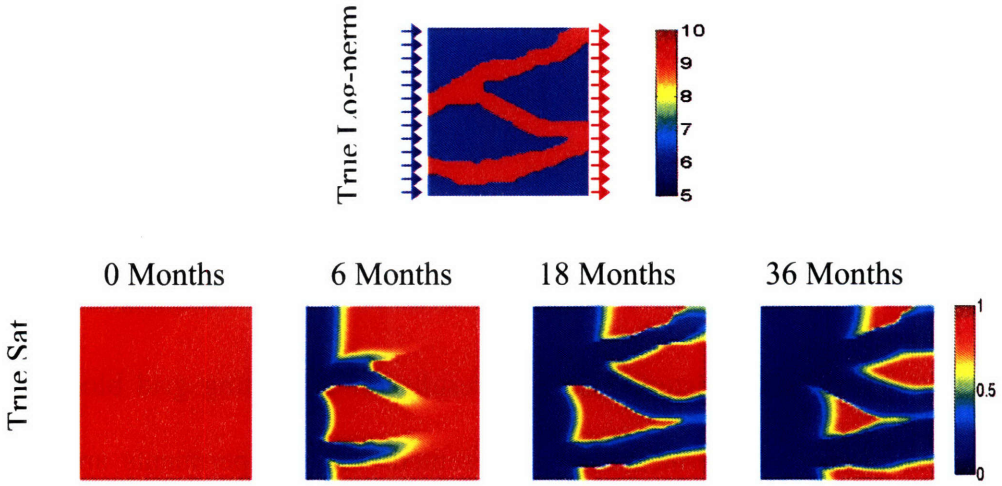
x direction. However, the permeability variance is selected to cover the expected range of permeability values in the SPE10 data set.

4.4.3 History-Matching Results for Grid Block-Oriented Estimation

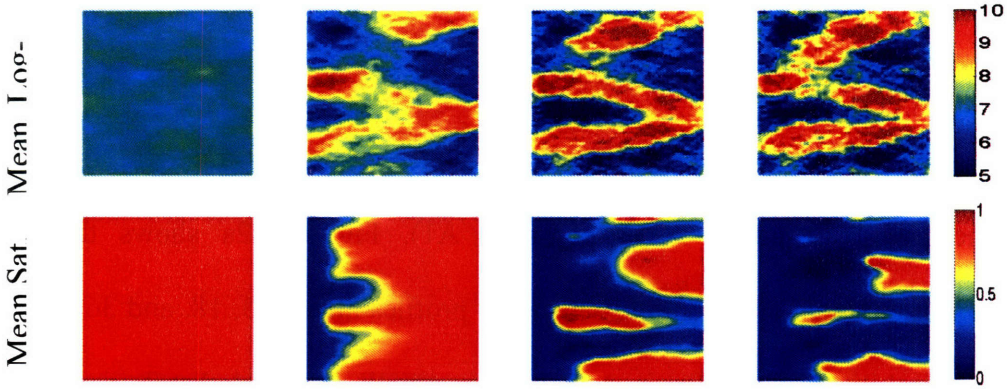
In this section the EnKF history matching results for estimating grid block values of pressure, saturation, and permeability is presented. Two experiments are considered based on Reservoirs A and B, respectively.

Figure 4.2 summarizes the EnKF history matching results for grid block-oriented estimation with Reservoir A (Experiment A1). Figure 4.2a shows the true log-permeability and its corresponding saturation profiles after 0, 3, 18, and 36 months. The prior permeability for Reservoir A follows a strictly bimodal distribution. Moreover, saturation profiles also tend to show a bimodal behavior except near the moving water front. This bimodality can make the estimation of facies distribution a challenging task for the linear update in Equation (4.2), which is known to be optimal only under jointly Gaussian priors and observations [4.12]. Nevertheless, the strong correlations between states and observations help to insure that the covariance-based Kalman updates are in the right direction, provided that correlation information is correctly specified.

a) Experiment A1 "True" log-permeability and saturations



b) Experiment A1 mean log-permeability and saturation estimates



c) Experiment A1 production forecast

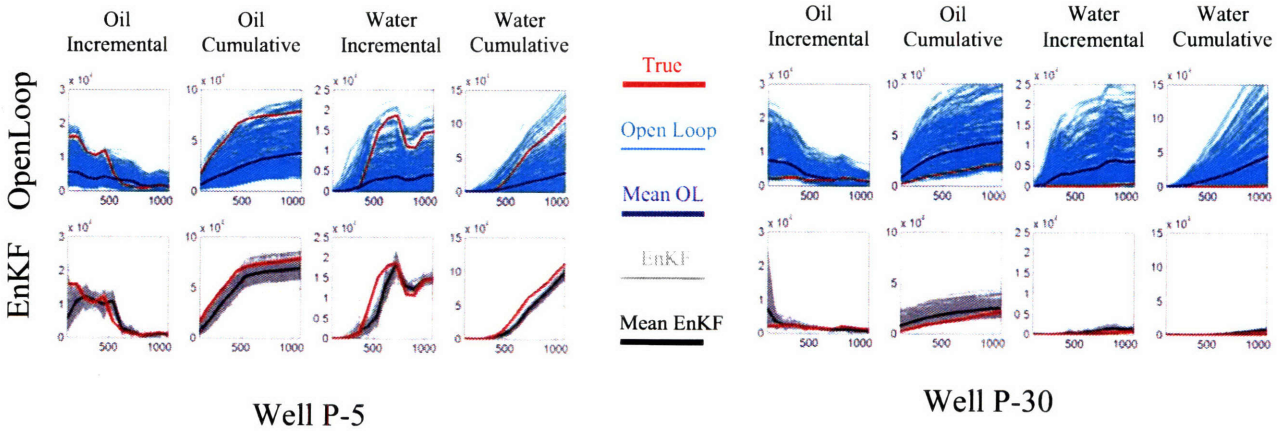


Figure 4.2 Results for Experiment A1 (estimating grid-block properties with the EnKF for Reservoir A): (a) the true log-permeability field and snapshots of its corresponding synthetic true saturation profiles generated by ECLIPSE reservoir simulator; (b) pixel-based EnKF mean log-permeability and saturation estimates at selected update times; (c) production observation (red), openloop forecast and its mean (cyan and blue), and EnKF forecast and its mean (grey and black).

Figure 4.2b shows the mean over the ensemble of the updated permeability and saturation profiles for Reservoir A at the EnKF measurement (update) times. While the estimated permeability field has captured the main large scale permeability trends present in the true field, the channels are somewhat disconnected in the interior of the field. The EnKF update appears to have difficulty capturing permeability connectivity (channel structure), even though the prior replicates are all connected. A contributing factor to this loss of connectivity is, at least in part, the absence of observations away from the boundaries where the wells are located. Covariances derived from the prior permeability replicates are used to interpolate between the dynamic observations at the ends of the domain where wells (and measurements) are located. Estimates in the interior far from the wells tend to have higher uncertainties (not shown).

A close inspection reveals that the locations of the "channel" boundaries produced by the history matching procedure are slightly offset. For example, in the final permeability estimate (rightmost plot) the top channel is closer to the top boundary than it is in the true field. As a result the estimated saturation profile in the top right region shows no residual oil, in contrast to the true saturation profile that reveals that a considerable amount of oil remains. A similar trend is observed in the low-permeability region in the middle of the reservoir between the bottom and middle channels. Since the estimated low-permeability feature is too narrow compared to the true permeability field the leftover oil in the middle of the reservoir is underestimated. While this visual analysis of estimation results is not a precise performance measure, it provides a useful qualitative assessment, particularly when considered in conjunction with more quantitative measures.

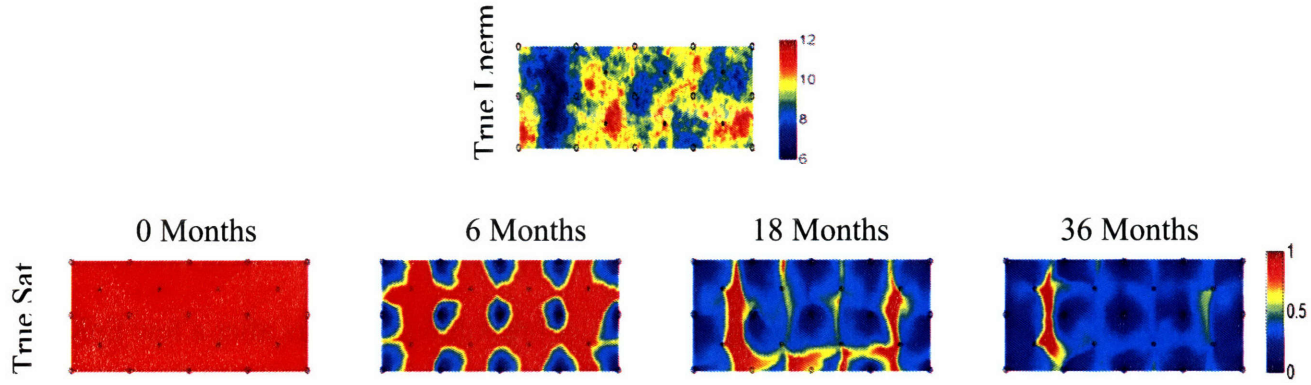
Particularly relevant quantitative performance measures for history matching can be derived from the differences between the predicted and measured oil and water production (prediction errors) at different time steps. When the predicted values are open-loop (i.e. when they are obtained without history matching) the root-mean squared prediction error defines a reference performance level. If the history matching procedure is helpful then predictions produced by the ensemble Kalman filter should give a smaller root-mean squared error. Moreover, the filter performance should improve over time, as more measurements are collected. This improvement should be accompanied by a decreased ensemble spread around an ensemble mean that is close to the true value.

The production forecast curves for Reservoir A are shown in Figure 4.2c. Plots for incremental and cumulative oil (first and second columns) and water (third and fourth columns) productions are shown for wells P-5, P-30. These wells correspond to two separate regions (inside and outside the channel facies) in the production end of the field (on the right). Well P-30 is located in a low permeability area and does not show any sign of water breakthrough. This is in contrast to Well P-5, which is inside a high permeability region and experiences early water breakthrough. The first and second rows in each plot show the open loop and EnKF forecast results, respectively. Figure 4.2c indicates that oil production predictions improve with time, suggesting that the updates continuously improve the predictive power of the reservoir model after each analysis step. This can also be observed from the improvement in the permeability field (Figure 4.2b).

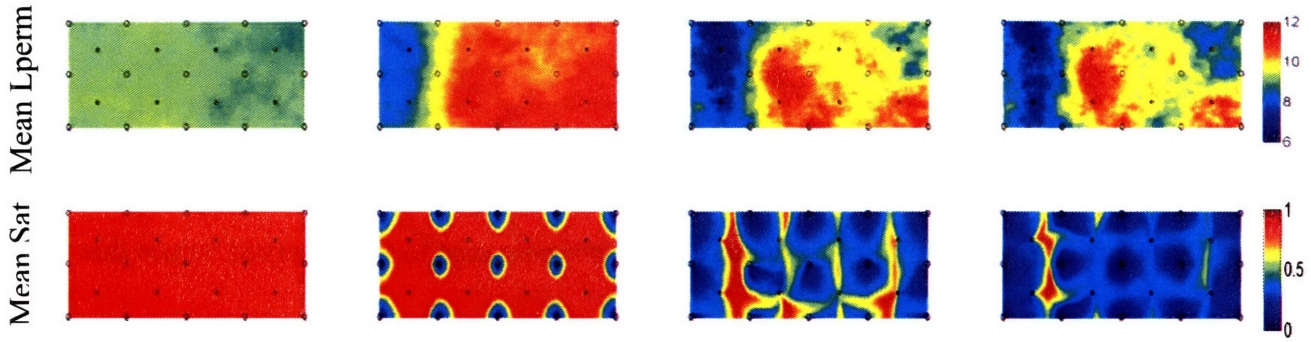
While the rate of oil production generally decreases in time, water production is initially close to zero until breakthrough occurs and increases thereafter. As a result, initial deviations between predicted and true production curves are larger for oil and smaller for water. The third and fourth columns of Figure 4.2c indicate that the water production has less ensemble spread and better prediction quality than oil production even at later times. This can be explained by observing that the water breakthrough occurs after a few EnKF updates have already been carried out, when a more accurate model is available for generating water production predictions.

It is important to note that even though the later updates improve forecast quality, the cumulative production plots (second row) do not converge to the observed values due to the initial forecast errors. This highlights an important implication in production optimization. In general, early injection and production control decisions are based on poorer knowledge of reservoir properties and are therefore suboptimal. This implies that more frequent monitoring and model updating should be performed at early stages of the reservoir's lifetime.

a) Experiments B1 "True" log-permeability and saturations



b) Experiment B1 mean log-permeability and saturation estimates



c) Experiment B1 production forecast

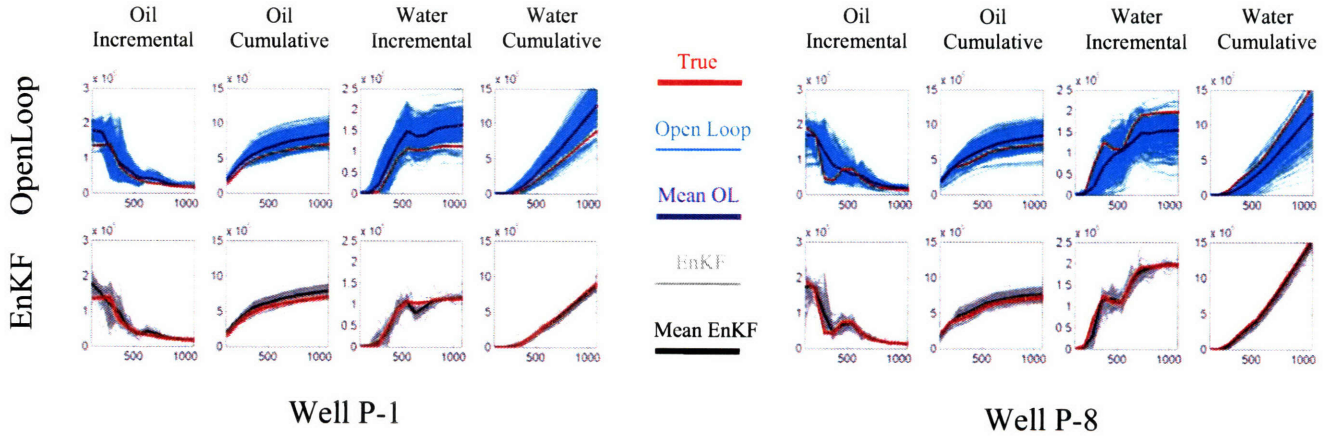


Figure 4.3 Results for Experiment B1 (estimating grid-block properties with the EnKF for Reservoir B): (a) the true log-permeability field and snapshots of its corresponding synthetic true saturation profiles generated by ECLIPSE reservoir simulator; (b) pixel-based EnKF mean log-permeability and saturation estimates at selected update steps; (c) production observation (red), openloop forecast and its mean (cyan and blue), and EnKF forecast and its mean (grey and black).

Figure 4.3 summarizes the EnKF history matching results for grid block-oriented estimation with Reservoir B (Experiment B1). The true log-permeability and saturation profiles after 0, 3, 18, and 36 months are shown in Figure 4.3a while the ensemble mean permeability and saturation updates are shown in Figure 4.3b. The third panel in the figure (Figure 4.3c) shows the production plots for wells P-1 and P-8 (see Figure 4.1). P-1 is a production well located in a low permeability region (refer to Figure 4.1) while P-8 is in a high permeability zone.

From an estimation viewpoint, Experiment B1 differs in several ways from Experiment A1. The ratio of observation points to the total number of grid-blocks for Reservoir B is much less than for Reservoir A, i.e. 23/13200 compared with 128/4096. On the other hand, the Reservoir B observation points are distributed more evenly throughout the field and cover a larger area. This observing strategy can be expected to provide more information content (better observability) than the one used for Reservoir A. Another major difference between the two experiments is the more nearly Gaussian character of the prior permeability distribution in Experiment B1 (although the saturation profile still shows bimodal characteristics). This nearly Gaussian behavior is more compatible with the implicit assumptions of the EnKF.

It is clear from Figure 4.3b that the filter is able to distinguish high and low permeability regions in the field. The updates at earlier times appear to have only found the two major high and low permeability areas. However, further details of the permeability field are

retrieved at later update times. As expected, permeability trends in the vicinity of observation points are estimated more accurately than in regions between the wells.

The estimated saturation results after 18 and 36 months appear to be very close to the true saturation profiles at those times. However, the saturation profile at 6 months is very different from true saturation. This is explained by the inaccurate permeability estimates at the initial update steps.

Production plots in the second row of Figure 4.3c show the incremental improvement in the quality of the EnKF estimates (first and third columns) compared to open loop results (first row). It is clearly seen that the EnKF ensemble spread is substantially reduced in comparison with the open loop production ensemble. The mean EnKF production forecast approaches the true production curve, indicating an improvement in estimation accuracy.

The results presented in Figure 4.2 and 4.3 confirm the effectiveness of the EnKF for history matching. However, while updating grid-block values of states and parameters may be computationally reasonable for small problems, there are a few drawbacks to this approach. First, the computational cost of applying the update to individual grid-block values can become considerable for large problems ($\sim 10^6$-10^7 grid blocks). Second, neighboring blocks have strongly correlated values both within and across states. In this

case grid block-oriented estimation can be both wasteful and geologically unrealistic. For example, much work can be expended to obtain locally disconnected estimates for problems with continuous facies (Figure 4.2b). It is possible to obtain a better posed and less computationally demanding history matching problem by changing the way that spatial variability is represented. This is the topic of the next section.

4.5 Parameterization with the Discrete Cosine Transform

This section provides a brief overview of the discrete cosine transform (DCT), with an emphasis on its use for parameterizing spatially variable reservoir properties such as permeability. The DCT was first introduced in [4.22] for signal decorrelation and has been widely used for image compression after its introduction [4.23-4.24]. A detailed discussion of the use of DCT for parameterization and its properties is given in [4.8].

The DCT is a Fourier-based image compression technique that uses orthonormal cosine functions as bases for representing a given image (see Figure 4.4a). Suppose that the original image is represented by a grid of n equal size blocks (or pixels), with a distinct variable value (e.g. grey scale intensity or permeability) assigned to each block. When a DCT parameterization is used the variable value in each block is represented as the sum of r of the cosine basis functions. Since the basis functions are fixed the r DCT weighting coefficients constitute an alternative way to describe the image (i.e. once these coefficients are specified the block values can be reconstructed). If $r = n$ the DCT

coefficient representation can reproduce the image perfectly. When $r < n$ some information is usually lost but the DCT coefficient representation is more compact. But in many cases the compressed DCT representation of the image is still quite good, even for r much less than n.

The discrete cosine transform can be concisely summarized with a pair of linear equations. The inverse transformation from r DCT basis function coefficients ($\mathbf{v}$) to n grid block values ($\mathbf{u}$) is given by:

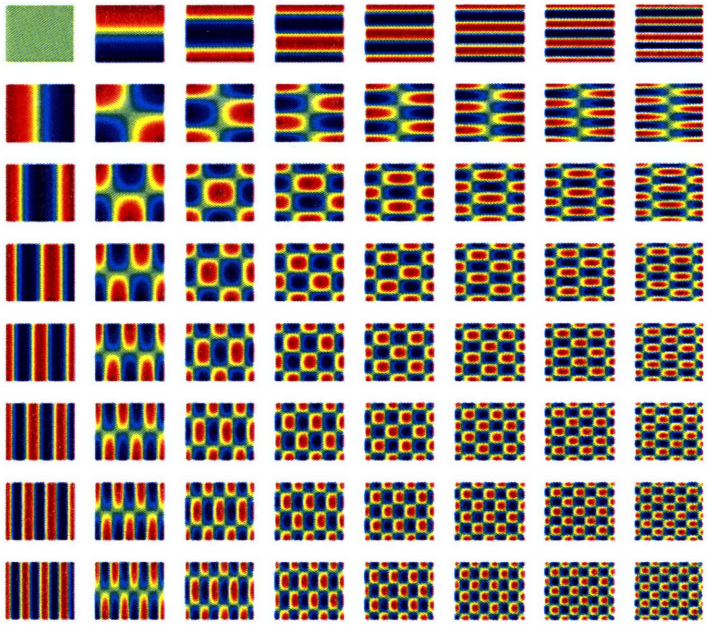
$$\mathbf{u} = \mathbf{\Phi}\, \mathbf{v} \tag{4.5}$$

The forward transformation from n grid block values ($\mathbf{u}$) to r DCT basis function coefficients ($\widetilde{x}_t$) is given by:

$$\mathbf{v} = \mathbf{\Phi}^T\, \mathbf{u} \tag{4.6}$$

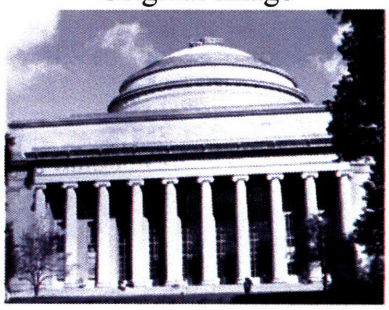
Where $\mathbf{\Phi}$ is an n by r matrix with columns consisting of the r retained DCT basis functions. Equation (4.6) follows from Equation (4.5) and the fact that the DCT basis functions are orthonormal. In practice the DCT coefficients are computed with a sequence of one-dimensional Fast Fourier transforms.

(a) Low-frequency discrete Cosine transform modes

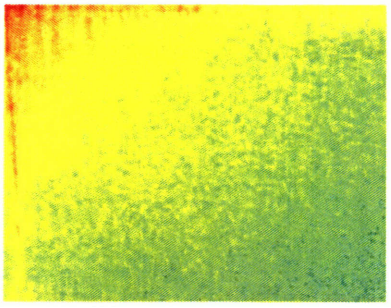


(b) An example image and its log-DCT coefficients

Original Image



Log-DCT Coefficients



(c) Low-rank approximations with increasing number of DCT modes

0.1% Coeffs. 1% Coeffs. 5% Coeffs. 10% Coeffs.

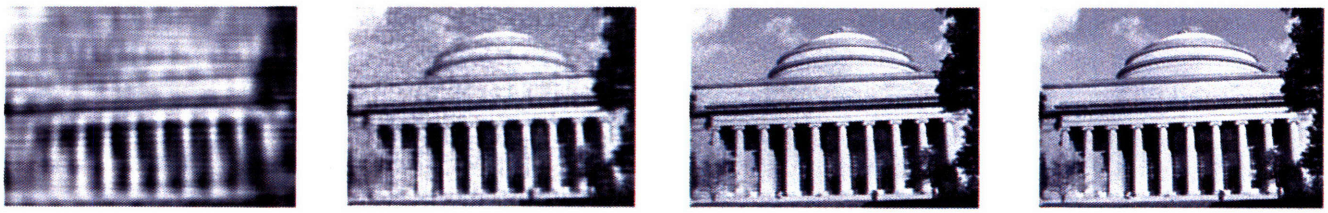


Figure 4.4 Compression power of the discrete cosine transform: (a) sample discrete cosine transform modes (bases); (b) an example image (the famous MIT dome) with its log-DCT coefficients; (c) approximate representations of the example image with increasing number of included modes.

Figure 4.4b shows an example to illustrate the compression capabilities of the DCT. An image of MIT's symbolic dome is shown with a truck on top of it. The logarithm of the magnitude of the DCT transform coefficients of this image is shown next to it. It is clear that only a few of the DCT coefficients in the upper left corner have large magnitude (red areas), implying that the corresponding basis functions are responsible for most of the variability in the image. Note that the upper left corner DCT basis functions are those with the largest periods (or lowest wave numbers), as indicated in Figure 4.4a. Figure 4.4c shows a series of approximations for the original image, obtained by keeping 0.1%, 1%, 5%, and 10% of the total number of DCT basis function coefficients. A reasonably good approximation of the image is obtained by including only 5% of the original coefficients (the truck on the top of the dome is clearly visible when only 5% of the coefficients are kept). This powerful compression property of the DCT transformation is responsible for its widespread use in JPEG image compression standard [4.24].

While size reduction (efficiency) is a common theme for image compression and parameterization of spatially distributed reservoir properties, the two problems are fundamentally different. In image compression, the weighting coefficient assigned to each basis function is known since the DCT coefficients can be computed directly from the known image. Consequently, it is easy to rank all n basis coefficient values and select the set of r bases that make the greatest contribution (i.e. have the largest coefficients).

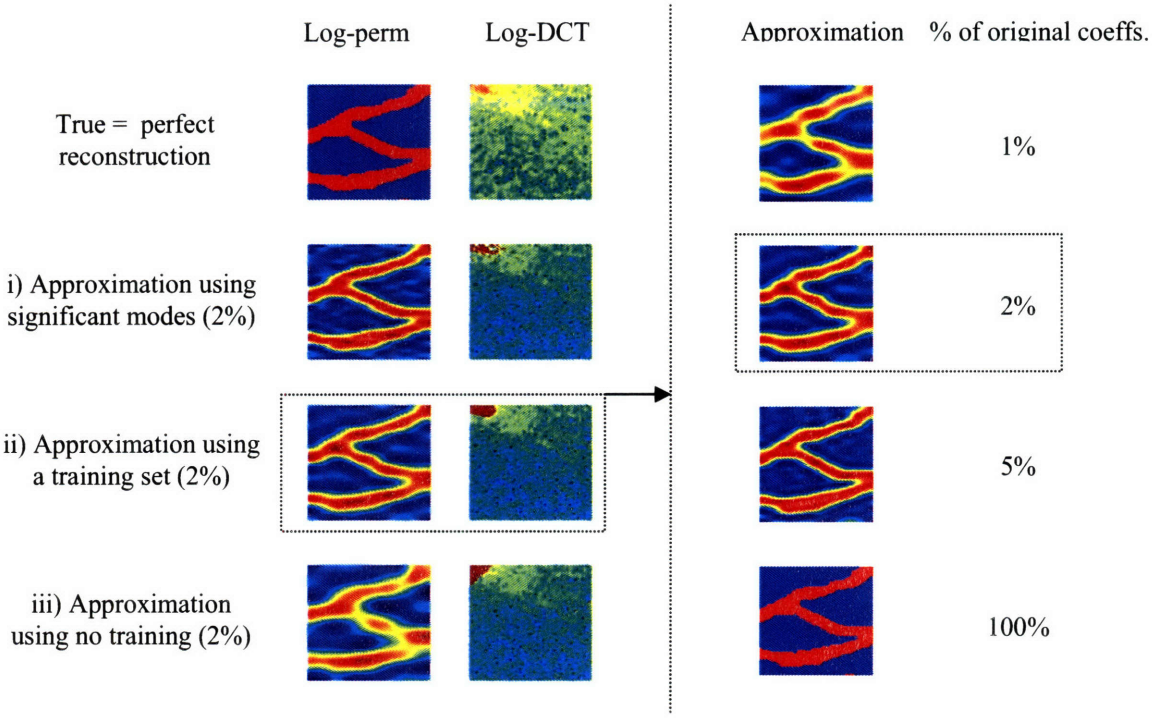
The parameterization problem is different because the image is not known and the r leading weighting coefficients cannot be identified in advance. In this case, it is common to retain the r basis function terms that are most important in the aggregate, over some set of specified images (called a training set) or, alternatively, to simply specify, *a priori*, the r basis function terms that are believed to be most important. In the first case the DCT is said to be "trained" while in the second case it is "untrained". In either case, the unknown basis function coefficients must be estimated from a history matching procedure.

We can use the two permeability models discussed earlier to demonstrate alternative methods for selecting dominant DCT basis functions. Figures 4.5a and 4.5b show in Row 1 the true permeabilities along with their DCT coefficients for Reservoirs A and B, respectively.

.

Rows 2, 3 and 4 of Figures 4.5a and 4.5b show approximations using, respectively, 2% and 5% of the original coefficients for Reservoirs A and B. The retained basis functions are associated with i) the most significant coefficients of the true permeability field (Row 2), ii) a training set of prior permeability fields (Row 3), or iii) an *a priori* selection that does not rely on a training set

a) DCT parameterization for Reservoir A



b) DCT parameterization for Reservoir B

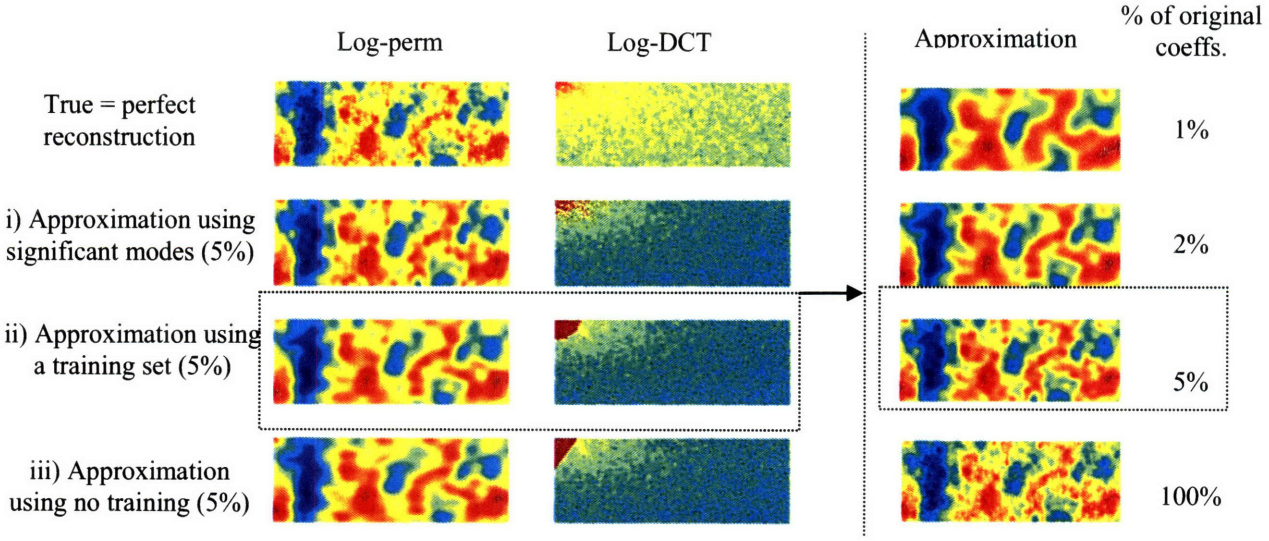


Figure 4.5 Mode selection and approximation with the DCT: (a) flexible strategies for mode selection with or without prior training; (b) approximate representation of a sample permeability with increasing number of modes.

(Row 4). In the subsequent history matching examples a training set was used to select the r retained basis function terms (Option ii). This is a reasonable approximation to the approach that would likely be used in practice. However, it was shown in Chapter 3 that an untrained approach for selecting basis function terms may be more robust than a trained approach since it is not misled by incorrect training images. The most appropriate choice depends on the application.

Figures 4.5a and 4.5b also show how image approximation quality changes with an increasing number of retained basis functions. It should be noted that perfect reconstruction (synthesis) of the original permeability field is possible if no truncation is carried out ($r = n$). For the history matching examples 2% of the original set of trained basis function coefficients were kept for the channelized model of Reservoir A ($r = 0.02n$) and 5% for the SPE model of Reservoir B ($r = 0.05n$).

4.5.1 History-Matching Results for the DCT Parameterization

In this section the block-oriented history matching experiments described in Section 4 to illustrate how the DCT method can provide low-dimensional parameterizations suitable for history matching is discussed. Various history matching parameterization methods have been investigated in the past [4.6-4.10]. Typically, these methods have only been used to reduce the number of unknown geological properties (e.g. permeabilities and/or porosities) adjusted in the history matching procedure. In this case the grid block values

of pressure and saturation states are obtained from the reservoir simulator, using grid block property values derived from the parameterization.

When the EnKF is used for history matching the forecast pressure and saturation states are derived from the simulator but the updated pressure and saturation are derived from the Kalman update of Equation (4.2). Therefore, history matching with the EnKF lends itself to more parameterization options. In particular, an efficient parameterization can be used to reduce the number of updated pressure and saturation states as well as the number of unknown permeability or porosity states.

The DCT parameterization of the EnKF update replaces the forecast grid block values of pressure, saturation, and permeability (a total of $3n$ states) contained in $x_{t|t-1}$ by their corresponding DCT coefficients (a total of $3r$ states). These DCT coefficients are contained in $\widetilde{x}_{t|t-1}$ and computed from Equation (4.6). The forecast DCT coefficients are updated with an appropriate version of Equation (4.2) (expressed in terms of the DCT state vectors $\widetilde{x}_{t|t-1}$ and $\widetilde{x}_{t|t}$ rather than the grid block-oriented state vectors $x_{t|t-1}$ and $x_{t|t}$). The updated grid block values of pressure, saturation, and permeability are then constructed from the updated DCT coefficients and the specified DCT basis functions, using Equation (4.5).

The implications of the reduction in state vector dimension provided by the DCT parameterization can be significant for large problems. If the state dimension is large the EnKF update scheme can be very time-consuming. With our proposed parameterization of all the states (pressure, saturation, and geological properties) the size of the EnKF update is considerably reduced, making the filtering approach much more practical for realistic large-scale applications.

In Section 4 (Experiments A1 and B1), the values of the permeability, pressure, and saturation fields for each individual grid block were estimated. In general this approach can result in discontinuous estimates that are geologically inconsistent with prior knowledge. The loss of structural continuity observed in the block-oriented approach can be mitigated if we use a truncated parameterization such as DCT that emphasizes large scale (low wave number) features over small scale (high wave number) features. In this case the estimated permeabilities are more correlated over space and continuous channels are more likely to be properly identified.

These properties of the DCT are illustrated in Figures 4.6 and 4.7, which show the history matching results from the ensemble Kalman filter with DCT parameterization for Reservoirs A and B (Experiments A2 and B2), respectively. Figure 4.6a and 4.7a show the true log-permeability and its corresponding saturation profiles after 0, 3, 18, and 36 months for the two reservoirs. Figures 4.6b and 4.7b show the updated ensemble mean permeability and saturation for each case. Finally, Figures 4.6c and 4.7c compare the

incremental and cumulative oil and water production forecast with the open loop and true values for the same production wells as in Figures 4.2c and 4.3c.

It is clear from a comparison of Figures 4.2b and 4.6b that the estimated permeability and saturation in Figure 4.6b are smoother than those in Figure 4.2b. The smoothness and continuity introduced by the DCT parameterization appear to have improved the connectivity between the channel facies. The saturation estimates in Figure 4.6b seem to be slightly more accurate than those in Figure 4.2b. In particular the residual oil values observed in the top right and bottom right regions of the reservoir after 36 months are closer to the true saturation in Figure 4.6b than they are in Figure 4.2b.

The production plots in Figure 4.6c for the DCT parameterization are almost identical to those in Figure 4.2c for the block-oriented alternative. This is not surprising as the permeability and saturation profiles are only marginally different in the two experiments, apart from the smoothness in the case of parameterized estimates. Our analysis of the production plots in Figure 4.6c also applies to Figure 4.2c as there is no visible difference between the results.

Figure 4.7 shows the results for Reservoir B when DCT parameterization is used to concisely describe the saturation, pressure, and permeability fields. Again in this case the results are in good agreement with those from the block-oriented history matching results

shown in Section 4 (see Figure 4.3). In particular, the saturation and production plots are almost identical for the block-oriented and DCT approaches.

These results imply that estimation of the fine details in the parameterization does not contribute significantly to the overall estimation results. In other words, the observations do not provide the information needed to resolve fine scale details and there is no added value in including those details in the estimation problem. An attempt to estimate grid-block state and parameter values increases the computational cost of the EnKF update while possibly giving spurious and geologically unexpected outcomes.

The fact that our reduced state DCT parameterization gives comparable or even better results than a grid block-oriented approach has important computational implications. The dimensions of the DCT state vectors, which are $82\times3 + 192 = 438$ and $660\times3 + 31 = 2011$ for Reservoir A and B, respectively, are only 3.5% and 5.1% of the dimensions of the grid block-oriented states, which are $64\times64\times3 + 192 = 12480$ and $60\times220\times3 + 31 = 39631$. Since the DCT bases are calculated only once with a very efficient fast Fourier transform [4.25] the overhead cost associated with the parameterization is much less than the gained computational benefits.

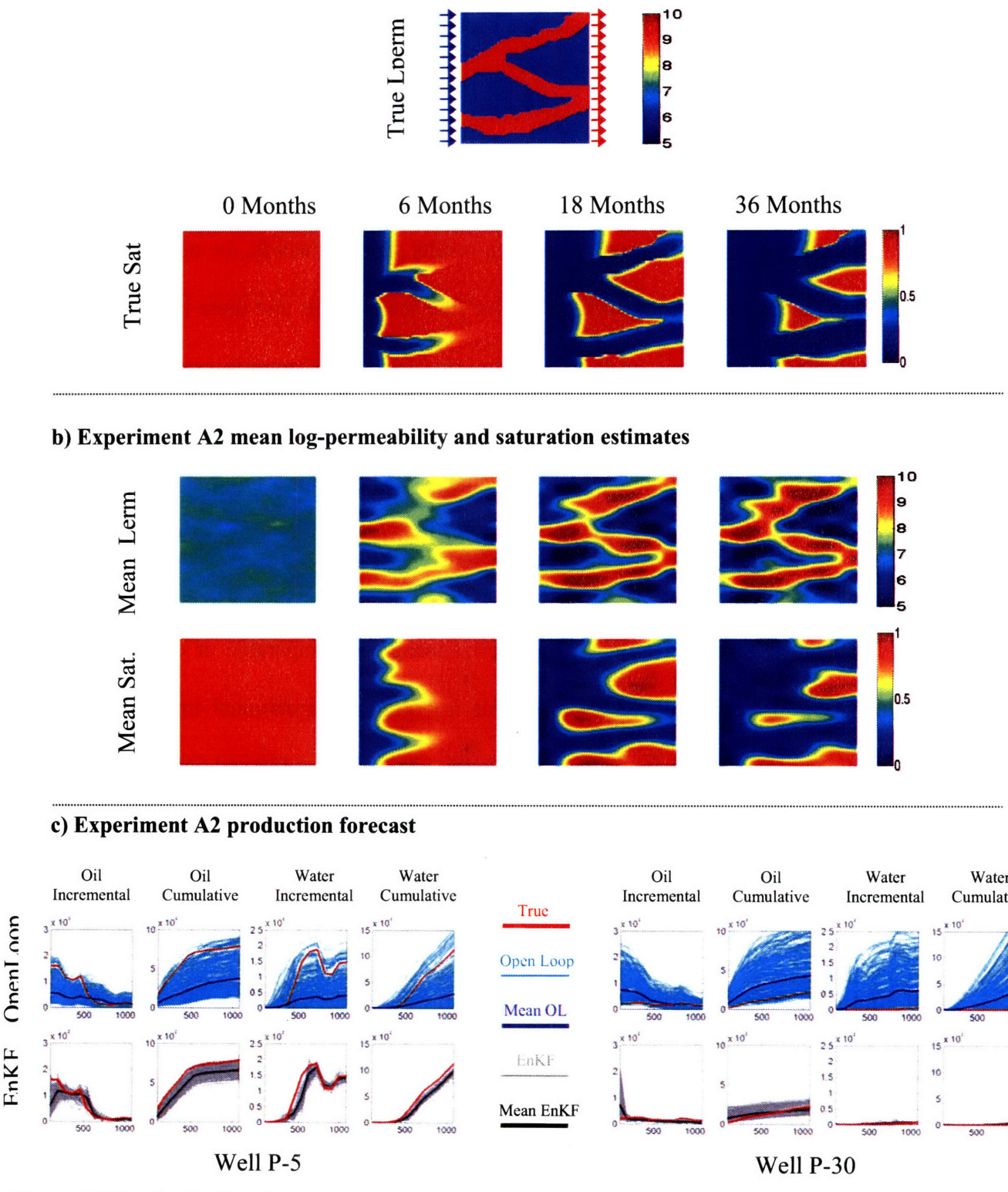
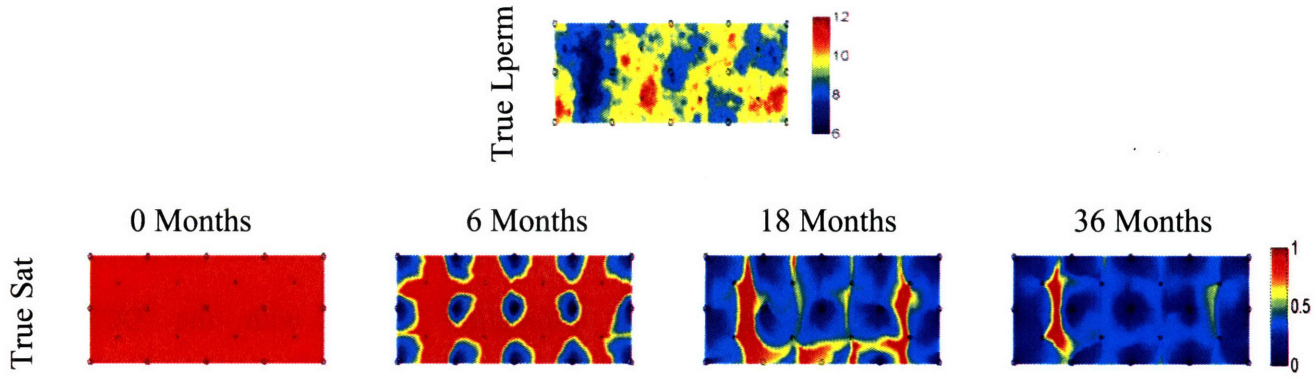
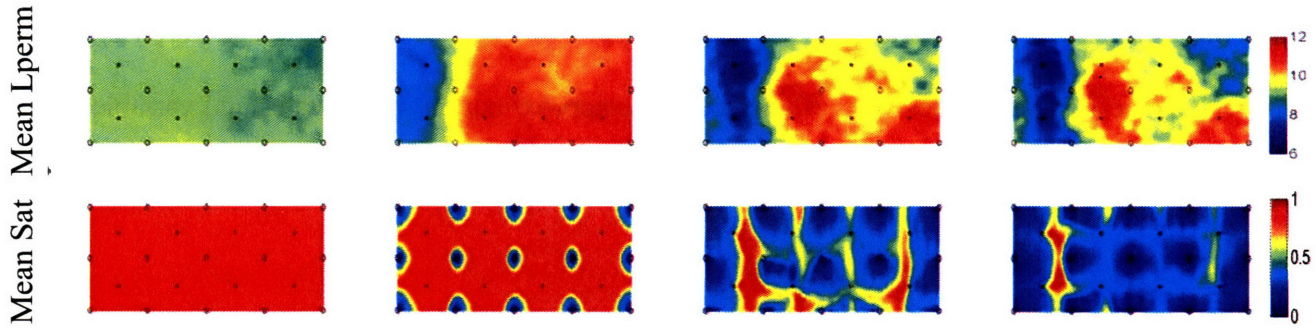


Figure 4.6 Results for Experiment A2 (estimating DCT coefficient with the EnKF for Reservoir A): (a) the true log-permeability field and snapshots of its corresponding synthetic true saturation profiles generated by *ECLIPSE* reservoir simulator; (b) the mean EnKF log-permeability and saturation estimates at selected time-steps using the DCT parameterization; (c) production observation (red), openloop forecast and its mean (cyan and blue), and EnKF forecast and its mean (grey and black).

a) Experiments B2 "True" permeability and saturations



b) Experiment B2 mean permeability and saturation estimates



c) Experiment B2 production forecast

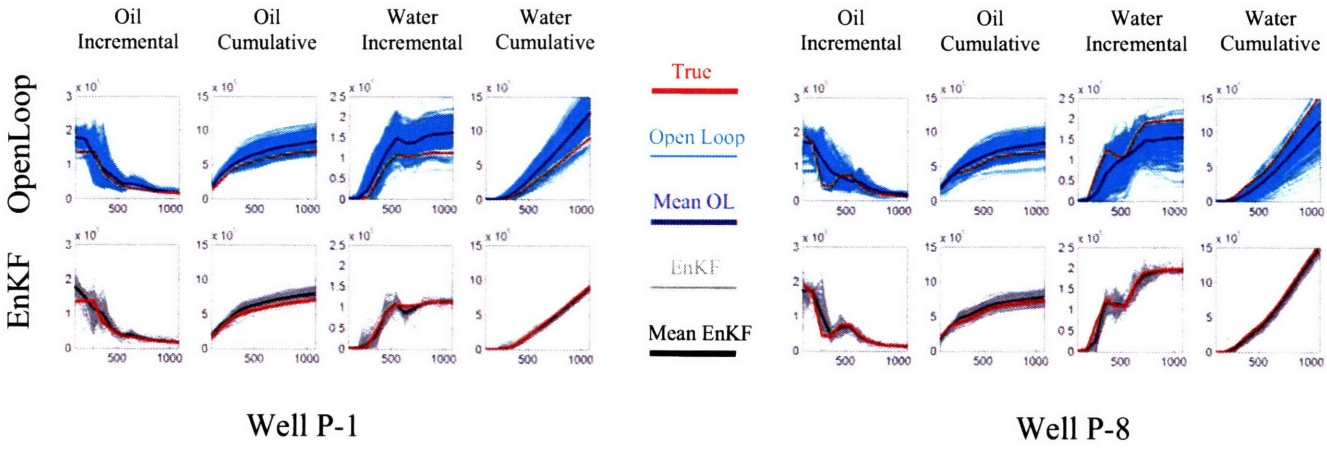


Figure 4.7 Results for Experiment B2 (estimating DCT coefficient with the EnKF for Reservoir A): (a) the true log-permeability field and snapshots of its corresponding synthetic true saturation profiles generated by *ECLIPSE* reservoir simulator; (b) EnKF mean log-permeability and saturation estimates using the DCT parameterization; (c) production observation (red), openloop forecast and its mean (cyan and blue), and EnKF forecast and its mean (grey and black).

4.6 Conclusions

In this chapter an efficient history matching approach was presented by combining the ensemble Kalman filter with a flexible and effective parameterization method, the Discrete Cosine Transform (DCT). Two waterflooding examples with structurally different geological formations show that estimation results with the DCT parameterization are almost identical to the results obtained with a much more expensive approach that estimates states in every block of the simulator computational grid. This implies that the computational effort spent in estimating individual grid-block properties is unnecessary and undesirable if an acceptable approximation of the states and variables can be obtained with smaller number of parameterized coefficients.

The approximation introduced by truncating the DCT representation of permeability seems to provide a smoother estimate of the original permeability field, as might be expected given the fact that smaller scale details are omitted in the truncated DCT. However, our examples suggest that the estimated saturations are less sensitive to approximations introduced by the truncated DCT, at least in determining the global pattern of oil displacement. This has important implications for the optimization and control of oil production operations, where knowledge of the movement of saturation profile is used to adjust reservoir control variables. Overall, it looks like a DCT version of the ensemble Kalman filter provides an accurate, efficient, robust, and flexible option for reservoir history matching, particularly for real-time operations.

References

4.1. Deutsch, C.V. (2002): *Geostatistical reservoir modeling.* Oxford University Press, New York.

4.2. Deutsch C.V., Journel A.G. (2002): *GSLIB: Geostatistical software library.* Oxford University Press.

4.3. Gavalas G.R., Shah P.C., Seinfeld J.H. (1976): "Reservoir history matching by Bayesian estimation". *Soc. Petrol. Eng. J.*, 16(6), 337–350.

4.4. Oliver D.S., Reynolds A.C., Bi Z., Abacioglu Y. (2001) "Integration of production data into reservoir models". *Petroleum Geosciences,* vol. 7, no. SUPP, pp. 65-73(9).

4.5. Aziz K., Settari A. (1979): *Petroleum Reservoir Simulation.* Applied Science Publishers LTD, London.

4.6. Reynolds A. C., He N., Chu L., Oliver D. S. (1996): "Reparameterization techniques for generating reservoir descriptions conditioned to variograms and well-test pressure data". *Soc. Petrol. Eng. J.,* 1(4) , 413-426.

4.7. Jacquard P., Jain C. (1965): "Permeability distribution from field pressure data". *Soc. Pet. Eng. Journal,* 281-294.

4.8. Jafarpour B., McLaughlin D., (2007): "Efficient permeability parameterization with discrete cosine transform". Paper SPE 106453 in the proceedings of the *SPE Reservoir Simulation Symposium*, Houston, TX.

4.9. Sahni I., Horne R.N. (2005): "Multiresolution wavelet analysis for improved reservoir description". *SPEREE*, 8(1), 53-69.

4.10. Sarma P., Durlofsky L., Aziz, K. (2007): "A new approach to automatic history matching using Kernel PCA". paper SPE 106176 in the proceedings of the *SPE Reservoir Simulation Symposium,* Houston, TX,.

4.11. Evensen G. (1994): "Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics". *J Geophys. Res.* 99(C5): 10 143–10 162.

4.12. Kalman R.E. (1960): "A new approach to linear filtering and prediction problems". *Transactions of the ASME--Journal of Basic Engineering.* vol. 82, pp. 35-45.

4.13. Evensen G. (2003): "The ensemble Kalman filter: theoretical formulation and practical implementation". *Ocean Dyn.* 53, 343– 367.

4.14. Nævdal G., Johnsen L.M., Aanonsen S.I., Vefring E.H. (2003): "Reservoir monitoring and continuous model updating using ensemble Kalman filter". Paper SPE-84372.

4.15. Gu Y., Oliver D.S. (2005): "History matching of the PUNQ-S3 reservoir model using the ensemble Kalman filter". Paper SPE 89942.

4.16. Wen X.H., Chen W. (2005): "Real-time reservoir model updating using ensemble Kalman filter". Paper SPE-92991 in the *proceedings of the SPE Reservoir Simulation Symposium*, pp. 1 – 14.

4.17. ECLIPSE 100 (Black Oil), Reference manual and technical description, Schlumberger GeoQuest, Houston (2006).

4.18. Jafarpour B., McLaughlin D.: 'Estimating reservoir permeabilities with the ensemble Kalman filter: the importance of the ensemble design". in review, *SPE Journal.*

4.19. http://www.spe.org/csp/

4.20. Strebelle S., Journel A.G. (2001): "Reservoir modeling using multiple-point statistics". paper SPE 71324 in the proceedings of the *SPE Annual Technical Conference and Exhibition*, New Orleans.

4.21. Remy N. (2004): "S-GeMS: A geostatistical earth modeling library and software". Ph.D. thesis, Stanford University.

4.22. Ahmed N., Natarajan T., Rao K.R. (1974): "Discrete cosine transform". *IEEE Trans. Computers*, 90-93.

4.23. Rao K.R., Yip P. (1990): "Discrete cosine transform: Algorithms, advantages, applications". Academic Press, Boston.

4.24. Gonzalez R.C., Woods R.E. (2002): *Digital image processing*. 2nd ed., Prentice Hall, Upper Saddle River, NJ.

4.25. Brigham E. O. (1988): *The fast Fourier transform and its applications*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 448 pp.

Chapter 5

History Matching of the 3-Dimensional SPE-10 Reservoir Model using EnKF with DCT

5.1 Introduction

History matching of realistic three-dimensional reservoirs is challenging due to the large number of unknowns relative to available measurements, the complexity in the reservoir facies structure due to variability in depth, and the associated computational burden. Realistic reservoirs often present complex heterogeneity in three dimensions, which can make inverse and forward modeling more demanding. In particular, retrieval of three-dimensional geological features (such as faults and channels) from sparsely measured state variables and poorly known priors is difficult.

A viable history matching approach should be able to provide reasonable performance under realistic field settings. Therefore, the history matching framework developed in

previous chapters is tested with a three-dimensional virtual reservoir model in this chapter. The example model is taken from a widely used data set [5.1] by the *Society of Petroleum Engineers* known as SPE-10 model. This reservoir model was originally developed for up-scaling studies [5.2] and has interesting geological characteristics that are described in the next section.

5.2 SPE-10 Reservoir Model

The 2001 SPE Comparative Solution Project [5.1, 5.2] was a project proposed by the Society of Petroleum Engineers to provide benchmark datasets that could be used to compare the performance of different reservoir simulators and up-scaling algorithms. Specifically, the SPE-10 model was developed by Christie et al. [5.2] to compare performance (accuracy and computational cost) of different up-scaling approaches in a waterflood example with over a million cell geological model. In this study, an up-scaled version of the SPE-10 model is used for history matching as described in the following section.

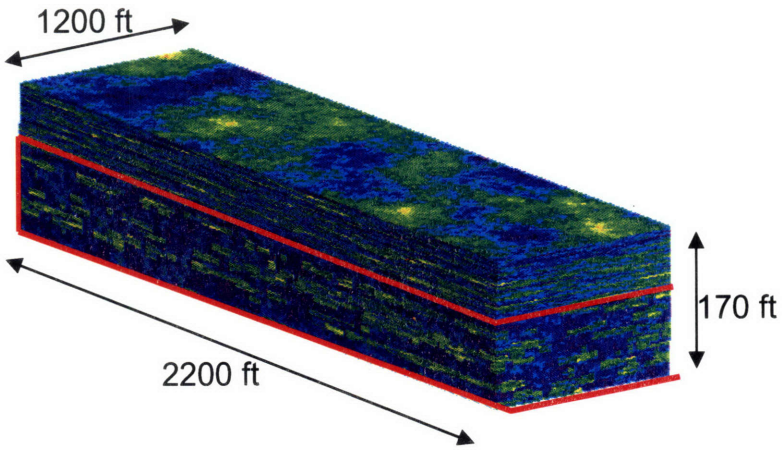
5.2.1 Original SPE-10 Model

The original SPE-10 model represents a heterogeneous reservoir with dimensions 1200 × 2200 × 170 ft in a Cartesian coordinate system (Figure 5.1). It comprises a Brent sequence with two distinct formations in depth. The model is discretized into a 60 × 220 × 85 grid block mesh with the top 35 layers (70ft) consisting of Tarbert formation (representing a prograding near shore environment) and the bottom 50 layers (100ft) including an Upper Ness formation (resembling fluvial deposits). The

fine scale model contains 1,122,000 grid blocks of size 20 × 10 × 2 ft. In the original model, the porosity varies between 0 and 0.5 and the horizontal permeability varies over eight orders of magnitude (6.65×10^{-4} - 2×10^{4} mD). In the channelized part of the reservoir (bottom 50 layers) the ratio of vertical to horizontal permeabilities inside and outside the channels are 0.3 and 10^{-4}, respectively. This causes vertical fluid displacement only through channels.

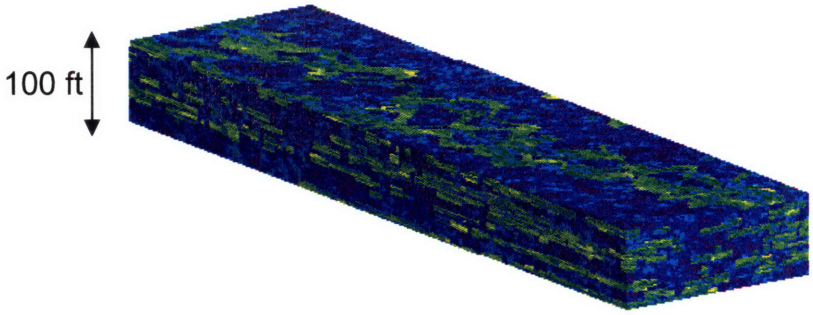
The connate water saturation is 0.2 and the initial pressure is 6,000 psi, assumed constant through the reservoir. A five-spot waterflooding well configuration with one injector in the middle and four producers (one on each corner) is considered in the original model. All wells are vertical and fully completed through the entire formation. The injection well is operated under rate control with injection rate of 5,000 bbl/day (under reservoir condition) and a maximum pressure of 10,000 psi while the production wells are kept under pressure control with a bottom-hole pressure specified at 4,000 psi. A two phase (dead oil) waterflooding simulation for 2,000 days is considered. More details including oil, water, and rock properties as well as relative permeability functions can be found in [5.1, 5.2]. In this Chapter a modified up-scaled version of the SPE-10 model is used for history matching with the EnKF and DCT parameterization.

a) Original SPE-10 model



$60 \times 220 \times 85 = 1{,}122{,}000$

b) Bottom 50 layers of SPE-10 model



$60 \times 220 \times 50 = 660{,}000$ blocks

Figure 5.1 SPE-10 model: a) original model with porosity distribution shown; b) bottom 50 layers resembling a channelized fluvial environment.

5.2.2 Modified SPE-10 Model

The original SPE-10 model has over a million grid blocks and is developed to test different up-scaling algorithms. The channeling structure of the bottom 50 layers in this model provides vertical connectivity only within the channels (vertical permeability values outside the channels are insignificant), which results in three dimensional fluid displacement inside the channel network. While this makes the SPE-10 model an interesting three-dimensional virtual reservoir for history matching studies, the existing contrast in the permeability field presents a challenging flow simulation problem that can take a long time to solve with grid-based simulation techniques on a PC machine. Therefore, the original SPE-10 model was computationally too demanding to consider for a history matching study in this thesis. Furthermore, the well configuration in the original model provided sparse observation points within the reservoir, which can be insufficient for constraining the reservoir model accurately. Because of these issues in the original SPE-10 model a modified version of it was considered for history matching in this chapter.

To mitigate the computational burden of running the SPE-10 model an up-scaled version of the channelized fluvial formation (bottom 50 layers) was used in this study. Following [5.3], a simple geometric averaging for log-permeability and arithmetic averaging for porosity were applied in up-scaling the original model. An up-scaling factor of 2 was used in horizontal direction while the up-scaling ratio in vertical direction was 5. This reduced the number of grid blocks in the model to 33000 (= 30 × 110 × 10). The size of

each grid block was adjusted to 40 × 20 × 10 ft in order to preserve the original volume of the reservoir section. An up-scaled version of the true reservoir (with vertical permeability distribution) is shown in Figure 5.2.

In addition to up-scaling, a new 13-spot well configuration was considered to provide a better observability for the history matching problem (Figure 5.3). In the new configuration, the wells were completed in all 10 vertical layers and it was assumed that observations of rate and pressure were available at each individual layer.

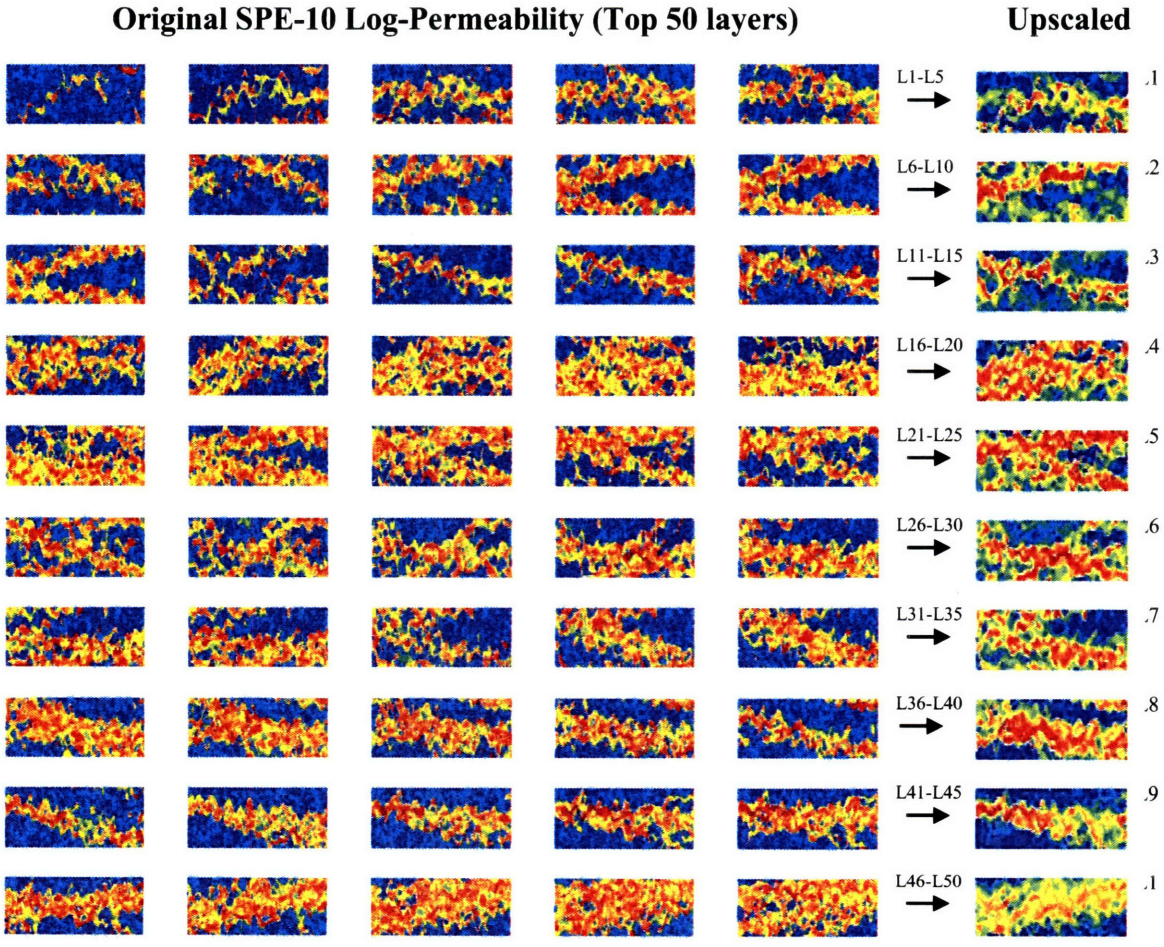
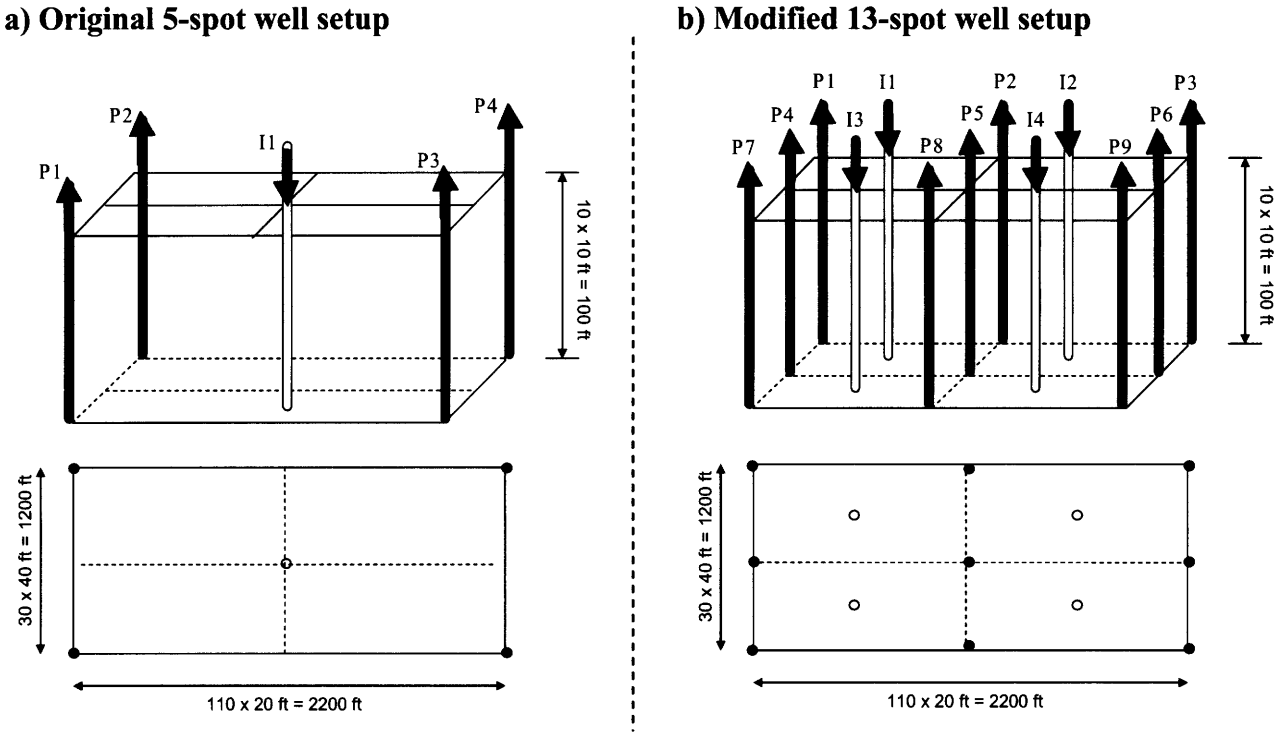


Figure 5.2 Original (first 5 columns) and up-scaled (last column) horizontal log-permeability in the SPE-10 model. Original model size (bottom 50 layers) is 60 × 220 × 50, which is up-scaled to a 30 × 110 × 10 model.



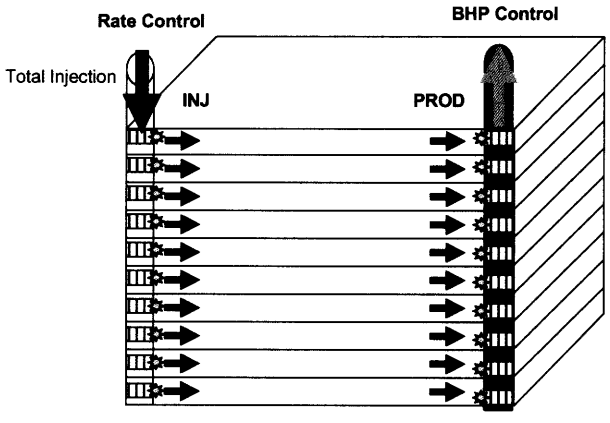


Figure 5.3 Reservoir well configurations for a) original 5-spot well pattern in SPE-10 model: b) a 13-spot well distribution in the modified SPE-10 model used for history matching; c) well completion assuming each connection acts independently and rates and pressures can be measured at each port.

5.3 Three-Dimensional DCT Parameterization

The DCT parameterization approach described in Chapter 3 can be easily extended to three dimensional problems. This is due to the separability property of the DCT bases that reduces the transformation of an *N*-dimensional signal to a sequence of *N* one-dimensional transforms [5.4-5.6]. This procedure is briefly described in this section.

5.3.1 Three Dimensional DCT Bases

Following the one dimensional notation introduced in Chapter 3, the three-dimensional DCT transform of a signal *u(m,n,p)* can be represented as:

$$v(i,j,k) = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\sum_{k=0}^{P-1} a(i,j,k;m,n,p).u(m,n,p) \qquad (5.1)$$

where *M*, *N*, *P* is the number of grids in *x*, *y*, and *z* directions; *a(i,j,k; m,n,p)* is the three dimensional basis function that transforms a point *(i,j,k)* in original domain to a point in the transform coordinates *(m,n,p)*; *v(i,j,k)* are the coefficients representing the signal in the new transformed coordinate system.

As in (3.1), the bases of the transformation can be written as follows:

$$a(i,j,k;m,n,p) = \alpha(i,j,k)\cos\left[\frac{\pi(2m+1)i}{2M}\right]\cos\left[\frac{\pi(2n+1)j}{2N}\right]\cos\left[\frac{\pi(2p+1)k}{2P}\right] \qquad (5.2)$$

where $\alpha(i,j,k)$ can be written as

$$\alpha(i,j,k) = \sqrt{\frac{2^{(\mathrm{sgn}(i)+\mathrm{sgn}(j)+\mathrm{sgn}(k))}}{MNP}} \tag{5.3}$$

The three-dimensional basis functions for a cube of size 8 are shown in Figure 5.4. A linear combination of these bases can be used to approximate any volume of size 8 (or any 8 frames of 8 × 8 images in a video file--where time is used as the third dimension) [5.4-5.6]. In practice, implementation of the 3D-DCT using equations (5.1-5.3) is computationally inefficient and separability of the bases is used to provide a simpler and more efficient one-dimensional implementation. Using the separability of the DCT bases equation (5.1) can be reduced to:

$$v(i,j,k) = \sum_{i}^{M1} a(i,m)\{\sum_{j}^{N} a(j,n)\{\sum_{k}^{P} a(k,p).u(m,n,p)\}\} \tag{5.4}$$

Here the three-dimensional basis functions *a(i,j,k; m,n,p)* are replaced with the following one-dimensional basis functions *a(i;m), a(j;n), a(k;p)* that implement the transformation sequentially:

$$a(i;m) = \sqrt{\frac{2^{\mathrm{sgn}(i)}}{M}} \cos\left[\frac{\pi(2m+1)i}{2M}\right]$$

$$a(j;n) = \sqrt{\frac{2^{\mathrm{sgn}(j)}}{N}} \cos\left[\frac{\pi(2n+1)j}{2N}\right] \tag{5.5}$$

$$a(k;p) = \sqrt{\frac{2^{\mathrm{sgn}(k)}}{P}} \cos\left[\frac{\pi(2p+1)k}{2P}\right]$$

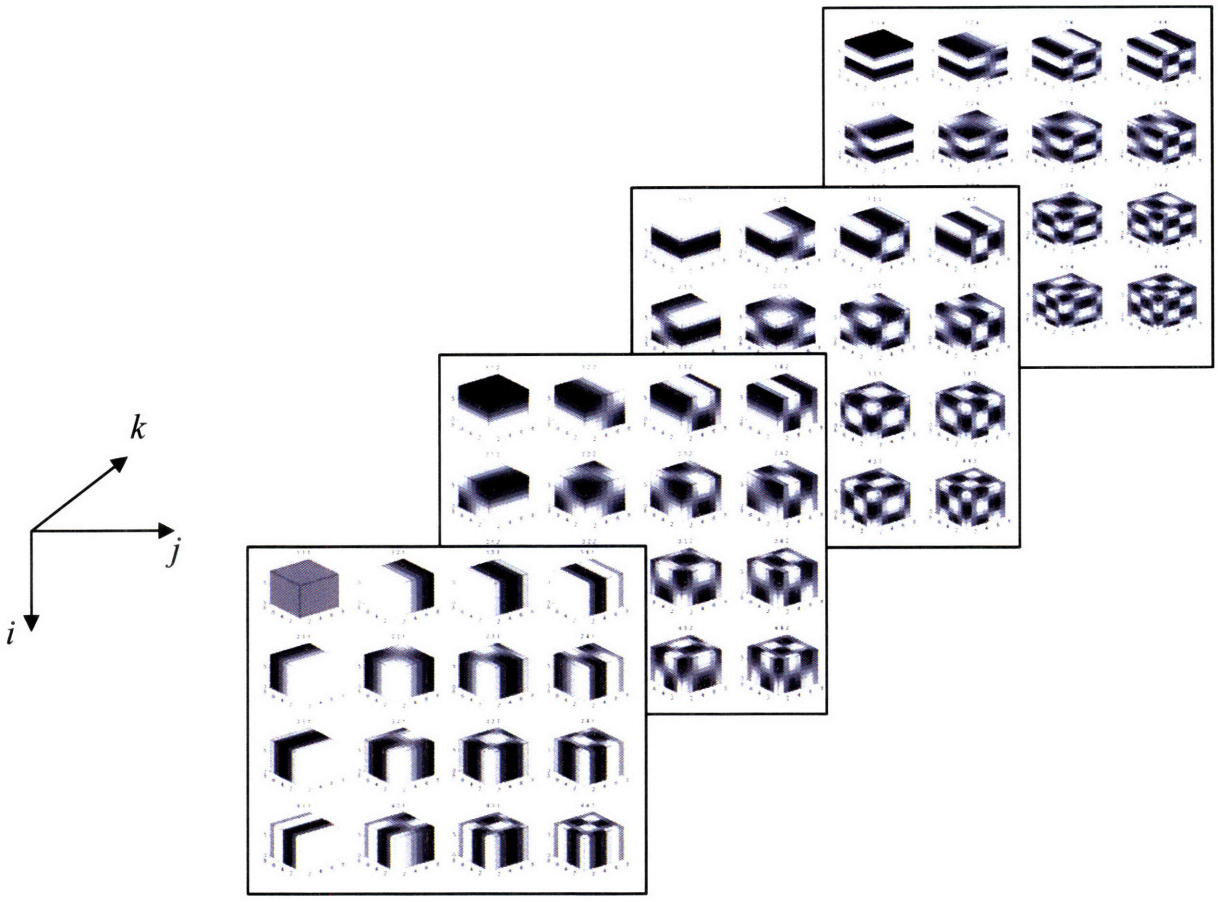


Figure 5.4 Three-dimensional DCT bases for a cube of size 8. The top figure shows bases variations in Cartesian coordinate and the four panels on the bottom show the first 64 low frequency bases.

Implementation of the 3D-DCT in the following section is carried out using the one-dimensional bases of equation (5.4-5.5) and the three-dimensional bases that are shown in Figure 5.4 are for illustration purpose.

5.3.2 Representation of the SPE-10 with the 3D-DCT

To illustrate the effectiveness of the 3D-DCT, the modified SPE-10 example of Figure 5.2 is approximated in this section. Figure 5.5 shows the result of applying the 3D-DCT to represent the modified SPE-10 horizontal permeability field. The first column shows the permeability field in all 10 layers before transformation. The second column shows the logarithm of the magnitude of DCT coefficients after the 3D-DCT transformation. It should be notated that the logarithm is used for better visibility of the spread in the coefficients; and contrast in the original significant coefficients is more pronounced. Columns three to five show the 3D-DCT approximation of the permeability volume in the first column using 0.1%, 1%, 5%, of the *significant* coefficients, respectively. The last column shows the approximation with 5% of the *leading* (prior-trained) coefficient.

For history matching purpose in the later sections of this chapter the approach in Chapter 4 is followed, where all the state and parameters are parameterized before the EnKF update step using prior-trained bases of size equal to 5% of the original coefficients. Therefore, the last column in Figure 5.5 corresponds to the best achievable horizontal permeability estimate in the history matching problem. The augmented state vector for the EnKF update includes predicted saturations, pressures, horizontal permeabilities, vertical permeabilities, porosities, and observations at well locations. After DCT transformation, the size of the augmented state vector reduces from ~165,000 (= 5 × 33000) to ~6750 (= 5 × 1350).

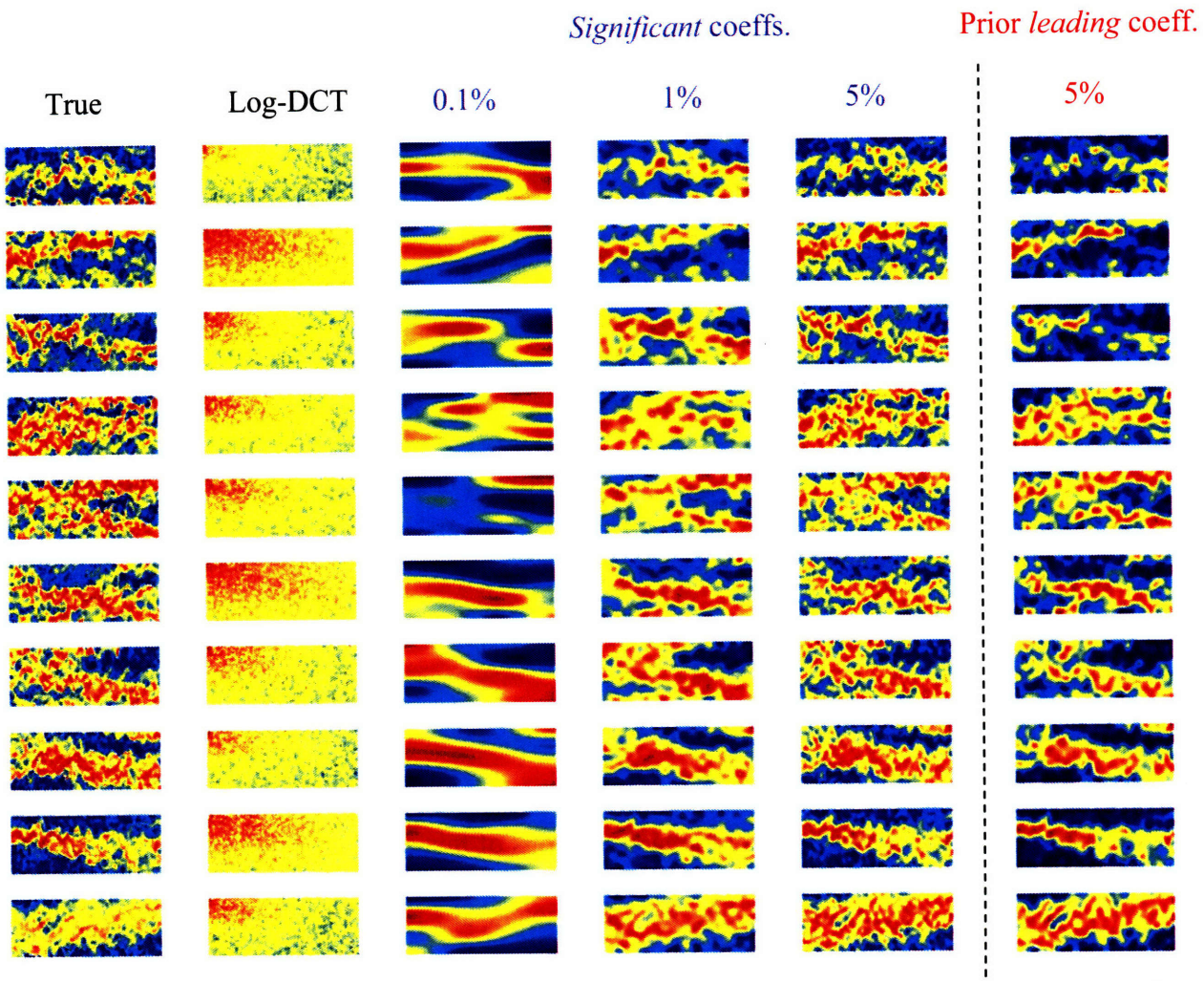


Figure 5.5 Three-dimensional DCT approximation of horizontal permeability for modified SPE-10 model using increasing number of *significant* coefficients and prior-trained leading coefficients. Modified SPE-10 model (column 1), logarithm of magnitude of DCT coefficients (third column), and 3D-DCT approximation of the horizontal permeability in the first column with *significant* DCT coefficient size of .1%, 1%, and 5% of original coefficients (third to fifth columns, respectively). The last column shows the approximation using 5% of leading (prior-trained) DCT coefficients (best achievable estimate in the history matching example).

In applying the 3D-DCT, training of the bases with prior information for each state or parameter is, in general, different. However, in history matching applications (as seen in Chapter 4) the same bases used to represent the permeability field can be used to approximate other parameters and states. This approximation is reasonable because the important features in states (pressure and saturation) are strong functions of the permeability field. In addition, parameters such as porosity are correlated with the permeability field and are expected to behave similarly. This assumption has been used in the history matching section of this chapter.

5.4 Prior Reservoir Ensemble

The prior model for uncertain parameters of history matching, i.e. horizontal and vertical permeabilities, and porosity are discussed in this section. In generating these priors, a training-image-based facies modeling was used as discussed in previous chapters and in [5.7]. In addition, hard data at well locations and a probability map (assumed to be available from seismic survey) were used to generate conditional facies realizations. The facies realizations were then used to generate initial replicates of petrophysical properties.

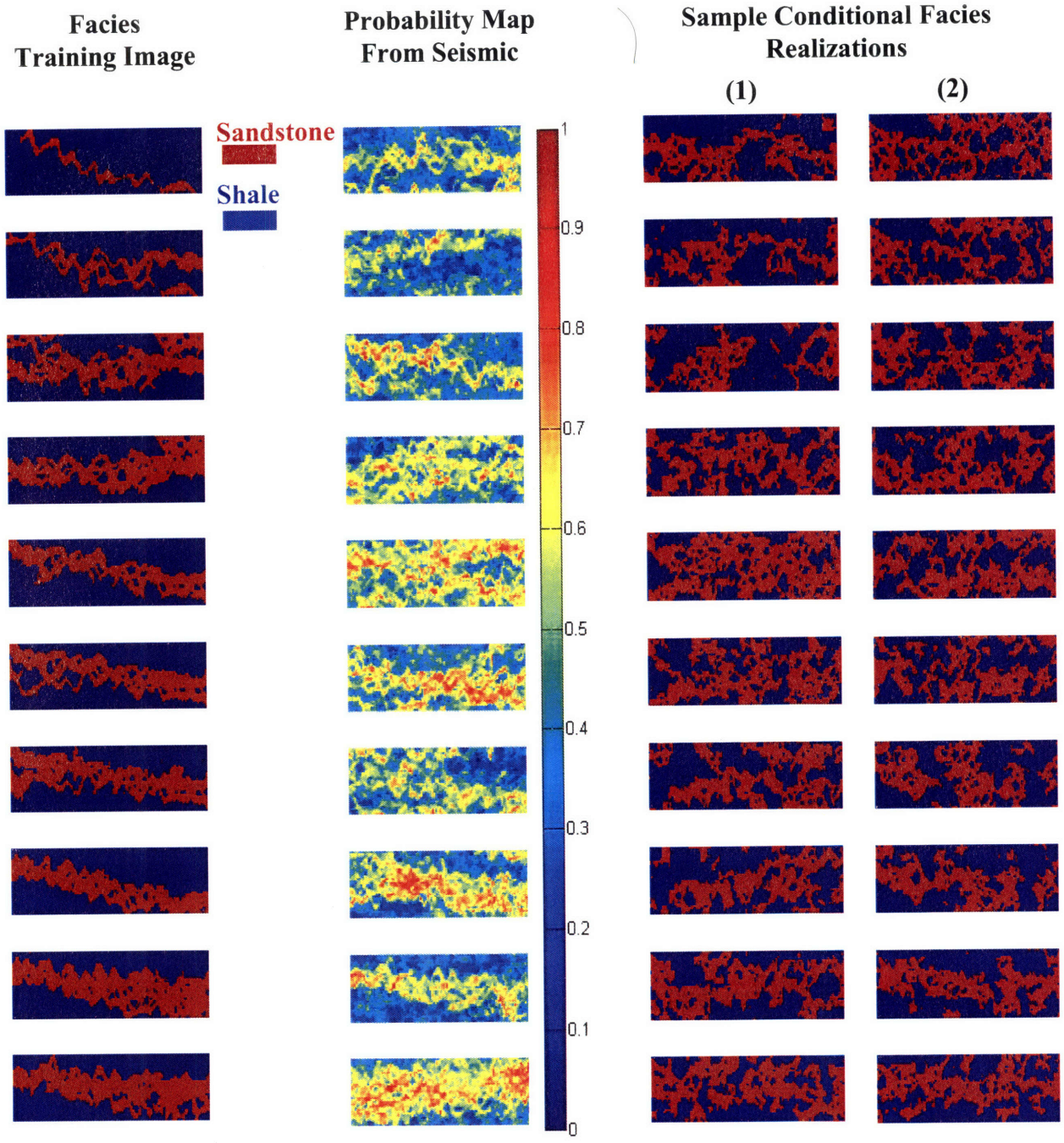


Figure 5.6 Prior conditional facies generation using a binary training image (first column) and a probability map from seismic (second column) and hard data at well locations (not shown). Two sample realizations are shown in third and fourth columns.

5.4.1 Facies Model

The training image used to generate the initial facies models through multiple-point geostatistics is shown in Figure 5.6. This training image assumes anisotropy in the facies orientation where commingled channels are mainly in east-west directions. Measurements of the permeability/porosity at wells were used to constrain the facies type at well locations. In addition, a probability map was used to guide facies generation from the training image. Probability maps contain information about the likelihood of occurrence of different lithofacies at a given location. They can be used as soft data, along with hard measurements from well bore-holes and logs, to condition generation of facies models.

In practice, probability maps can be obtained from geological interpretation of depositional information and seismic studies [5.8]. However, in the synthetic example of this section the probability map was generated by perturbing and rescaling the true log-permeability field. The true log-permeability was first scaled to [0 1] range; then correlated random noise with 0.0 to 0.5 range of variation was added to introduce uncertainty to the probability map. The final probability map was obtained by rescaling the result to cover a range of [0.2 0.8]. This range was selected to avoid unrealistically high certainty (>80%) in inference of facies distribution from seismic data. The process of generating the probability map in this example is shown in Figure 5.7.

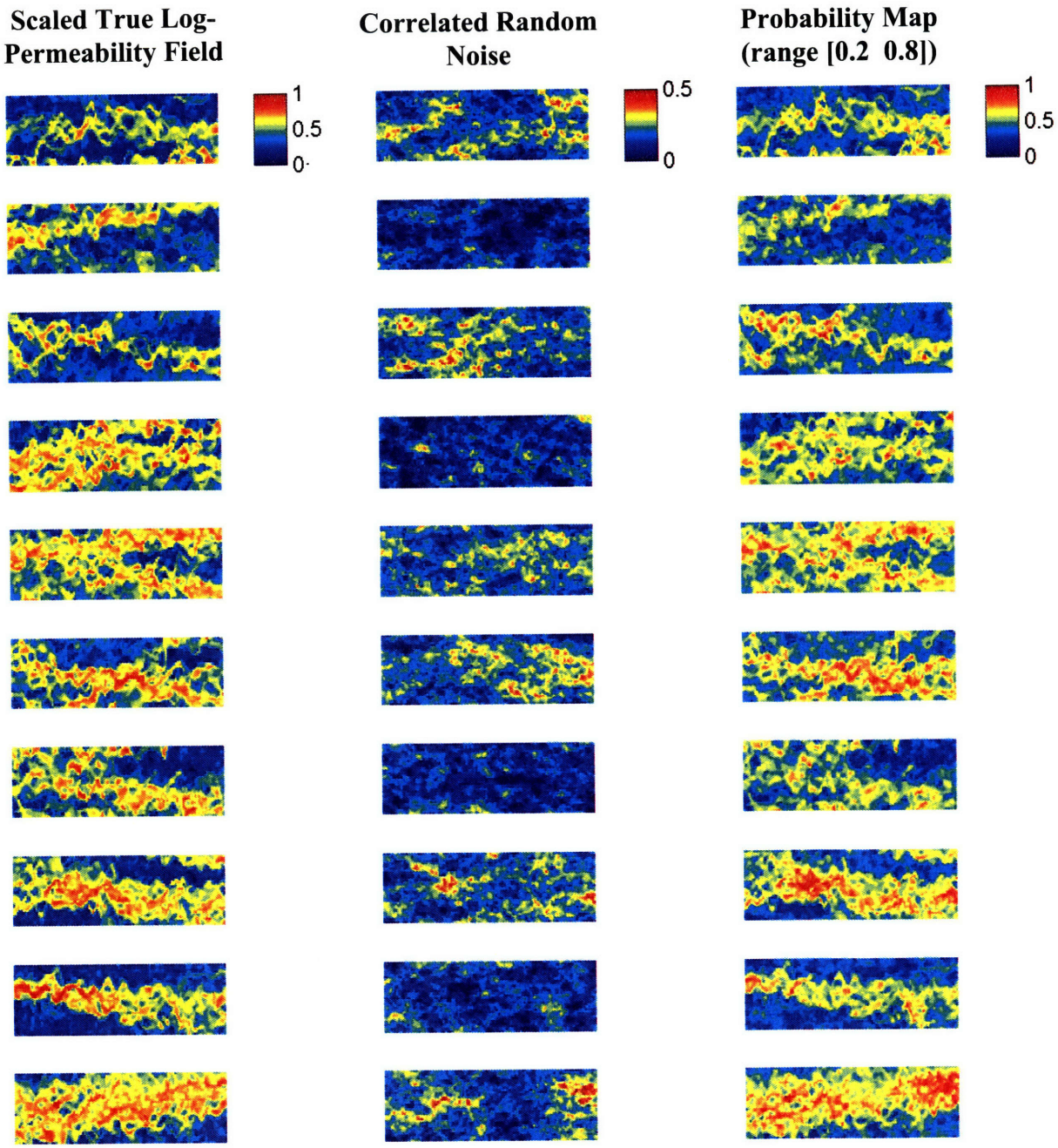


Figure 5.7 Probability map used for conditioning facies generation on soft seismic data. The first column shows the true log-permeability field rescaled to vary between 0 and 1; the second column shows a correlated random noise that is added to the true log-permeability to account for uncertainty in seismic data; and the third column shows the final probability map that ranges between 0.2 and 0.8 to reflect uncertainty in seismic data.

The *SNESIM* algorithm [5.9] was used to generate facies realizations conditioned on hard data at well locations and the probability map described above. The fraction of channel and non-channel facies types was specified as 0.30 and 0.70, respectively. Two sample realizations of the initial facies are shown in Figure 5.6. It is important to note that even though the probability map was used to guide facies generation, the resulting equi-probable facies realizations had different shapes and connectivity structures. This was due to the modest range of probability map ([0.2 0.8]) to allow for more uncertainty in facies generation while preserving connectivity in the desired directions.

The facies realizations generated using the above procedure was then used to build prior petrophysical properties. The variability within each facies type was modeled using correlated Gaussian random fields. It was assumed that the variability within each facies type is independent of that in the other facies type. This is a realistic assumption as the processes responsible for deposition and formation of these facies are fundamentally different and result in distinct intra-facies variabilities.

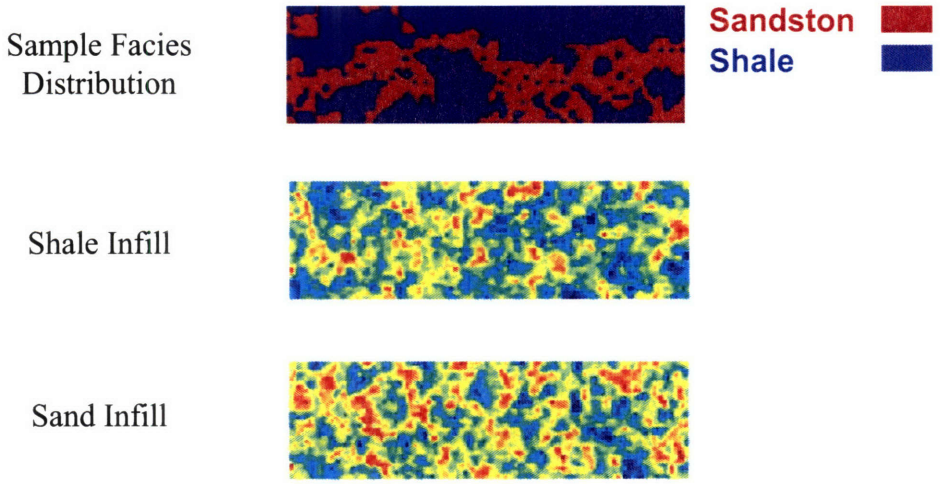
5.4.2 Prior Permeability and Porosity Models

The prior replicates for the permeability and porosity were generated from the initial facies models. The correlation between the porosity and permeability was introduced through the facies model; however, perfect correlation was avoided by choosing different

variability for permeability and porosity within each facies. Unconditional isotropic Gaussian random fields with zero-mean, unit variance, and isotropic correlation length of 3 pixels were generated to represent variability in channel and non-channel facies for permeability and porosity. Horizontal and vertical permeability replicates were generated from the same facies model; however, the horizontal permeability inside and outside the channels were 10 and 1000 times larger than those in vertical permeability. This provided vertical connectivity mainly through the channels. The final porosity was rescaled to vary in the range 0.0-0.40 while the permeability range in horizontal and vertical directions were $[10^{-1}\text{-}10^{4}]$ mD and $[10^{-3}\text{-}10^{3}]$ mD, respectively

Figure 5.8 shows sample horizontal and vertical log-permeabilities, and porosity generated using the above procedure. The common feature among these petrophysical properties is the facies distribution. However, the variability within the sand and shale facies distinguishes the permeability and porosity fields. The contrast in the vertical permeability inside and outside the channels is more pronounced than that in horizontal permeability or porosity as explained earlier. An ensemble including 200 of these initial replicates was used for history matching.

a) Facies map and variability inside and outside channels



b) Petrophysical properties generating from facies variability models.

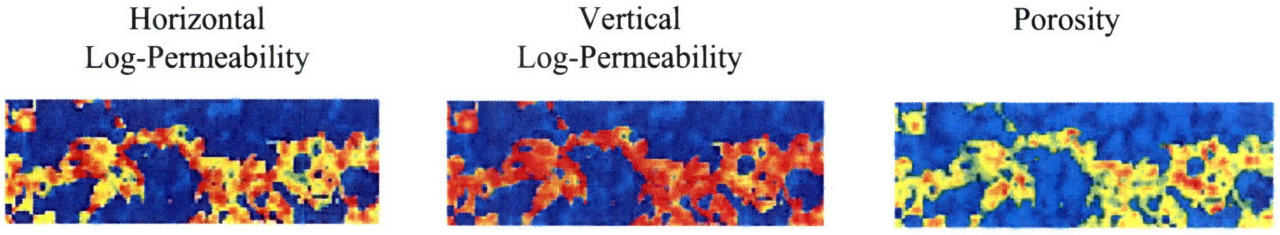


Figure 5.8 Cookie cutting procedure: top panel shows one sample layer of the prior facies, a layer showing variation inside shale and sand facies, respectively. The bottom panel shows the resulting horizontal (left) and vertical (center) log-permeability, and the corresponding porosity (right).

5.5 History Matching with Modified SPE-10 Model

This section presents the history matching results using the modified SPE-10 model with the prior permeability and porosity models described in the previous section. General information about the virtual field setup and data assimilation design is given first followed by the history matching results and discussion.

5.5.1 General Simulation and Assimilation Information

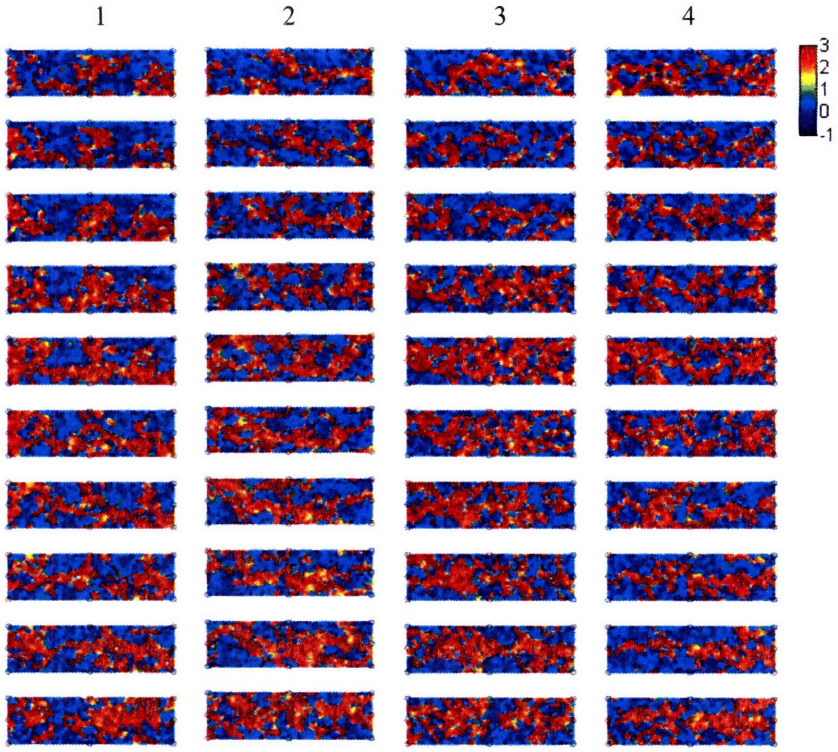
A total simulation time of 2000 days was considered in this study. Observations were integration into reservoir model after 10, 30, 60, 90, 180, 720, 1260, and 2000 days. A total of 13 wells (9 production and 4 injection wells, see Figure 5.3b) were completed in all 10 layers of the reservoir. As shown in Figure 5.3c the wells had connections in all 10 layers and flow and pressure measurements at each connection was used in the history matching. During the entire simulation/assimilation time a water injection rate of 2000 barrels/day was specified at each injection well while the production wells were operated under a bottom-hole pressure of 4000 psi. Fluid and rock properties, including relative permeability model, remained the same as the original SPE-10 model. The observations were derived from forward simulation of the modified SPE-10 model properties with the stated control variables. Uncorrelated Gaussian random noise with variance equal to the larger of 200 psi or 50% of pressure variation (relative to initial pressure of 6000 psi) for pressure observations and the larger of 100 barrels/day or 50% of the forecast rate values were assumed.

5.5.2 Results and Discussion

This section summarizes the result of history matching using the EnKF and DCT with the modified SPE-10 model. Figures 5.9 and 5.10 show the horizontal log-permeability results. In Figure 5.9a prior permeability replicates are shown while Figure 5.9b shows the same replicates after EnKF update. Each column represents a sample replicate with rows corresponding to the layers in the model. Visual inspection of the updates indicates that the updated ensemble has less spread than the prior. While this is expected due to the constraining effect of the measurements, it can only be used as an indication of good filter performance when updates are confirmed to be closer to the true values using a specified measure. A straightforward way to demonstrate this is through analysis of the production plots and breakthrough curves. This is done in the next section.

Figure 5.10 shows the results for the mean of the ensemble before and after the update as well as the true log-permeability and its best DCT approximation with the prior-trained coefficients and their corresponding bases. The true log-permeability (first column), the DCT approximation with 5% of coefficients (second column), the initial ensemble mean (third column), and the EnKF mean estimate after 2000 days (fourth column) are shown. The estimation results should only be compared to the second column (DCT approximation) in Figure 5.10 knowing that even if the estimation performance is optimal this is the best result that the estimation can be expected to provide.

a) Prior Sample Log-Permeability (Horizontal) Replicates



b) EnKF Updated Sample Log-Permeability (Horizontal) Replicates

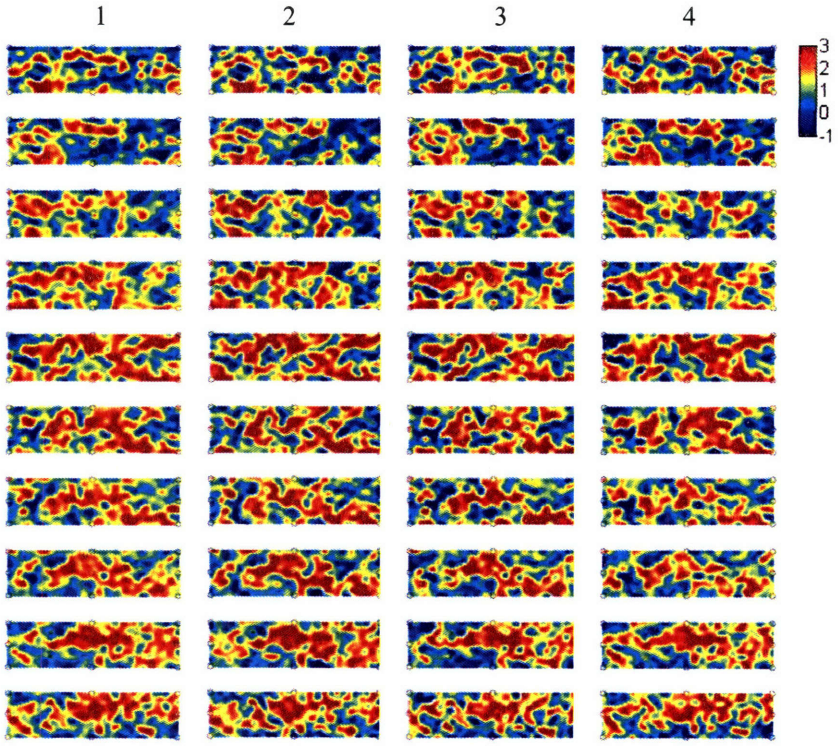


Figure 5.9 Horizontal log-permeability estimation results after 2000 days: a) four sample prior replicate; b) EnKF updated replicates corresponding to the samples in part (a). Rows correspond to layers 1-10 respectively.

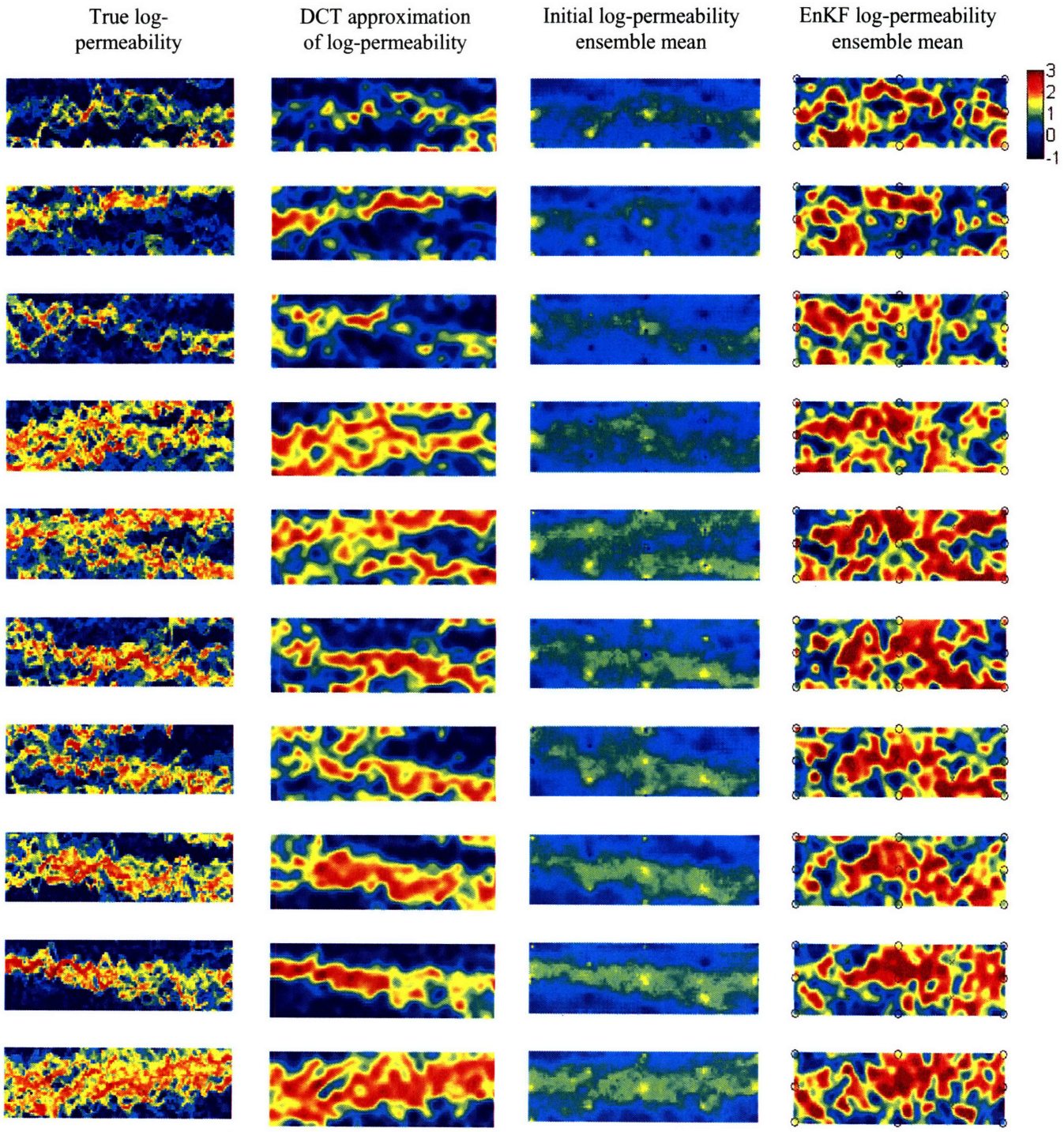


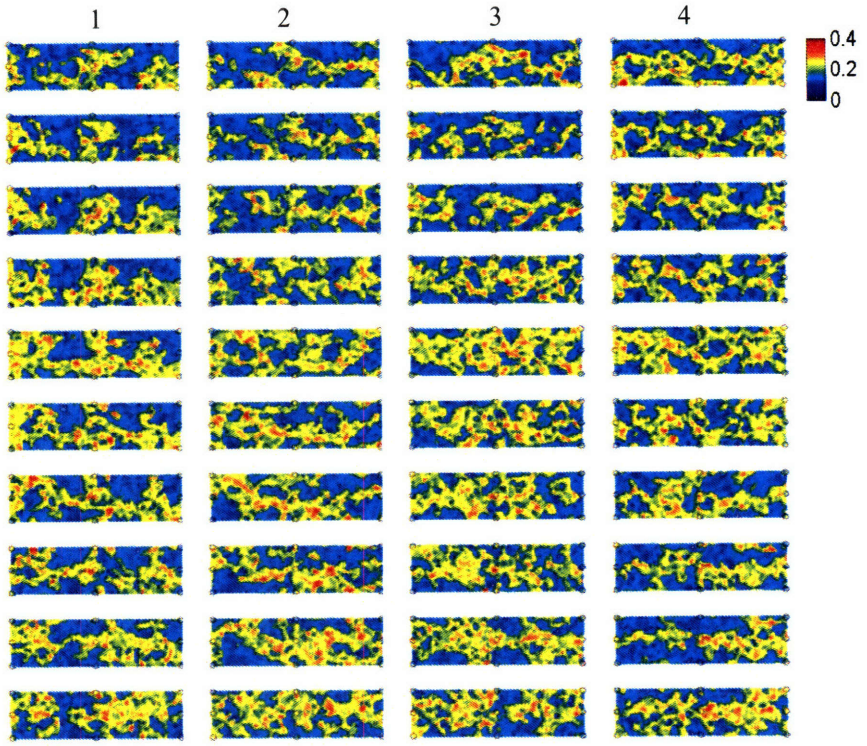
Figure 5.10 Horizontal log-permeability estimation results after 2000 days: the true log-permeability (first column), the DCT approximation of the true permeability field using significant bases and coefficients (second column), initial log-permeability ensemble mean (third column), and final EnKF log-permeability estimates mean (fourth column) are shown. Rows correspond to layers 1-10 respectively.

While it is difficult to comment on the quality of the estimates based on visual inspection, the overall high permeability and low permeability trends are more visible in the updates compared to prior. Connectivity in the vertical directions is an important aspect that is not easily inferred from the plots shown in here. As for horizontal connectivity, in some layers the channel connectivity is detected while in some others the channels exhibit disconnectivity. Part of this disconnectivity can be seen in the true field. This is, in part, due to up-scaling in which some of the channels have been combined with non-channels and resulted in local discontinuity. An additional contribution is the limited spatial content of the measurements that seems insufficient to constrain the DCT coefficients. In general, the estimation results are more accurate at the vicinity of the wells than far from them. Similar results are obtained for vertical permeability. A major difference between this example and examples from the previous chapters is that in here control variables (rates/pressures) are fixed and do not change in time. This introduces less excitation in the reservoir, which is expected do decrease the information content of the measurements.

Figures 5.11 and 5.12 show the results for the porosity. Figure 5.11a contains sample prior replicates for porosity and Figure 5.11b shows the same replicates after the EnKF update. Again in this case the ensemble spread has been reduced and only slight variations can be observed within the replicates. Figure 5.12 shows the true porosity (first column) and its DCT representation using significant coefficients (second column) as well as the prior ensemble mean (third column) and EnKF updated ensemble mean

(fourth column). In some regions the updated mean porosity appears to be somewhat overestimated and the exact shapes of continuous features are roughly captured. However, the performance of the filter is left to be assessed through a relevant quantitative measure such as production plots.

a) Prior Sample Porosity Replicates



b) EnKF Updated Sample Porosity Replicates

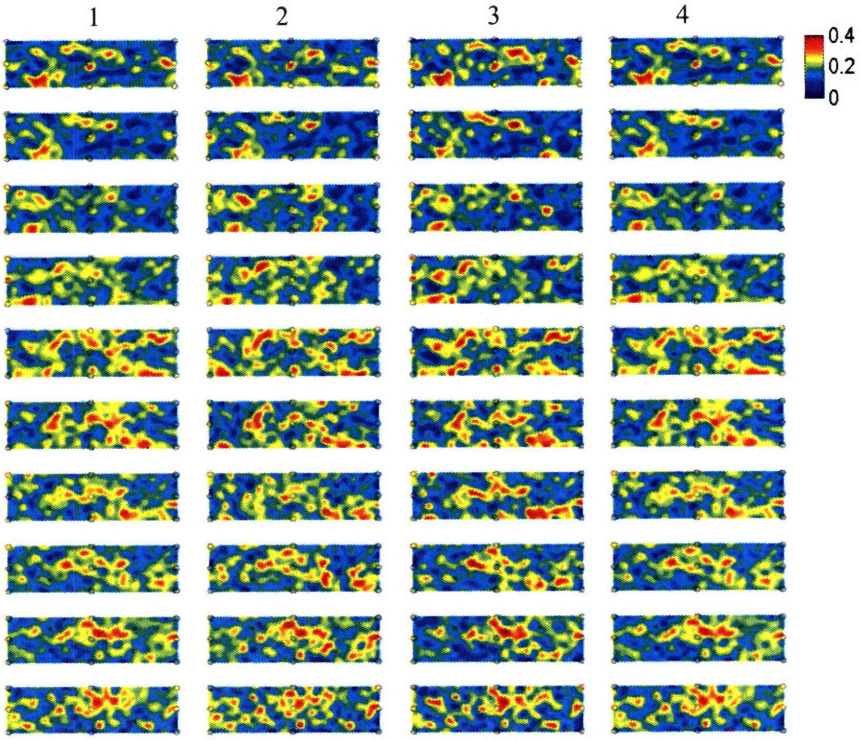


Figure 5.11 Porosity estimation results after 2000 days: a) four sample prior replicates; b) EnKF updated replicates corresponding to the samples in part (a). Rows correspond to layers 1-10 respectively.

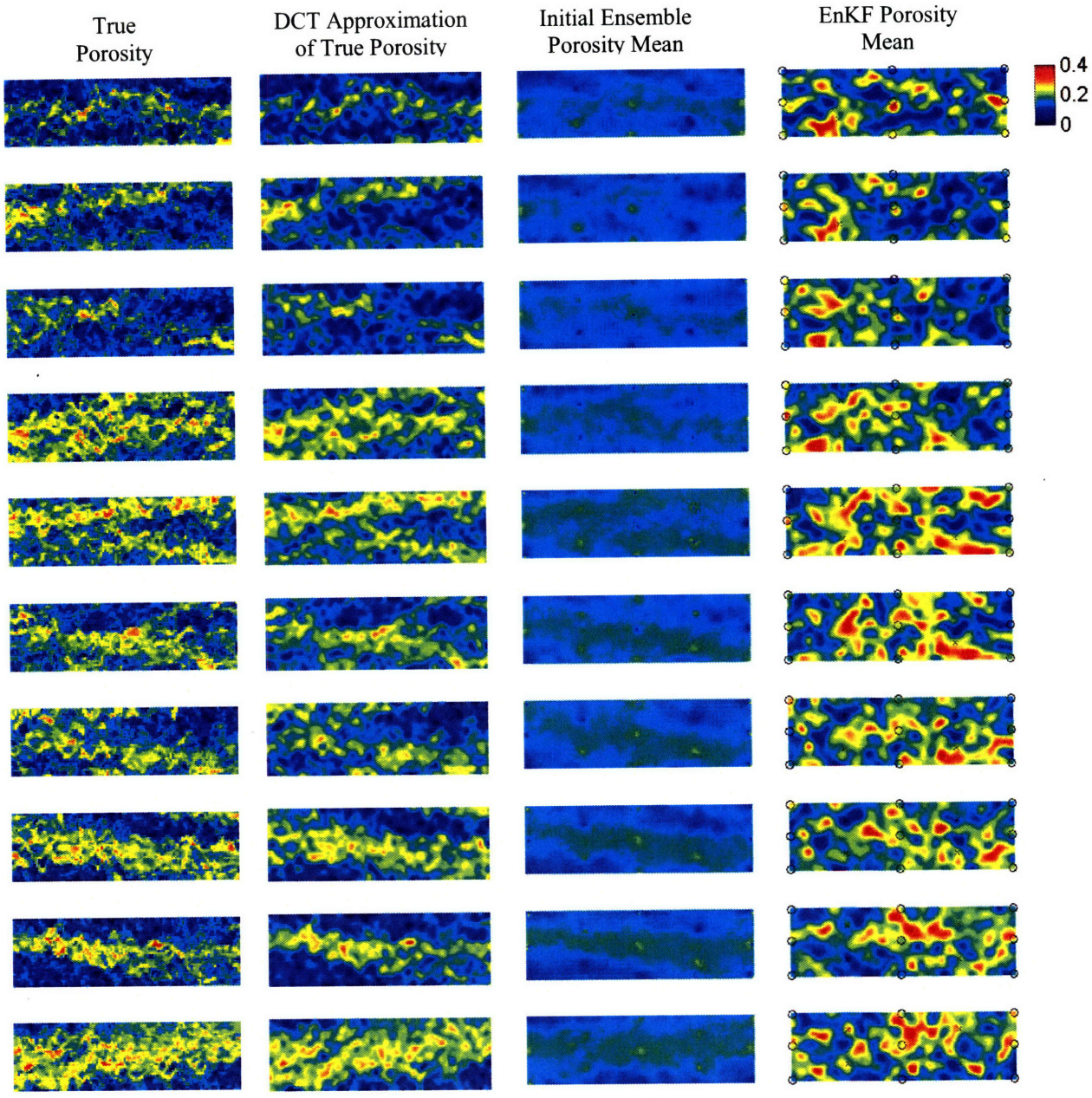


Figure 5.12 Porosity estimation results after 2000 days: the true porosity (first column), initial porosity ensemble mean (second column), and final EnKF porosity estimates mean (third column) are shown. Rows correspond to layers 1-10 respectively.

Figure 5.13 contains the true and EnKF ensemble saturation mean estimate after 2000 days. In general, the saturation profiles agree in terms of global water front movement directions. It is important to note that the DCT bases that are used to span the saturation profiles are trained using the horizontal permeability replicates. This assumption is not accurate at the beginning of the simulation where water distribution generates local and small scale features (while permeability feature are large scale and global). The training of saturation bases and other states/parameters can be done using their respective prior ensemble. Since saturation and pressure are dynamic states, the important DCT bases are expected to be time-dependent. Therefore, a more accurate bases selection should be based on forecast states than the prior permeability field. This alternative bases selection procedure adds some computational overhead to the overall complexity of the algorithm.

The smoothness of the estimated saturation profile in Figure 5.13 is due to the DCT's low frequency description and the effect of averaging over the ensemble replicates. Appearance of the water saturation front in regions where it is not expected could be due to the averaging effect or the vertical water displacement. Inspection of the individual saturation replicates after 2000 days indicated small variations. This implies that the local appearance of water is mainly due to vertical water movements.

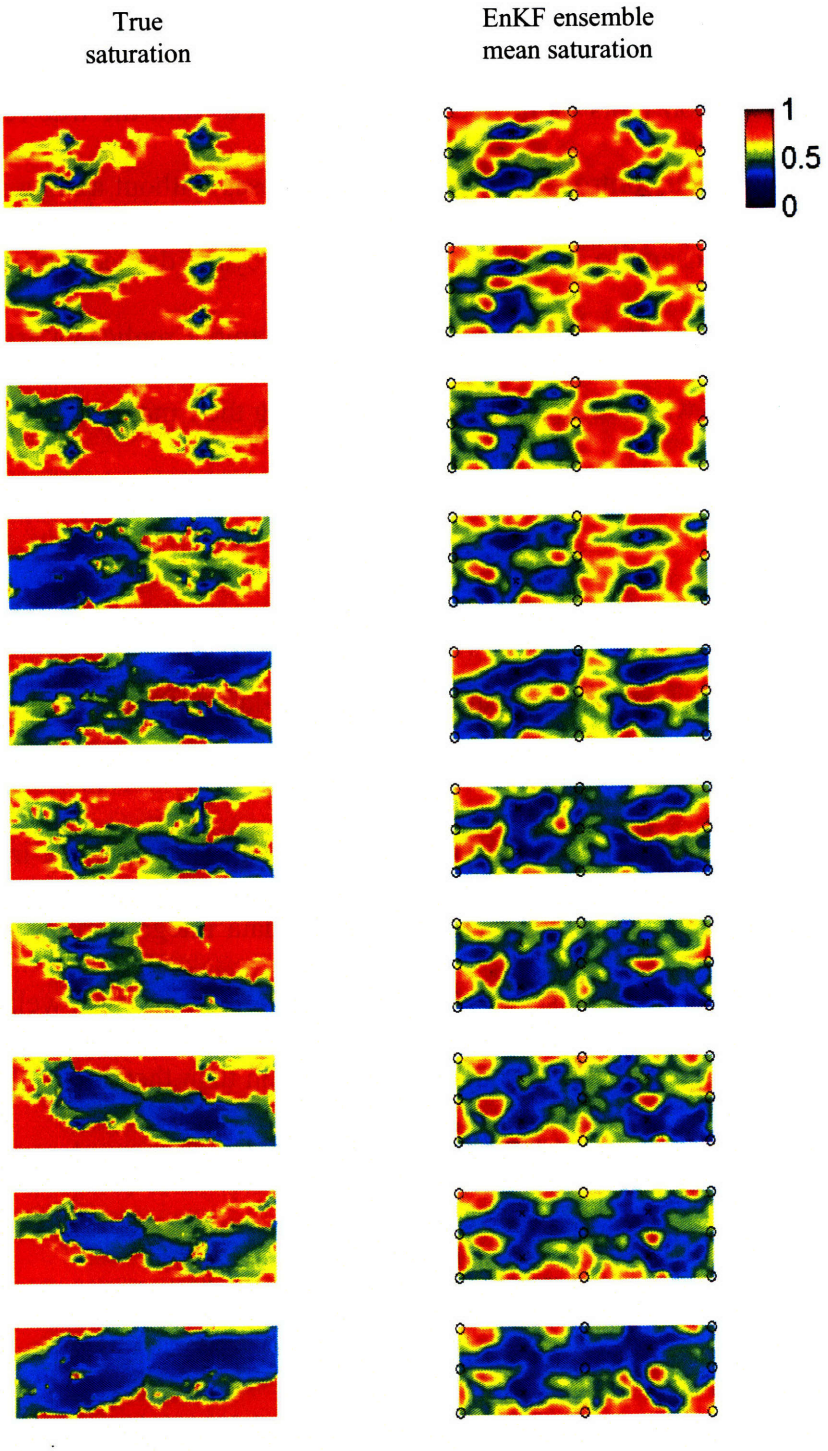
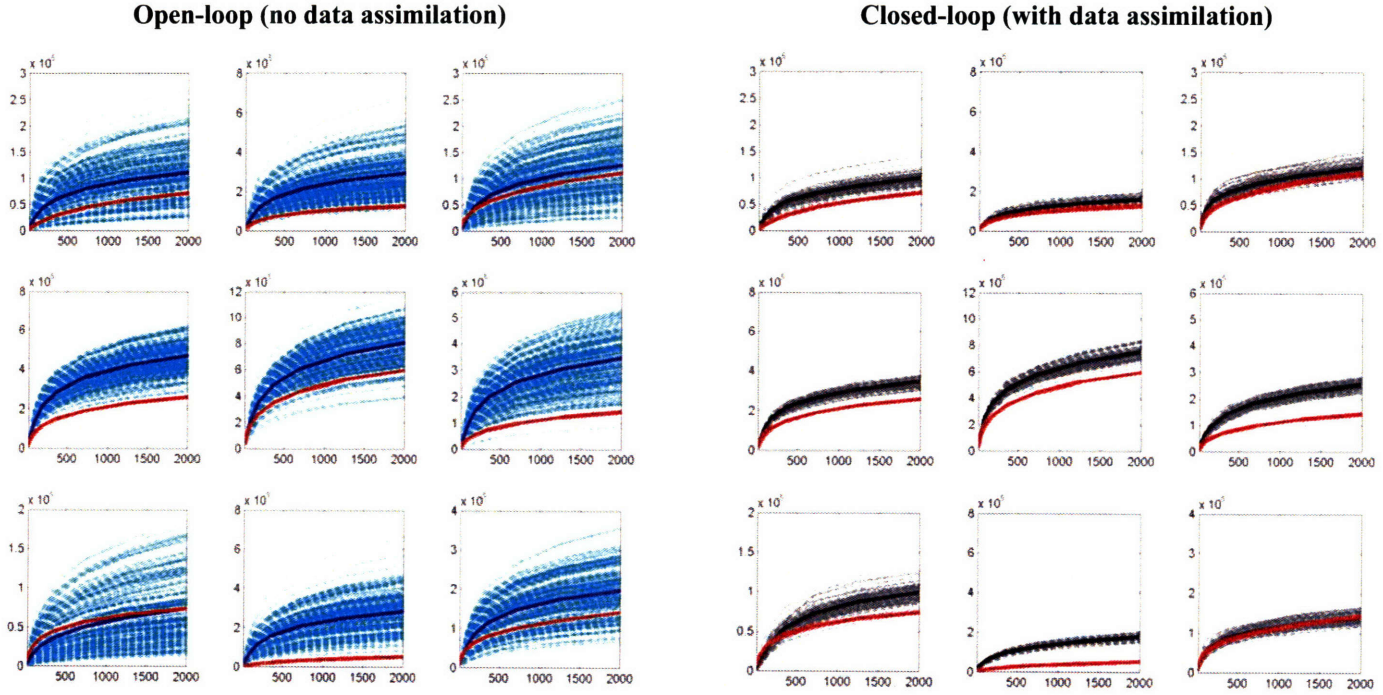


Figure 5.13 Saturation results at the end of the assimilation (after 2000 days): the first column shows the true saturation profile and while the second column shows the EnKF ensemble mean saturation.

Finally, Figure 5.14 shows the open-loop (forecast without data integration), closed-loop (forecast after data integration), and true production curves for selected wells in the field. The true production plots (solid red line) are derived from the true reservoir properties using the same well controls as in the other two cases without adding any perturbations. The open-loop ensemble members are shown in cyan (dashed lines) with their mean shown by the solid blue line. The closed-loop ensemble replicates are shown using the dashed grey lines with the solid black line indicating their mean value.

From Figure 5.14 it is clear that the closed-loop production estimates are overall more accurate and less uncertain compared to the open-loop estimates. In general, the EnKF estimates are expected to have less uncertainty due to the formulation of the filter, however only when the uncertainty reduction is accompanied by the improvement in the estimation accuracy can it be concluded that the data integration has, indeed, improved reservoir description. In this case, it can be seen that in most wells the closed-loop production forecasts are closer to the true production than they are in the open-loop case. Hence, it can be concluded that data integration has improved reservoir's predictive power.

a) Cumulative Oil Production Curves



b) Cumulative Water Production Curves

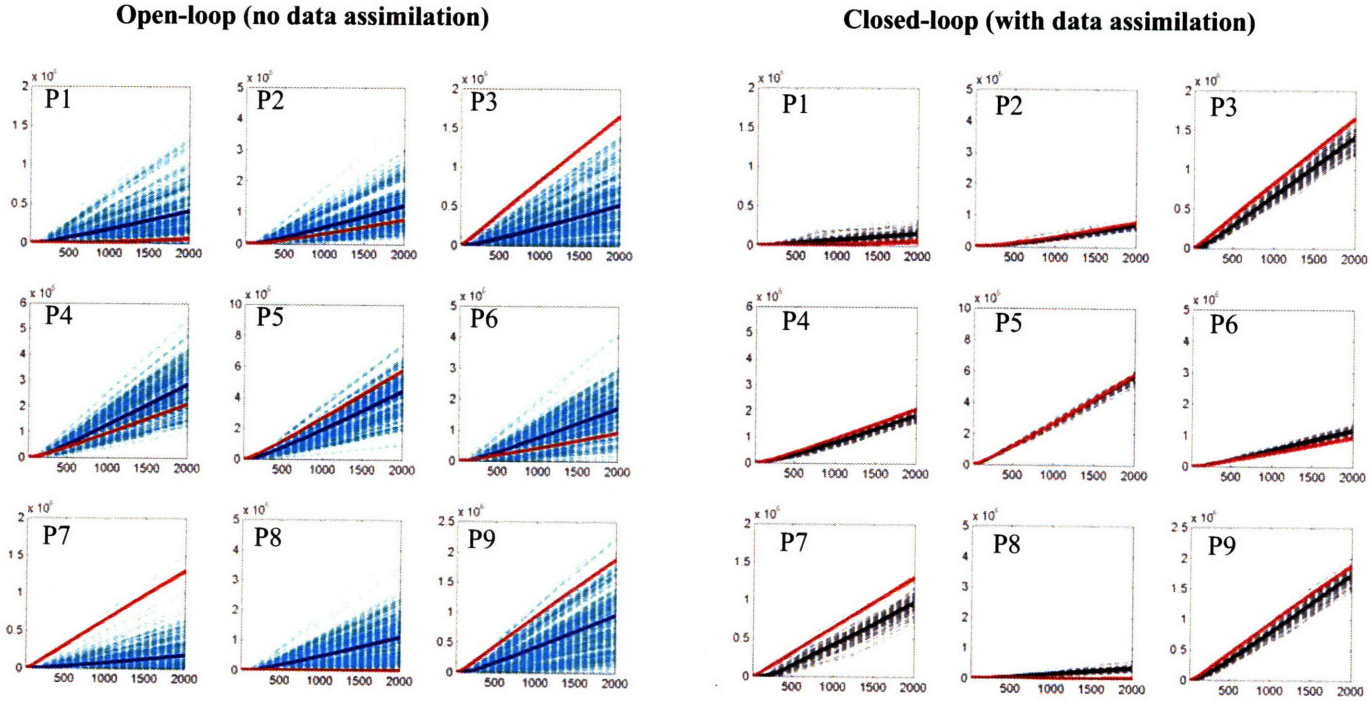


Figure 5.14 Production plots for modified SPE-10 reservoir: cumulative oil (a) and water (b) production observations for individual wells P1-P9: true production curves (red), open-loop forecast and its mean (cyan and blue - top left), and EnKF forecast and its mean (grey and black - top right) are shown.

A closer inspection of the results indicates that the closed-loop results for water production are generally more accurate than those for oil production. This is explained by noting that the closed-loop cumulative production results are not based on the final EnKF estimates (after 2000 days). They are obtained using the improving EnKF estimates in time; therefore, initial estimates are based on less accurate permeability estimates. This can result in inaccurate initial production results. Since it takes some time for water breakthrough to occur water production estimates are less affected by the initially inaccurate reservoir description.

The total amount of oil production forecasts in several wells is greater than the actual (true) oil production. This is mainly due to the high porosity values in the prior. Investigation of the prior replicates suggests that the channel facies proportion (high porosity areas) is generally greater in the prior ensemble replicates than in the true reservoir. Therefore, the initial oil-in-place is overestimated. The EnKF updates improve the porosity (hence the oil-in-place) estimates in almost all wells, however in some cases the improvement is not sufficient to capture the actual oil production. This effect is less seen in water production mainly because the initial water-in-place corresponds to the residual water that can not be extracted and therefore is less sensitive to errors in initial porosity specification.

Overall, the production results suggest that the history matching framework used in this chapter appears to be effective in improving reservoir production forecast.

5.6 Conclusions

In this chapter a three-dimensional virtual reservoir model was used to evaluate the performance of the EnKF-DCT approach for history matching. Construction of the prior facies distribution and reservoir properties was carried out by combing the multipoint geostatistics to generate prior facies distribution and the sequential Gaussian simulation to model reservoir properties variability within each facies. Horizontal and vertical permeabilities as well as porosity were considered as unknown parameters to be estimated from production measurements. Production forecast was used as a metric to quantify the history matching performance. The estimated properties (derived from the proposed history matching framework) resulted in an improved production forecast. This suggests that the proposed history matching framework in this thesis holds promise for application in realistic reservoir settings.

References

5.1. SPE-10 virtual reservoir model website: http://www.spe.org/csp/datasets/set02.htm

5.2. Christie M.A., Blunt M.J., (2001): "Tenth SPE comparative solution project: A comparison of upscaling techniques", *SPE* paper 66599.

5.3. Huang W., Di Donato G., and Blunt M.J., (2004): "Comparison of streamline-based and grid-based dual porosity simulation," *Journal of Petroleum Science and Engineering*, 43, 129-137.

5.4. Abdelhalim M. B., Salama A. E., (2003): "Implementation Of 3D-DCT Based Video Encoder/decoder System", *International Symposium on Signals, Circuits and Systems (SCS 2003)*, 389 – 392.

5.5. Yeo B., Liu B., (1995): "Volume rendering of DCT-based compressed 3D scalar data", *IEEE Trans. Visualization and Computer Graphics* 1 (1) 29{43}.

5.6. Rao K., Yeb P., (1990): *Discrere Cosine transform: algorithm, advantages, applications.* Academic Press Inc.

5.7. Strebelle S., Journel A.: "Reservoir Modeling Using Multiple-Point Statistics," paper SPE-71324 presented at the 2001 SPE Annual Technical Conference and Exhibition, New Orleans, Sept. 30-Oct. 3.

5.8. Avseth, P., Mukerji, T., Joerstad A., Mavko G. and Veggeland, T. 2001 (in press), Seismic reservoir mapping from 3-D AVO in a North Sea turbidite system: Geophysics.

5.9. Remy, N., 2004: S-GeMS, "A Geostatistical Earth Modeling Library and Software,": Ph.D. thesis, Stanford University.

Chapter 6

Conclusions and Future Research

This chapter presents a summary of the conclusions that can be drawn from the previous chapters of this thesis. It also highlights the major contributions of this work and points to some future research directions in reservoir characterization.

6.1 Thesis Conclusions and Contributions

6.1.1 Thesis Conclusions

This thesis covered a number of topics related to state and parameter estimation in reservoir engineering. The ensemble data assimilation method that was studied in this thesis, i.e. the ensemble Kalman filter, is a promising history matching algorithm under reasonably specified prior information. Its ability to identify high permeability channels was confirmed in experiments where synthetic "true" channel configurations were

captured using a channelized training-image prior and production observations. It was also successful in capturing permeability trends in a Gaussian-type permeability field example. Features nearer the well measurements were generally estimated more accurately than features located further away.

In general, it is concluded that the ensemble Kalman filter requires proper specification of the permeability ensemble to yield successful results. The filter's performance was promising in capturing the facies shape and geometry when the prior statistics (up to second order) were correctly specified. While the filter performed reasonably well despite tolerable errors in the prior, it could not correct a completely wrong prior and filter divergence (performance deterioration in time) was observed. Overall, the sample covariances and their corresponding updates gave better results when the underlying ensemble members had similar geometrical shapes and directionality to the true permeability.

When dealing with channelized reservoirs, it is essential to include compatible channel dimensions and geometrical properties. A training image approach with biased channel width produced relatively poor EnKF estimates that degraded over time, in a manner similar to the estimates described by Brouwer et al. [6.1]. When the bias in the channel width was removed the resulting permeability field estimate was remarkably improved. Results from experiments where a mixture of correct and incorrect channel geometries were used indicated that including uncertainty in the prior ensemble does not result in

filter performance degradation, unlike the case where sufficient uncertainty is not included and the prior is biased. Therefore, it is important to have a realistic quantification of the uncertainty in the prior and reflect it in generating initial ensemble members.

While the original Kalman update equation (2.2) is a second order characterization and optimal only in case of jointly Gaussian prior and observation error statistics, in the channelized permeability examples of the previous chapters the EnKF could reproduce the bimodal true permeability field reasonably well. Even when the permeability field is not bimodal, saturation profiles show strong bimodality (with a sharp transition at the vicinity of the moving water front). However, the EnKF performed well in estimating saturation profiles in the examples of this thesis. This indicates the usefulness of the EnKF update form in reservoir history matching application, even though the linear Gaussian assumptions are violated.

Real world reservoirs can have very complex channelized and faulted geometries that cannot generally be reproduced with statistically simple random fields, such as stationary Gaussian fields. Facies generation methods based on realistic training images and multi-point geostatistical methods seem to be better able to generate geologically credible realizations in channelized environments. When these realizations were used with the ensemble Kalman filter to derive sample covariances, the resulting estimates were reasonable even if the underlying distributions were far from Gaussian.

The performance of the EnKF is also dependent upon the number of realizations used in the ensemble. In some cases, an ensemble size of 100 was too small to give reliable results while an ensemble size of 300 appeared to be sufficient. The sample covariances that control the filter updating procedure is expected to improve when the ensemble replicates are realistic and when the ensemble is large enough to provide an adequate characterization of uncertainty. However, the computational cost of forward simulation runs is high and, therefore, it is important to adopt proper sampling strategies to select realistic replicates before applying the EnKF in practice.

Ensemble generation is definitely more challenging in a real application than it was in the examples of this thesis. In more realistic situations it is important that the training image (or images) used for permeability/porosity estimation be derived from field data and represent all the features likely to occur at the site of interest. At locations where the geology is highly uncertain the training image channels should vary in width, tortuosity, connectedness, and complexity, so that this uncertainty is reflected in the ensemble. Proper ensemble design is a critical part of the parameter estimation process. If the filter's ensemble reflects the likely range of true conditions, the resulting estimates can be expected to be more accurate and robust. This is an important topic that deserves further investigation from both research and application perspectives.

The results from examples discussed in Chapters 3, 4, and 5 indicate that the discrete cosine transform is an attractive parameterization method for spatially correlated fields

(such as permeability). While the method is widely used for image compression, it is also well-suited for parameterization of ill-posed estimation (or history matching) problems. The coefficients of an appropriately truncated DCT are the parameters to be estimated in the history matching procedure and the complete block permeability field can be reconstructed from the inverse DCT. The DCT parameterization approach is robust and computationally efficient and has the flexibility of including both quantitative and qualitative prior information through prior bases screening. It can also be used for parameterization in situations where prior information is unavailable.

When combined with the EnKF, the DCT provides an efficient history matching framework, in which the entire state vector can be reduced before the update step. The estimated production forecasts using the DCT parameterization were almost identical to the results obtained with the block-oriented reservoir description. However, the DCT approach provides better facies connectivity and improved history matching results than the alternative block-oriented approach. This implies that the computational effort spent on estimating individual grid-block properties is unnecessary and undesirable if an acceptable approximation of the states and variables can be obtained with smaller number of parameterized coefficients.

The approximation introduced by using DCT representation of permeability seems to provide a smoother estimate of the original permeability field, as might be expected given the fact that smaller scale details are omitted in the truncated DCT. However, the

examples of Chapters 4 and 5 suggest that the estimated saturations are less sensitive to approximations introduced by the truncated DCT, at least in determining the global pattern of oil displacement. This has important implications for optimization and control of oil production operations, where knowledge of the movement of saturation profile is used to adjust reservoir control variables. Overall, it looks like a DCT version of the ensemble Kalman filter provides an accurate, efficient, robust, and flexible option for reservoir history matching.

6.1.2 Thesis Contributions

The original contributions of this thesis are summarized in this section. While several topics were discussed in the body of this thesis, the major contributions can be listed as follows:

1) *Assessment of the suitability of the ensemble Kalman filter for oil reservoir characterization using dynamic production observations.*

Several waterflooding examples were used to illustrate the applicability of the ensemble Kalman filter for history matching. The examples included bimodal and Gassuian type parameter distributions. The filter performance in both cases was promising, despite the evident violation of the assumptions used in the update equation. This indicates that, for

the history matching examples used in this study, second order characterization that combines flow dynamics and parameter/state distributions with production observations is promising.

2) *Addressing the previously observed divergence and performance issues of the ensemble Kalman filter for non-linear, non-Gaussian history matching problems.*

The divergence of the filter reported in earlier studies was shown to be mainly due to incompatible prior specification. While it is difficult to prove filter convergence for nonlinear models, examples were used to show that under incorrect prior specifications, divergence is likely to occur. In contrast, when prior uncertainty was properly accounted for, the divergence behavior did not occur. It was concluded that the filter design and inputs have to be realistic and handled appropriately to achieve robust and consistent results.

3) *Introduction of an efficient, flexible, and robust parameterization approach, using the discrete cosine transform, for solving history matching problems with and without including prior knowledge.*

The discrete cosine transform was shown to have attractive properties that can be taken advantage of in reservoir parameterization context. The computational efficiency (inherited from the basis seprarability property), flexibility to include prior information (due to its bases data independence), and robustness when prior information is highly uncertain are among the prominent features of the proposed parameterization approach. Application of the DCT for parameterization was introduced, for the first time, in this thesis.

4) *Development of an efficient and reduced-order history matching framework using the ensemble Kalman filter and the discrete cosine transform to represent states and parameters during the update step.*

Previous parameterization techniques were only used to represent parameters. In principle, state variables with redundant description can similarly be reduced in order to provide computational savings and better geological continuity without sacrificing accuracy. This was shown with several examples in Chapters 4 and 5. The combination of the ensemble Kalman filter with the discrete cosine transform parameterization provides an efficient history matching framework for large problems.

6.2 Future Research

The conclusions of this thesis raise several important research questions. Some of these questions lead to future research directions that are presented in this section. In general, the sequential ensemble filtering consists of three elements: constructing a prior ensemble, providing observations with realistic uncertainty, and merging the prior ensemble with observations using a model updating scheme. A realistic representation of each of these components is crucial for success of the overall algorithm. This realization leads to several research directions to address different aspects of ensemble data assimilation. Some of these subjects are briefly discussed below.

It is important to appreciate the importance of realistic prior generation in implementing the model inversion. Realistic generation of prior ensemble and its uncertainty in reservoir application is critical as the limited production observations lead to significant prior-based spatial interpolations. For update schemes that depend heavily on prior models (this is often the case due to required regularization to improve ill-posedness), the quality of the results is a direct function of the accuracy of the prior model. Therefore, prior ensemble generation is an important aspect of data assimilation in reservoir engineering that should be handled carefully.

Realistic assessment of the prior uncertainty is also an important topic. It is natural to ask *"how uncertain is the prior relative to observations?"*. In the Bayesian estimation

framework, prior and observations are weighted based on their uncertainties. In practice, these uncertainties must be specified based on our knowledge about the reservoir. However, placing too much importance on the prior without adequate justification can handicap the estimation algorithm and result in estimates that are forced to look like the prior when the observations suggest otherwise. In reservoir engineering applications, there is typically significant uncertainty inherent in the prior information. Further work is needed on realistic quantification and incorporation of uncertainty in estimation problem.

Another important issue regarding the prior ensemble is characterizations of the key features (e.g. channels and faults) that are believed to be present in the physical reservoir. Characterization of these important features and their incorporation in the prior ensemble (with their respective uncertainty) can result in a more realistic sampling strategy. This can improve the estimation results and lead to computational savings associated with avoiding forward model integration of unrealistic ensemble replicates. While characterization of key features in the prior is critical, it should be complemented with appropriate data integration algorithms that are designed to preserve these key structures. For instance, when channels are believed to be present in the field, an ensemble featuring a channelized environment is more appropriate than a Gaussian representation of the field. Further, it is also important to adopt a model updating algorithm that can preserve channel connectivity. In this thesis, DCT parameterization was used to enforce this effect.

While the ensemble Kalman filter is a promising history matching approach, the update equation of the filter requires sample prior covariances that are expensive to compute for large problems because it requires forward simulations of multiple prior reservoir models. In history matching where significant uncertainties are dominant exact and time-consuming computation of state and production forecasts using grid-based numerical simulation is hard to justify. This can present serious setbacks in practice and should be addressed properly. An important research direction is developing approximate methods to derive the same information in the sample covariance without having to run expensive forward runs. Alternatively, one could alter the form of the update equation to eliminate the need for expensive forward simulations.

Technological advances in reservoir monitoring can provide an array of observed quantities at different resolution and coverage. A specific example of this is the time-lapsed seismic, which is hoped to provide information of fluid displacement in time throughout the reservoir. This can significantly improve our knowledge about regions of reservoirs that are away from well locations and result in more accurate and confident forecasts about reservoirs behavior. Incorporation of seismic inversion in ensemble based approaches is more convenient than iterative algorithms. However, there are major issues regarding use of time-lapsed seismic that require research. Seismic data have extensive coverage and lower resolution that makes them highly uncertain. Additionally, modeling the relationship between measured seismic attributes and reservoir states and properties is not straightforward and require pterophysical modeling. Therefore, further research is

required to assess the feasibility of using time-lapsed seismic data in history matching with ensemble filtering algorithms.

As discussed in chapter 1, several sources of uncertainty are present in reservoir description. Ensemble data assimilation schemes can accommodate all uncertainty sources, yet it would be overwhelming to estimate all uncertain parameters from a limited set of measurements. This thesis concluded that permeability and porosity can be estimated from reasonable prior information and production observations. Further studies are needed to evaluate the performance of ensemble data assimilation in characterizing other uncertain parameters such as relative permeability, initial oil-water distribution and contact level, well skin factors, and other uncertain quantities.

Following the discussion on proper prior generation, it is important to use an update procedure that honors and preserve the existing features in the field. The Kalman update form that was used in this study is a second order characterization. For channelized examples second order statistics may not be adequate to describe channel connectivity. In the examples of this thesis, the initial Kalman updates broke down the channels present in the prior. It was only after several updates (when dynamic information implied presence of channel) that the "true" channels were formed. More effective updates that can preserve the channel structures from the beginning may be found to update channelized permeabilities.

Another relevant topic for research is to understand and improve the robustness of the ensemble Kalman filter. Here, robustness refers to the sensitivity of the filter to prior information. In history matching application, only a sparse set of observations are used to update a large number of unknowns. The EnkF update equation relies entirely on prior sample covariances to estimate the unobserved states and parameters (using point correlations). The original form of the EnKF update equation is derived based on linear-Gaussian assumptions that are clearly violated in the history matching application. Therefore, the updates are no longer optimal, leading to the question *"why should the form of the update equation remain unchanged for non-Gaussian problems?"*, or alternatively *"How can the EnKF update equation be altered to satisfy other desirable filter attributes, such as robustness?"*. These questions may be answered by analyzing the properties of the update form (Kalman gain), its dependence on the physical system of interest, and error analyses after the update step.

In parameterization of permeability fields in Chapter 3, the important bases were selected only based on the prior information, while the observations were used to estimate their corresponding transform coefficients. In this case, if the prior parameter distribution is incorrect, the selected bases can not represent the main features of the true field, regardless of the performance of the estimation algorithm. An alternative approach is to include prior information as well as observations in selecting important bases. This leads

to adaptive bases selection algorithms that can help mitigate the effects of incorrect prior specification and improve estimation results.

In this thesis, the discrete cosine transform was used to transform permeability fields that were already generated using geostatistical simulations. It is reasonable to ask whether the DCT can be used for generating permeability realizations by for instance using assumptions about the distribution of the DCT coefficients. To pose the question differently, *"Can we use the spectral content of a training image to generate realizations that are similar to the existing features in the training image?"*. This is an interesting research direction that can lead to computational efficiency in consistent facies generation from training images.

Finally, while improving the accuracy and computational efficiency of the EnKF update were studied in this thesis, the forward simulation runs were the most expensive part of the EnKF history matching process. Fast streamline simulation technique [6.2, 6.3] is one way of improving the computational aspects of the forward runs. Wavelet transforms have been previously used for reservoir parameterization [6.4], but they can also be useful for reducing forward model computations because of their multiscale property. In applications with the EnKF several forward model runs have to be performed to provide prior sample covariances. Given the approximate nature of sample covariances and the existing uncertainty in the prior, it may be possible to use coarse scale (faster) simulation runs to derive these covariances. Wavelets also provide adaptive local grid refining

capabilities, which may be used to perform faster computations while preserving important fine scale variabilities.

In summary, the results presented in this research suggest that the ensemble Kalman filter is a promising history matching technique under reasonably specified prior information. While it is straightforward to implement the method, its computational requirements may present a challenge when applying it in realistic settings. There are several issues to be studied before successful application of the current form of this method is achieved in practice. Generation of a realistic ensemble of reservoir models may not be trivial in practice; several sources of uncertainty may be present and have to be incorporated in a consistent manner. Studies of the method in the literature, including this thesis, have been limited to fewer sources of uncertainty that can exist in reality. Therefore, the successful results and conclusions reported about the performance of the filter in here (and other studies) can not be generalized to realistic situations where all possible sources of uncertainty should be taken into account. Furthermore, computational cost of model forward runs to construct the forecast states before the analysis step of the EnKF can be very involved. While this was not the focus of this thesis, approximate solution of the dynamic model integrations can be expected to mitigate the computational time and provide a reasonable alternative. The efficiency of the DCT bases computation and independence of its bases from prior information provide great advantages over other parameterization alternatives. The DCT is a general parameterization approach that is

recommended for use with any history matching algorithm with or without including prior information.

Finally, the ensemble Kalman filter application for history matching is currently at research stage. Its performance and implementation cost are being studied by several researchers in the field. While promising preliminary results are being reported in the literature, it seems too early to reach a final conclusion regarding its general applicability to large scale realistic reservoirs.

References

6.1. Brouwer D. R., Nævdal G., Jansen J.D., Vefring E.H., van Kruijsdijk C. P. J. W.: "Improved reservoir management through optimal control and continuous model updating," paper SPE 90149 presented at the SPE Annual Technical Conference and Exhibition, Houston, TX (2004).

6.2. Datta-Gupta A., King M.J. (2007), *Streamline Simulation: Theory and Practice*, Society of Petroleum Engineers, Textbook Series.

6.3. Devegowda D., Arroyo E., Datta-Gupta A., (2007): "Efficient and Robust Reservoir Model Updating Using Ensemble Kalman Filter With Sensitivity Based Covariance Localization," SPE 106144 presented at the *SPE Reservoir Simulation Symposium*, Houston, TX.

6.4. Sahni, I., and Horne R.N.: "Multiresolution Wavelet Analysis for Improved Reservoir Description," *SPEREE* (February 2005), 8(1), 53-69.