

**STREAMLINE ASSISTED ENSEMBLE KALMAN FILTER – FORMULATION
AND FIELD APPLICATION**

A Dissertation

by

DEEPAK DEVEGOWDA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2009

Major Subject: Petroleum Engineering

**STREAMLINE ASSISTED ENSEMBLE KALMAN FILTER – FORMULATION
AND FIELD APPLICATION**

A Dissertation

by

DEEPAK DEVEGOWDA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Akhil Datta-Gupta
Committee Members,	Steve Holditch
	Ahmed Ghassemi
	Bani K. Mallick
Head of Department,	Steve Holditch

August 2009

Major Subject: Petroleum Engineering

ABSTRACT

Streamline Assisted Ensemble Kalman Filter – Formulation and Field Application.

(August 2009)

Deepak Devegowda, B.Tech., Indian Institute of Technology, Madras;

M.S., Texas A&M University

Chair of Advisory Committee: Dr. Akhil Datta-Gupta

The goal of any data assimilation or history matching algorithm is to enable better reservoir management decisions through the construction of reliable reservoir performance models and the assessment of the underlying uncertainties. A considerable body of research work and enhanced computational capabilities have led to an increased application of robust and efficient history matching algorithms to condition reservoir models to dynamic data. Moreover, there has been a shift towards generating multiple plausible reservoir models in recognition of the significance of the associated uncertainties. This provides for uncertainty analysis in reservoir performance forecasts, enabling better management decisions for reservoir development. Additionally, the increased deployment of permanent well sensors and downhole monitors has led to an increasing interest in maintaining ‘live’ models that are current and consistent with historical observations.

One such data assimilation approach that has gained popularity in the recent past is the Ensemble Kalman Filter (EnKF) (Evensen 2003). It is a Monte Carlo approach to generate a suite of plausible subsurface models conditioned to previously obtained measurements. One advantage of the EnKF is its ability to integrate different types of data at different scales thereby allowing for a framework where all available dynamic data is simultaneously or sequentially utilized to improve estimates of the reservoir model

parameters. Of particular interest is the use of partitioning tracer data to infer the location and distribution of target un-swept oil. Due to the difficulty in differentiating the relative effects of spatial variations in fractional flow and fluid saturations and partitioning coefficients on the tracer response, interpretation of partitioning tracer responses is particularly challenging in the presence of mobile oil saturations.

The purpose of this research is to improve the performance of the EnKF in parameter estimation for reservoir characterization studies without the use of a large ensemble size so as to keep the algorithm efficient and computationally inexpensive for large, field-scale models. To achieve this, we propose the use of streamline-derived information to mitigate problems associated with the use of the EnKF with small sample sizes and non-linear dynamics in non-Gaussian settings. Following this, we present the application of the EnKF for interpretation of partitioning tracer tests specifically to obtain improved estimates of the spatial distribution of target oil.

DEDICATION

To my lovely wife, Roopa, and my beloved parents and brother. And to all the people along the way who made this journey fun, meaningful and a source of future nostalgia. And my pride and joy, my bicycle, giving me miles of fun away from all the worries.

ACKNOWLEDGEMENTS

I had the enviable opportunity to learn a lot and interact with many people during my PhD and I would like to acknowledge each one of them. First and foremost, I would like to thank my academic advisor, Dr. Akhil Datta-Gupta for his encouragement, academic guidance and support throughout my study. I will be forever indebted to him for everything that he has done for me.

I would also like to thank the members of my committee, Dr. Steve Holditch, Dr. Ahmed Ghassemi, and Dr. Bani K. Mallick for serving as members on my committee and their help in shaping the work in this dissertation.

Special thanks to my colleagues in the research group, Eduardo Jimenez and Adedayo Oyerinde with whom I started this program, for the constructive discussions and knowledge sharing over the years. Also, to past member of the group, Dr. Ahmed Daoud, Dr. Hao Cheng, and present members, Xianlin, Ahmed, Kim, Ajitabh, Pranay, Sarwesh, Alvaro, Shingo, Matt, Jiang, Jichao, thanks for your friendship and making my graduate years memorable.

NOMENCLATURE

A	Matrix specifying spatial derivatives
β	Scalar weighting on regularization terms
$B_{o,w,g}$	Phase formation volume factor for oil, water, and gas
c	Divergence of flux
C_D	Data covariance matrix
C_M	Model covariance matrix
$C_{M,d}$	Cross-covariance matrix between data and model variables
D, d	Data vector
e_k	Error in the model state at time k
ε	Noise in the data
f	Fractional flow
$g(\mathbf{m})$	Simulator response to vector of reservoir parameter \mathbf{m}
K	Absolute permeability
k_r	Relative permeability
\mathbf{m}	Reservoir model parameter vector
\mathbf{m}^s	Vector of static model variables
\mathbf{m}^d	Vector of dynamic model variables
N_d	Number of observation data
N_e	Number of ensemble members
N_m	Number of model variables
φ	Porosity
P	Pressure
ρ	Density
ρ	Covariance localizing function
R	Precision of the coarse-scale data
R_s	Solution gas-oil ratio
$S_{o,w,g}$	Phase saturation for oil, water, gas

S_{ij}	Stacked sensitivity of response i to model parameter j
t	Time
τ	Time of flight
θ	Covariance inflation factor
\bar{u}	Darcy velocity
y	Model state vector
μ	Viscosity
λ	Mobility
σ	Standard deviation

TABLE OF CONTENTS

	Page
ABSTRACT.....	i
DEDICATION.....	v
ACKNOWLEDGEMENTS.....	vi
NOMENCLATURE.....	vii
TABLE OF CONTENTS.....	ix
LIST OF FIGURES.....	xi
1. INTRODUCTION.....	1
1.1 The Ensemble Kalman Filter (EnKF).....	2
1.1.1 Components of the EnKF.....	2
1.2 Minimum Variance Estimator - MVE.....	5
1.3 EnKF Workflow.....	9
1.4 Application of the EnKF: Synthetic Example.....	9
1.5 Difficulties in Data Assimilation Using the EnKF.....	11
1.5.1 Key Aspects of Reliable Data Assimilation.....	11
1.5.2 Assessment of the EnKF.....	12
1.6 Conclusions and Research Objectives.....	17
2. DATA ASSIMILATION: IMPROVED VARIANTS OF THE EnKF.....	19
2.1 Distance Dependant Covariance Localization.....	19
2.2 Covariance Inflation.....	24
2.3 Smoothness-Constrained 2-Stage EnKF.....	26
2.4 Section Summary.....	30
3. STREAMLINE BASED COVARIANCE LOCALIZATION.....	32
3.1 Streamline Trajectory Based Covariance Localization.....	32
3.2 Streamline-Sensitivity Based Covariance Localization.....	37
3.2.1 Motivating Example.....	38
3.2.2 Sensitivity Computation.....	40
3.2.3 Construction of the Localizing Function.....	43

	Page
3.3 Application: Synthetic Example	46
3.4 Field Application: The Goldsmith Field Case	51
3.5 Section Summary	55
4. A HIERARCHICAL MULTI-SCALE EnKF.....	56
4.1 Introduction.....	56
4.2 Methodology	57
4.3 Composite Tracers and Partitioning Interwell Tracer Tests (PITT)	60
4.4 Application: Synthetic 2-D Example	62
4.5 Application: 3-D Example	69
4.6 Section Summary	82
5. CONCLUSIONS AND RECOMMENDATIONS	83
5.1 Conclusions.....	83
5.2 Recommendations.....	85
REFERENCES	87
VITA.....	91

LIST OF FIGURES

		Page
Figure 1.1	Reference log-permeability field and location of the 4 wells.	10
Figure 1.2	(a) is the chart of initial water-cut at 3 wells for the synthetic example for 50 members and (b) shows the water-cut from the history matched ensemble. The reference water cut is the profile in red.	10
Figure 1.3	Comparison of the updated permeability fields on the right with the initial fields on the left for four realizations of the ensemble of 25 members.	13
Figure 1.4	(a) is the initial permeability histogram on the left with the updated histogram on the right in (b) for an ensemble size of 25.	14
Figure 1.5	Comparison of the updated permeability fields on the right with the initial fields on the left for four realizations of the ensemble of 50 members.	14
Figure 1.6	(a) is the initial permeability histogram on the left with the updated histogram on the right in (b) for an ensemble size of 50.	15
Figure 1.7	Comparison of the updated permeability fields on the right with the initial fields on the left for four realizations of the ensemble of 100 members.	15
Figure 1.8	(a) is the initial permeability histogram on the left with the updated histogram on the right in (b) for an ensemble size of 100.	16
Figure 1.9	Comparison of the total variance in the permeability as a function of time during the assimilation process for 3 different ensemble sizes.	16
Figure 2.1	Cross-covariance between water-cut at one well (black dot) and the permeability for two different ensemble sizes.	20
Figure 2.2	Distance based localizing function centered at the well shown in a black dot (Gaspari and Cohn 1996).	21
Figure 2.3	(a) is the initial water-cut at each of the 3 wells and (b) is the matched water-cut. The results do not appear to be satisfactory.	22

Figure 2.4	(a) is the initial water-cut at each of the 3 wells and (b) is the matched water-cut. The matches with a cut-off radius of 400 feet are considerably better.	23
Figure 2.5	Plot of total variance versus time during assimilation comparing the covariance-inflated EnKF versus the standard EnKF for an ensemble size of 50.	24
Figure 2.6	The initial log-permeability histogram on the left compared to that from an updated realization clearly shows the transformation from bimodal to Gaussian.	25
Figure 2.7	Covariance inflation may not be suitable for a small ensemble size (~ 30 model realizations).	26
Figure 2.8	(a) is the water-cut prediction from the ensemble of 50 realizations and (b) shows the matches to the data from the updated models. The final model predictions are not within reasonable bounds.	29
Figure 2.9	Updated log-permeability histogram in (b) in comparison with the initial histogram in (a).	30
Figure 3.1	The area of influence for each observation on the right selected by the streamline pattern on the left.	34
Figure 3.2	Four arbitrarily selected realizations on the left are updated using the streamline trajectory based covariance localization and the results are displayed on the right.	35
Figure 3.3	The initial log-permeability histogram on the left compared with the updated log-permeability histogram on the right for one realization.	36
Figure 3.4	Comparison of the total variance as a function of time during the assimilation process.	36
Figure 3.5	This plot shows two realizations of a completely random permeability field and the corresponding streamline patterns for two wells.	39
Figure 3.6	A comparison of the cross-covariance estimation between water-cut at one well (black dot) and the permeability for various ensemble sizes.	40
Figure 3.7	The streamline pattern and time-of-flight contours for a inverted 5-spot pattern in a synthetic permeability field.	44

Figure 3.8	The streamline trajectory and the normalized streamline-derived sensitivities for 2 selected wells for the field shown in Figure 3.7.....	45
Figure 3.9	The normalized sensitivities computed for each well from an ensemble of 50 realizations at $t = 1000$ days. The significant areas impacting each well's water-cut data are highlighted.....	46
Figure 3.10	(a) is the initial spread in the predicted water-cut from the ensemble. (b) is the match to water-cut data achieved using EnKF with streamline-derived sensitivities for covariance localization.....	47
Figure 3.11	The updated permeability fields for 4 realizations on the right compared to the initial members on the left. Notice that changes to the initial models are minimal.....	48
Figure 3.12	The total variance amongst the members as a function of time during the assimilation process. Flow relevant conditioning of the cross-covariance clearly provides substantial benefits by not artificially reducing the variability.....	49
Figure 3.13	(b) is the histogram of log-permeability for one updated realization compared to (a) which is the initial histogram. We maintain the multi-modal nature of the histograms without overshooting.....	49
Figure 3.14	Eigen-spectrum of the covariance matrix using sensitivity based localization (in red) and without any form of localization (in blue).....	50
Figure 3.15	The location of the Goldsmith pilot area in the San Andreas Unit in West Texas. The location of the producers (in yellow) and the injectors (in blue) is on the right.	51
Figure 3.16	(a) is the initial permeability field showing the multi-modal permeability histogram. (b) illustrates the transformation to a normal histogram with the standard EnKF, while (c) shows the results from the streamline-sensitivity based covariance localization.....	52
Figure 3.17	Matches to production history at 3 selected wells in the bottom row. The initial predictions are plotted in the top row.....	54
Figure 3.18	On the left is the cross-covariance between water-cut data at the well (vertical line) and permeabilities in the field. Using streamline-derived sensitivities, we condition the cross-covariance matrix and on the right is the plot of the localized cross-covariance.....	54
Figure 4.1	The multi-scale EnKF procedure at any given assimilation step.....	60

Figure 4.2	Typical composite tracer responses showing the delay of the partitioning tracer (Oyerinde 2004).	62
Figure 4.3	The tracer profiles at each of the 4 producing wells showing the chromatographic delay of the partitioning tracer. The reference field is on the right.	63
Figure 4.4	The top row shows the initial water-cut predictions and the bottom row shows the matches to the data from the updated realizations.	64
Figure 4.5	Plot of composite tracer responses from the initial ensemble for two selected wells in the top row and the corresponding matches in the bottom row.	64
Figure 4.6	Comparison of the saturation contours of the reference field on the left with the mean of the ensemble on the right at two different times.	65
Figure 4.7	The coarse-scale constraint on the right serves as an observation to improve the estimates of the fluid saturation distribution in the ensemble prior on the left.	66
Figure 4.8	The top row shows the initial water-cut predictions and the bottom row shows the matches to the data from the updated realizations.	67
Figure 4.9	The top row shows the conservative and partitioning tracer responses from the initial ensemble for two selected wells and the bottom row shows the corresponding matches.	67
Figure 4.10	Comparison of the target oil saturations from the standard EnKF (center) with those obtained from the 2 stage EnKF (right) at two different times. The profile from the 'true' model is provided as a reference (left).	68
Figure 4.11	The 2-stage EnKF (right) outperforms the standard EnKF implementation (center) when compared to the reference saturation.	69
Figure 4.12	The reference log-permeability field showing the location of the producers and injection wells.	70
Figure 4.13a	Matches to the water-cut profile in the bottom row. The initial ensemble predictions are in the top row. The only wells shown are P1, P2 and P3.	71

Figure 4.13b	Matches to the conservative tracer data in the bottom row. The initial predictions are in the top row. The only wells shown are P1, P2 and P3.	72
Figure 4.13c	Matches to the partitioning tracer profile in the bottom row. The initial predictions are in the top row. The only wells shown are P1, P2 and P3.	72
Figure 4.14a	Saturation profile in Layer 1 at 1350 days.....	73
Figure 4.14b	Saturation profile in Layer 5 at 1350 days.....	73
Figure 4.14c	Saturation profile in Layer 1 at 2250 days.....	74
Figure 4.14d	Saturation profile in Layer 5 at 2250 days.....	74
Figure 4.14e	Saturation profile in Layer 1 at 4000 days.....	75
Figure 4.14f	Saturation profile in Layer 5 at 4000 days.....	75
Figure 4.15	A comparison of the prior mean saturation on the left with the imposed coarse-scale constraint on the right.	76
Figure 4.16a	Matches to water-cut data in the bottom row for wells P1, P2 and P3 using the 2-stage EnKF. The initial predictions are in the top row.	77
Figure 4.16b	Matches to conservative tracer data in the bottom row for wells P1, P2 and P3 using the 2-stage EnKF. The initial predictions are in the top row.	77
Figure 4.16c	Matches to partitioning tracer data in the bottom row for wells P1, P2 and P3 using the 2-stage EnKF. The initial predictions are in the top row.	78
Figure 4.17a	Contour plot of the bypassed oil saturation at 1350 days in Layer 1.....	79
Figure 4.17b	Contour plot of the bypassed oil saturation at 1350 days in Layer 5.....	79
Figure 4.17c	Contour plot of the bypassed oil saturation at 2250 days in Layer 1.....	80
Figure 4.17d	Contour plot of the bypassed oil saturation at 2250 days in Layer 5.....	80
Figure 4.17e	Contour plot of the bypassed oil saturation at 4000 days in Layer 1.....	81
Figure 4.17f	Contour plot of the bypassed oil saturation at 4000 days in Layer 5.....	81

1. INTRODUCTION

The economic impact of inaccurate predictions of future petroleum reservoir performance is substantial, therefore making proper characterization of the reservoir and uncertainty analyses in production forecasts crucial aspects in any reservoir development strategy. This goal is achieved through the use of history matching algorithms which reconcile reservoir models to measurements such as production data or time lapse seismic information. The workflow depends on constructing an appropriate initial reservoir simulation model and adjusting the values of the uncertain model variables so that the model performance is in reasonable agreement with actual measurements. These approaches can be broadly classified into two distinct categories: gradient based algorithms and stochastic approaches. Regardless of the approach, all previously recorded data is simultaneously used to update the reservoir model. The first class of algorithms generally utilizes parameter sensitivities or gradients of an appropriately constructed objective function to arrive at a solution to the inverse problem. To generate multiple history matched model realizations, these techniques involve the repeated application of the procedure to each realization of the reservoir, a procedure which can be computationally demanding. Furthermore, these approaches rely on the development of code that may not be capable of handling diverse types of dynamic data. On the other hand, stochastic algorithms like the Markov Chain Monte Carlo (MCMC) approach, simulated annealing and genetic algorithms rely on statistical approaches to arrive at solutions to the inverse problem. However, these approaches are slow to converge and require excessive turn-around times for model calibration. The associated difficulties with both categories of history matching algorithms are further compounded when it is

This dissertation follows the style of *SPE Journal*.

necessary to assimilate data frequently. Because of the increased deployment of permanent sensors or downhole monitors, it is important to maintain ‘live models’ that are progressively updated as soon as data is obtained.

Data assimilation for reservoir characterization studies therefore require computationally efficient algorithms that can provide uncertainty assessment by generating multiple plausible reservoir models, have the capability to integrate data when it becomes available and be flexible enough to assimilate diverse data types. The ease of implementation without the use of simulator-specific code is another key requirement. A recent and promising approach that combines these capabilities is the Ensemble Kalman Filter (EnKF) technique. This section presents an introduction to the EnKF and highlights the motivation and objectives of the research in this dissertation.

1.1 The Ensemble Kalman Filter (EnKF)

There are several variations to the original EnKF first proposed by Evensen (1994) and later by Houtekamer and Mitchell (1998). These include the Ensemble Square Root Filter (EnSRF) (Whitaker and Hamill 2002), the Ensemble Adjustment Filter (EAF) (Anderson 2001) and the Ensemble Transform Kalman Filter (ETKF) (Bishop et al. 2001). In this study, we use the initial formulation proposed by Evensen (1994). The following section provides the algorithmic details of the EnKF as it relates to data assimilation for the purpose of reservoir characterization.

1.1.1 Components of the EnKF

Broadly speaking, the EnKF comprises 3 distinct components:

The Ensemble of Model Realizations. The EnKF works with an ensemble of reservoir models. Geologic information is augmented with static well data to populate the models with rock properties inferred from measurements like well logs and core analysis. Additional information to improve our initial estimates of the models may include structural settings derived from 3-D seismic surveys. These types of data are often classified as static data to reflect the unchanging temporal nature of this information. To

incorporate all of this information, geostatistical interpolation techniques like Sequential Gaussian Simulation (Deutsche and Journel 1992) are often employed to generate multiple realizations of the reservoir that are consistent with the prior static data.

Such a set of initial model realizations may be defined as random samples from a multidimensional prior probability distribution function (pdf) of the unknown variables. Typically, the uncertain variables are drawn from a multi-Gaussian prior distribution with a specified mean and a covariance that reflects our uncertainty in the estimates of the model variables. The process of data assimilation attempts to improve the initial estimates of these unknown variables which could include the spatial distribution of porosity, permeability and phase saturations and/or the locations of fluid contacts and phase fronts. The EnKF progressively updates these model realizations so that at any time models are kept current with all previously obtained static information and production data. The updated models constitute samples from a posterior pdf that is consistent with the prior information and the observed data.

As will be shown in the following discussion, the EnKF entails running a reservoir simulator for each model realization to compute the model responses and also to estimate the various quantities associated with the EnKF formulation. This step may be computationally demanding and time consuming for field scale applications. Consequently, to aid in a reduction of computational burden, the size of the ensemble must be kept as small as practically feasible.

The Reservoir State Vector. The state vector, denoted by \mathbf{y} includes the vector of the model unknowns, \mathbf{m} , comprising static variables \mathbf{m}^s (e.g. permeability, porosity) and the dynamic variables \mathbf{m}^d (e.g. pressure, phase saturation). The state vector is augmented by the production data \mathbf{d} (e.g. bottom-hole pressure, water-cut and gas-oil ratio at the wells). The length of the vector \mathbf{m} is N_m and the number of measurements is N_d . Consequently, the relationship between the data and the model variables at time k is expressed as

$$\mathbf{d}_k = \mathbf{H}\mathbf{y}_k + \boldsymbol{\varepsilon} = \mathbf{H} \begin{bmatrix} \mathbf{m} \\ \mathbf{d} \end{bmatrix}_k + \boldsymbol{\varepsilon} \dots\dots\dots 1.1$$

The noise in the data, $\boldsymbol{\varepsilon}$, is a vector of the unknown measurement errors and is assumed Gaussian with a zero mean. Additionally, the component of the vector $\boldsymbol{\varepsilon}$ associated with different measurements is assumed independent. The measurement operator \mathbf{H} is a trivial matrix consisting of 0's and 1's and we can arrange \mathbf{H} as

$$\mathbf{H} = [\mathbf{0} \quad \mathbf{I}] \dots\dots\dots 1.2$$

\mathbf{I} is the identity matrix of dimension $N_d \times N_d$ and $\mathbf{0}$ has dimensions $N_m \times N_d$ so the measurement operator \mathbf{H} selects the vector \mathbf{d} from the augmented state vector \mathbf{y} .

The Measurement or Data Vector. As previously outlined, the vector \mathbf{d} typically consists of measurements at the well locations, for e.g. measurements of water-cut and gas-oil ratio and bottom-hole pressure. Additionally, 4-D time-lapse seismic data may also be used as a measurement. The EnKF is capable of assimilating diverse data types and the amount and type of measurement can vary with time.

Regardless of the source of the data, there is a noise component associated with each measurement. The noise is assumed to be Gaussian with a zero mean and is assumed to be uncorrelated. Moreover the measurement errors are assumed to be independent. Consequently, the data covariance matrix, \mathbf{C}_D , given by $\mathbf{C}_D = E[\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T]$ is principally a diagonal matrix. Or conversely, this can be written as $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{C}_D)$.

Since, the EnKF operates in an ensemble framework, additional quantities like the sample state vector covariance matrix may be estimated at any time during the assimilation sequence and is given by the following expression:

$$\mathbf{C}_y = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} [\mathbf{y}_i - \bar{\mathbf{y}}] [\mathbf{y}_i - \bar{\mathbf{y}}]^T \quad \dots\dots\dots 1.3$$

N_e is the number of model realizations or the ensemble size. The augmented state vector consists of the model state and the data vectors; therefore the state vector covariance matrix is equivalent to,

$$\mathbf{C}_y = \begin{bmatrix} \mathbf{C}_M & \mathbf{C}_{M,d} \\ \mathbf{C}_{M,d}^T & \mathbf{C}_{d,d} \end{bmatrix} \quad \dots\dots\dots 1.4$$

\mathbf{C}_M is the model state covariance and specifies the spatial correlation amongst the model parameters and $\mathbf{C}_{d,d}$ is the data covariance. The non-diagonal sub-matrix $\mathbf{C}_{M,d}$ is the sample estimate of the cross-covariance between the data and the model variables.

1.2 Minimum Variance Estimator - MVE

The goal of the EnKF is to find the ‘best’ estimate of the state vector that is capable of reproducing the observed production history. The ‘best’ estimate should be optimal in some pre-defined sense. The conditions of optimality are specified by the algorithmic details of the data assimilation procedure. Bayesian approaches to data integration (Vega et al. 2004) often attempt to seek the estimate with the maximum posterior probability or the mode of the posterior distribution, often known as the *Maximum A posteriori* (MAP) estimate. In a Kalman filter framework, however, the optimal estimate is one that reduces the error covariance matrix associated with the updated or posterior estimate. The estimate thus obtained is referred to as a *Minimum Variance Estimator*.

The estimation error at time k is,

$$\mathbf{e}_k = \mathbf{y}_{true} - \mathbf{y}_{k,prior} \quad \dots\dots\dots 1.5$$

The term \mathbf{y}_{true} is the true (but unknown) state vector. The prior state vector refers to the initial estimate of the true state. The observed data, $\mathbf{d}_{obs,k}$ is used to improve the prior estimate. In the Kalman filter approach, the posterior or updated estimate is chosen to be a linear combination of the prior estimate and the data misfit between the observed data and the model predictions and can be specified as,

$$\tilde{\mathbf{y}}_k = \mathbf{y}_k + \mathbf{K}(\mathbf{d}_{obs,k} - \mathbf{H}\mathbf{y}_{k,prior}) \quad \dots\dots\dots 1.6$$

The matrix \mathbf{K} is referred to as the *Kalman Gain*. The posterior estimate is the MVE if it minimizes the estimation error covariance matrix given by,

$$\tilde{\mathbf{C}}_y = E(\tilde{\mathbf{e}}_k^T \tilde{\mathbf{e}}_k) \quad \dots\dots\dots 1.7$$

$$\tilde{\mathbf{e}}_k = \mathbf{y}_{true} - \tilde{\mathbf{y}}_k \quad \dots\dots\dots 1.8$$

Using the definition of the optimal estimate in Eq. 1.5 we can rewrite the estimation error covariance matrix as

$$\tilde{\mathbf{C}}_y = E[\tilde{\mathbf{e}}_k^T \tilde{\mathbf{e}}_k] = E[(\mathbf{y}_{true} - \tilde{\mathbf{y}}_k)^T (\mathbf{y}_{true} - \tilde{\mathbf{y}}_k)] \quad \dots\dots\dots 1.9$$

Substituting the updated $\tilde{\mathbf{y}}_k$ by its equivalent expression (Eq. 1.6) into Eq. 1.9, we obtain

$$\tilde{\mathbf{C}}_y = E[(\mathbf{y}_{true} - \mathbf{y}_{k,prior} - \mathbf{K}(\mathbf{d}_{k,obs} - \mathbf{H}\mathbf{y}_{k,prior}))^T \dots\dots \dots\dots (\mathbf{y}_{true} - \mathbf{y}_{k,prior} - \mathbf{K}(\mathbf{d}_{k,obs} - \mathbf{H}\mathbf{y}_{k,prior}))] \quad \dots\dots\dots 1.10$$

Recalling that the vector $(\mathbf{y}_{true} - \mathbf{y}_{k,prior})$ is the prior estimation error \mathbf{e}_k , and assuming that this error is uncorrelated with the measurement error \mathbf{e}_k we can rewrite Eq. 1.10 as

$$\tilde{\mathbf{C}}_y = E[(\mathbf{e}_k - \mathbf{K}(\mathbf{H}\mathbf{e}_k - \boldsymbol{\varepsilon}_k))^T (\mathbf{e}_k - \mathbf{K}(\mathbf{H}\mathbf{e}_k - \boldsymbol{\varepsilon}_k))] \dots\dots\dots 1.11$$

The covariance matrices \mathbf{C}_D and \mathbf{C}_y are defined as $\mathbf{C}_D = E[\boldsymbol{\varepsilon}_k^T \boldsymbol{\varepsilon}_k]$ and $\mathbf{C}_y = E[\mathbf{e}_k^T \mathbf{e}_k]$ and the posterior estimation error covariance can be rewritten as

$$\tilde{\mathbf{C}}_y = \mathbf{C}_y - \mathbf{K}\mathbf{H}\mathbf{C}_y - \mathbf{C}_y\mathbf{H}^T\mathbf{K}^T + \mathbf{K}(\mathbf{H}\mathbf{C}_y\mathbf{H}^T + \mathbf{C}_D)\mathbf{K}^T \dots\dots\dots 1.12$$

The diagonal terms of the matrix $\tilde{\mathbf{C}}_y$ represent the estimation error variance for the elements of the posterior state vector and so to obtain the MVE, we need to find \mathbf{K} that minimizes the trace of the matrix.

$$\frac{d[\text{trace } \tilde{\mathbf{C}}_y]}{d\mathbf{K}} = -2(\mathbf{H}\mathbf{C}_y)^T + 2\mathbf{K}(\mathbf{H}\mathbf{C}_y\mathbf{H}^T + \mathbf{C}_D) \dots\dots\dots 1.13$$

Setting the above derivative to zero and solving for the optimal Kalman gain \mathbf{K} , we get,

$$\mathbf{K} = \mathbf{C}_y\mathbf{H}^T (\mathbf{H}\mathbf{C}_y\mathbf{H}^T + \mathbf{C}_D)^{-1} \dots\dots\dots 1.14$$

The computed value of \mathbf{K} minimizes the mean square estimation error and posterior estimated $\tilde{\mathbf{y}}$ obtained by Eq. 1.15 is also known as the *minimum variance estimator* or MVE.

$$\tilde{\mathbf{y}} = \mathbf{y}_{\text{prior}} + \mathbf{C}_y\mathbf{H}^T (\mathbf{H}\mathbf{C}_y\mathbf{H}^T + \mathbf{C}_D)^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{H}\mathbf{y}_{\text{prior}}) \dots\dots\dots 1.15$$

$$\tilde{\mathbf{C}}_y = \mathbf{C}_y - \mathbf{C}_y\mathbf{H}^T (\mathbf{H}\mathbf{C}_y\mathbf{H}^T + \mathbf{C}_D)^{-1} \mathbf{H}^T \mathbf{C}_y \dots\dots\dots 1.16$$

1.3 EnKF Workflow

The major steps involved in the EnKF are:

Ensemble Forecast Step. We compute the dynamic response of each realization in the ensemble to the time when we have observation data available. This step is the forecast step implying that we predict production response and the model state (pressures, phase saturations) at a particular time. It involves the use of a numerical reservoir simulator for each realization.

Estimating the Kalman Gain. The predicted production response from each realization and the variables of the state vector form the input to a cross-covariance estimation whereby the Kalman gain is computed.

The Ensemble Update. The data misfit vector is computed as the difference between observed production data and the calculated responses. The cross-covariances computed in the previous step and the data misfit constitute the key terms in the update expression, as shown in Eq. 1.15.

For the next set of observations, the updated parameters and state variables are used in a forecast step and the previous steps are repeated. The forecast step involving the simulator run does not however require performing the simulation from the initial time onwards, but only entails a restart run from the previous assimilation time.

1.4 Application of the EnKF: Synthetic Example

In this section, the EnKF is applied to a synthetic example to illustrate the previously outlined procedure. The test case generated uses a reference permeability on a 50x50x1 grid system to generate ‘observed’ water-cut data. The model has three producers and one injector. Production data is generated for a period of 4000 days and 50 realizations conditioned to permeability data at the well locations are generated to constitute our initial ensemble. The generated permeability realizations and the reference permeability fields have a non-Gaussian histogram and this was chosen to highlight some of the difficulties in the application of the EnKF as will be demonstrated in a later section. The

reference log-permeability field and the position of the 4 wells in the field are shown in Figure 1.1.

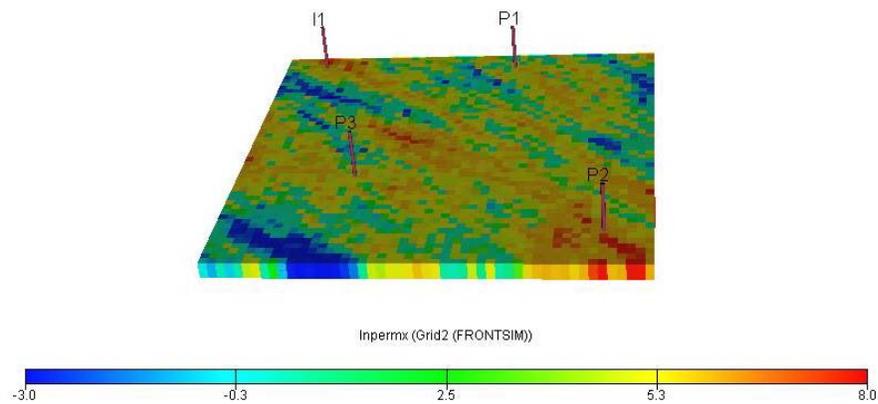


Figure 1.1 Reference log-permeability field and location of the 4 wells.

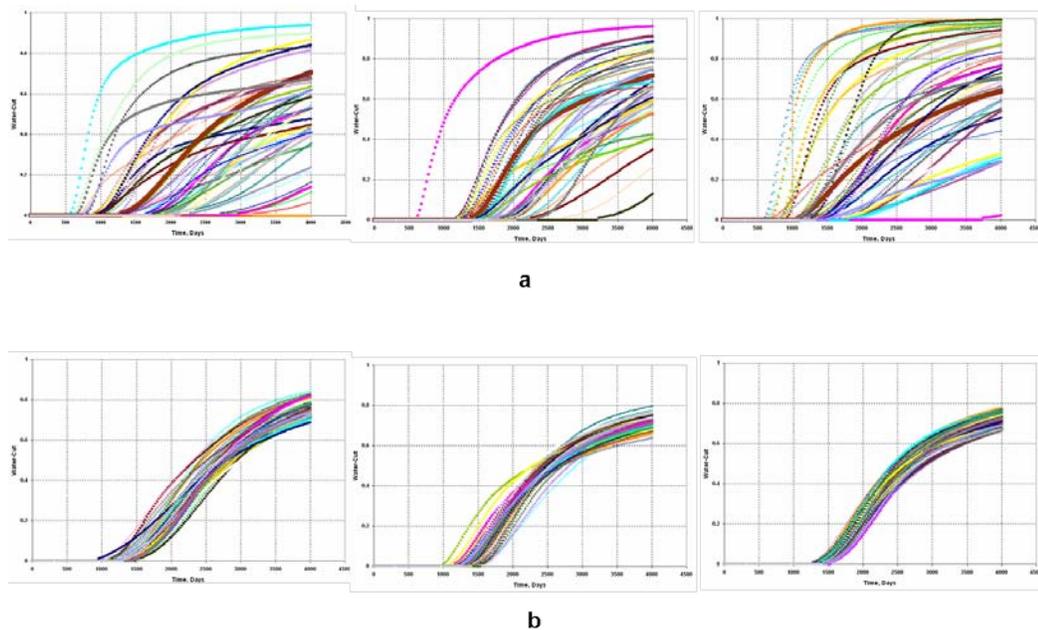


Figure 1.2 (a) is the chart of initial water-cut at 3 wells for the synthetic example for 50 members and (b) shows the water-cut from the history matched ensemble. The reference water cut is the profile in red.

Water-cut data was sequentially assimilated for a period of 3000 days to condition the permeability fields to the observed data from the reference field. After conditioning the models to observed data, flow simulation using the updated permeability fields provided the entire predicted water-cut for 4000 days as seen in Figure 1.2. The entire suite of models show that the updated realizations match observed history quite well in the first 3000 days and the variability in forecasts from 3000 days to 4000 days is reduced.

1.5 Difficulties in Data Assimilation Using the EnKF

In spite of all of its favorable properties, the EnKF implementation comes with its own share of challenges. This section highlights some of the difficulties arising from the use of the EnKF in data assimilation for the purposes of reservoir characterization. The following discussion focuses on the constituent elements of a reliable history matching and uncertainty assessment algorithm as an aid in understanding the issues associated with EnKF performance.

1.5.1 Key Aspects of Reliable Data Assimilation

An important aspect of history matching is to generate acceptable models that can reproduce historical production data and simultaneously maintain geologic realism. Significant time and effort employing extensive geologic studies, seismic interpretation studies and geostatistical methods are routinely spent in constructing the highly detailed initial reservoir description. Consequently, it is necessary to address the issue of geologic consistency between the pre and post assimilation reservoir models. Most inversion based schemes are formulated around norm and regularization constraints to prevent divergence of the solution from the initial geologic model (Cheng et al. 2004; Oyerinde 2008). However, in the EnKF framework, the explicit specification of the prior covariance matrix in the formulation of the equations is the only control to ensure geologic consistency between the prior and posterior models. In essence, the final updated model should retain the key underlying features of the initial model, for e.g.

connectivities between extremes and the general orientation of the significant geologic features.

The issue of geologic consistency is also tied to the non-uniqueness of the history matching problem. Non-uniqueness of the solution in data assimilation arises due to the fact that the number of parameters to be estimated is oftentimes orders of magnitude higher than the number of data available. Consequently, although a suite of updated realizations may be able to reproduce production history within acceptable tolerances, the data assimilation algorithm may not necessarily construct reliable reservoir performance models and may exhibit a loss of geologic realism. This has serious ramifications for future reservoir performance predictions and can lead to erroneous reservoir management strategies.

Another significant goal of any data assimilation methodology is to provide a reasonable framework for uncertainty assessment. The validity of the uncertainty quantification will strongly depend upon the distribution of the model realizations and whether they adequately represent the underlying uncertainties. In the context of the EnKF, the variability associated with the realizations within the ensemble provides a means for uncertainty assessment.

1.5.2 Assessment of the EnKF

This section focuses on the difficulties in the application of the EnKF to reservoir characterization studies. To assess EnKF performance, the model setup used is the same as in section 1.2.2. Water-cut data is assimilated to estimate the spatial distribution of permeability within the synthetic test reservoir. The reference permeability field used to generate the observed water-cut data is shown in Figure 1.1. An initial ensemble comprising 25 members is generated using SGSIM (Deutsche and Journel 1992) by conditioning the realizations to permeability data at the well locations. The initial and updated permeability fields for four representative realizations are shown in Figure 1.3. The most apparent and striking feature of the updated fields is the strong resemblance of each of the realizations. The loss of variability between ensemble members results in a

higher confidence placed on the prior models since the model covariance matrix is approximately zero ($C_y \sim \mathbf{0}$). As previously outlined, this results in a negligible Kalman gain as shown in Eq. 1.15 and therefore the EnKF tends to ignore subsequent data.

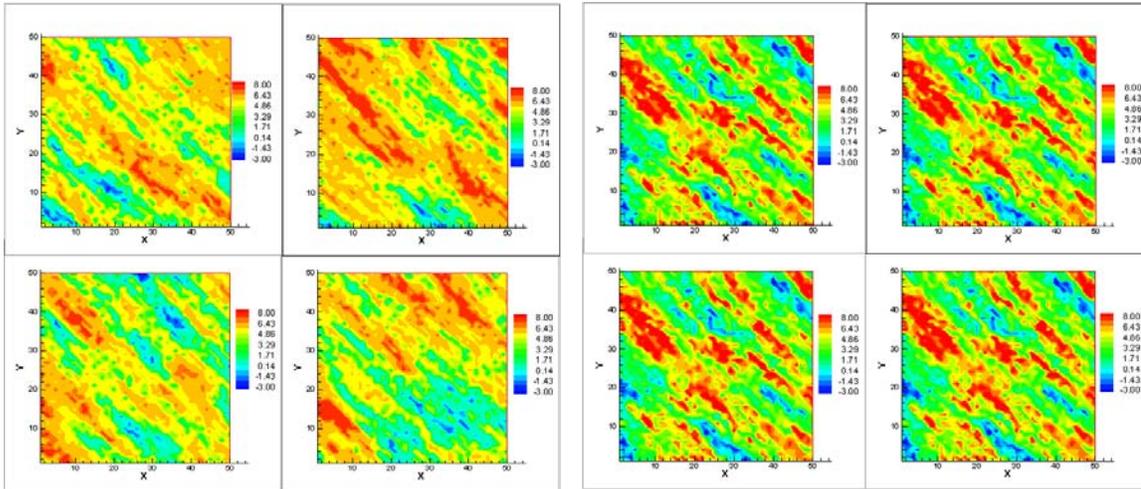


Figure 1.3 Comparison of the updated permeability fields on the right with the initial fields on the left for four realizations of the ensemble of 25 members.

The EnKF implementation also adversely impacts the permeability histograms. Figure 1.4 shows the histogram of log-permeability for one realization pre and post assimilation. The initial bi-model permeability histogram is clearly transformed to a more Gaussian histogram. This transformation implies that the EnKF methodology may not preserve the relative proportions of low quality and high permeability reservoir quality rock and the associated heterogeneities.

In addition to above mentioned problems, the updated models also display characteristics of unrealistic localized high and low permeabilities. This observation is also supported by the extremely wide range of log-permeabilities in Figure 1.4. A considerable number of model cells exhibit extremely high or low values of permeability. This observation is consistent with previous findings (Gu and Oliver 2006;

Arroyo et al. 2006, Devegowda et al. 2007) where the EnKF was observed to create localized zones of geologically inconsistent extremes of permeability values.

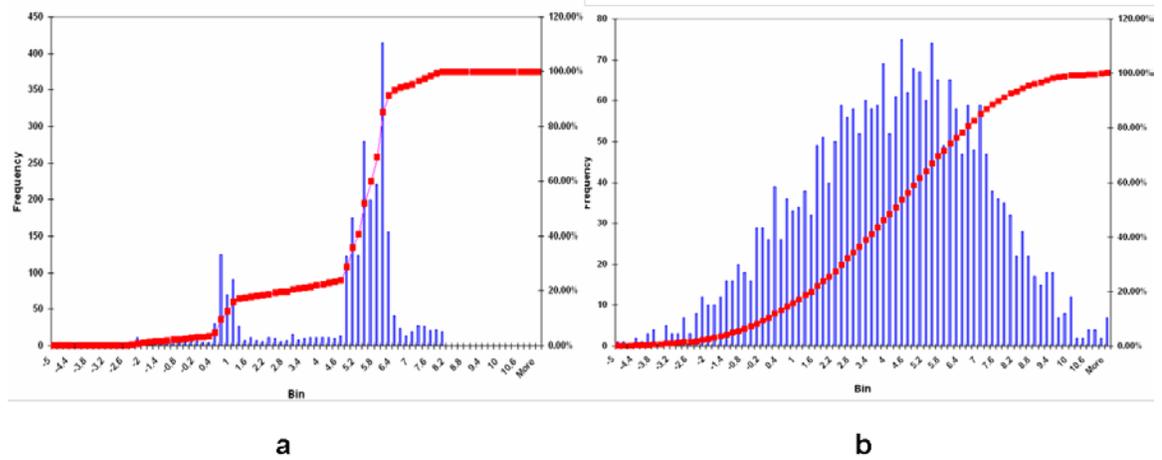


Figure 1.4 (a) is the initial permeability histogram on the left with the updated histogram on the right in (b) for an ensemble size of 25.

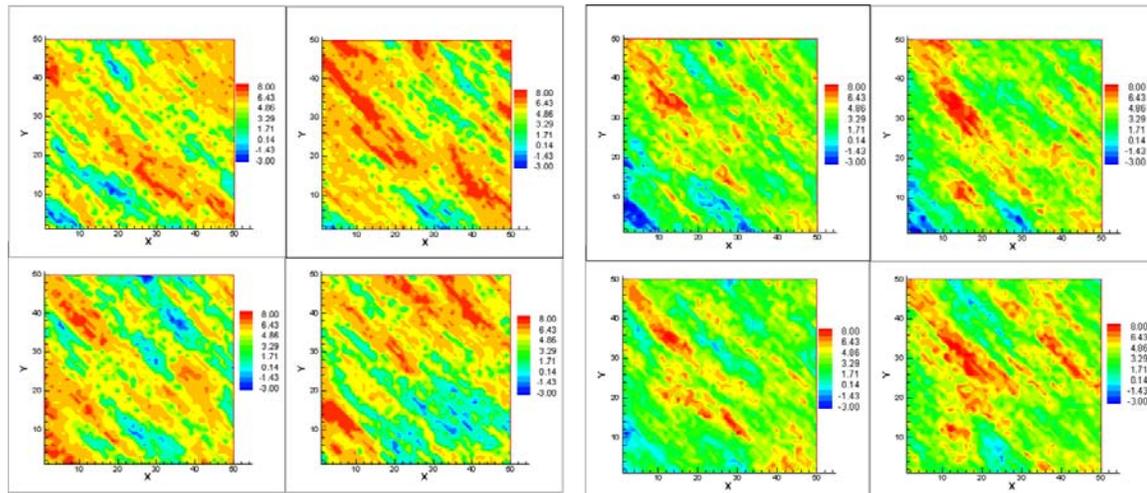


Figure 1.5 Comparison of the updated permeability fields on the right with the initial fields on the left for four realizations of the ensemble of 50 members.

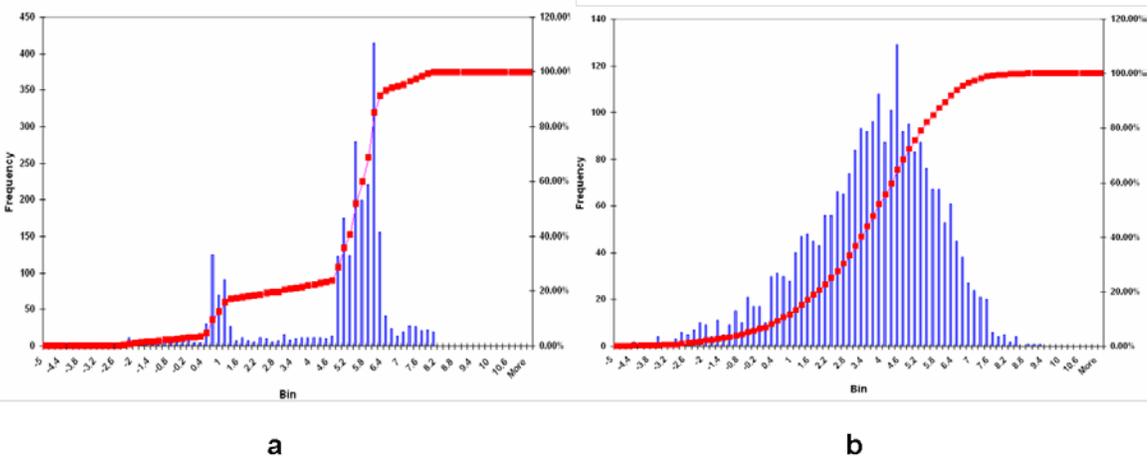


Figure 1.6 (a) is the initial permeability histogram on the left with the updated histogram on the right in (b) for an ensemble size of 50.

Although the severity of these problems is reduced with increasing ensemble size as seen in Figures 1.5 and 1.6 ($N_e = 50$) and in Figures 1.7 and 1.8 ($N_e = 100$), reliable reservoir description becomes difficult with the traditional EnKF implementation.

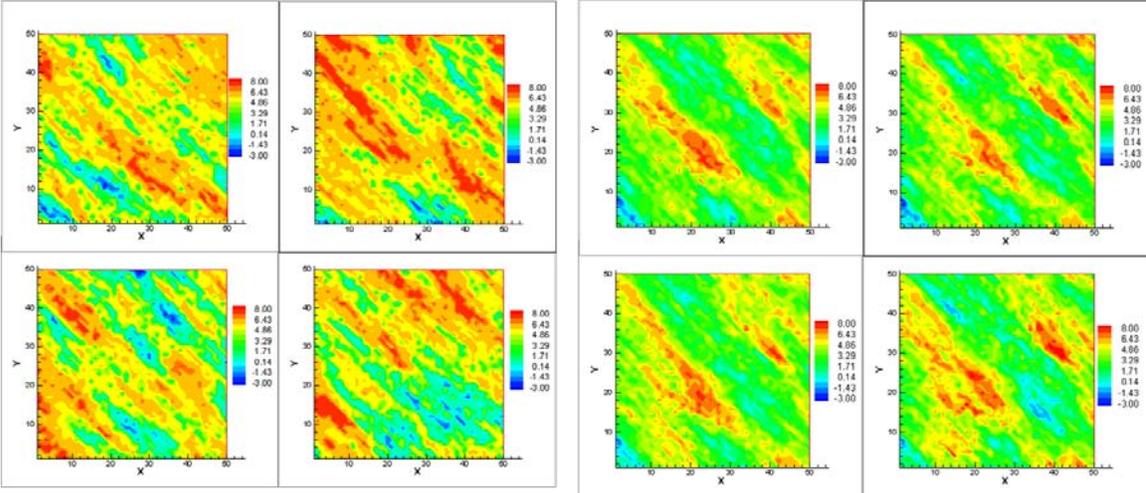


Figure 1.7 Comparison of the updated permeability fields on the right with the initial fields on the left for four realizations of the ensemble of 100 members.

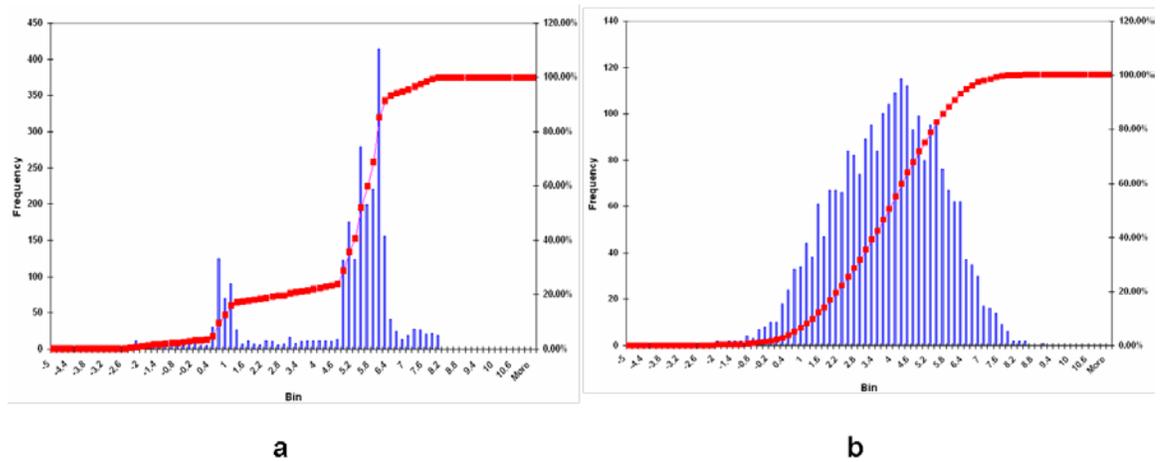


Figure 1.8 (a) is the initial permeability histogram on the left with the updated histogram on the right in (b) for an ensemble size of 100.

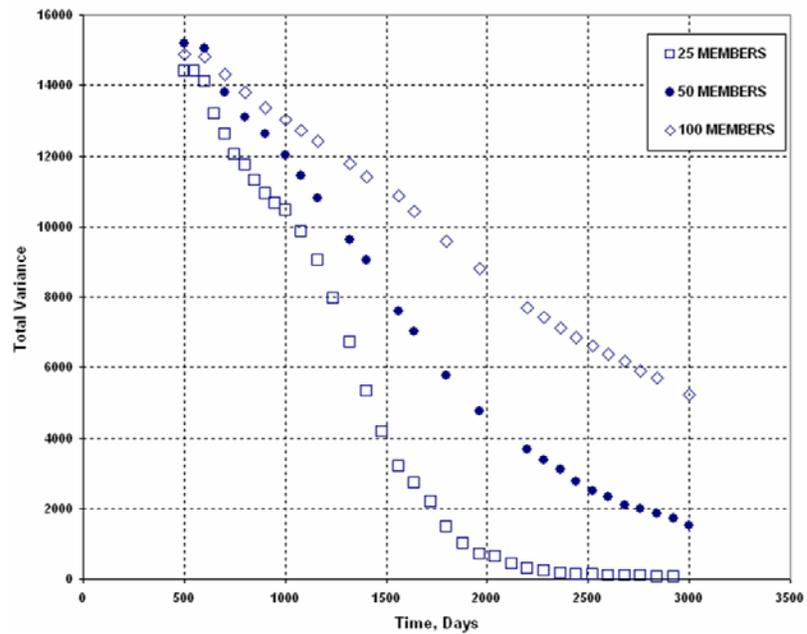


Figure 1.9 Comparison of the total variance in the permeability as a function of time during the assimilation process for 3 different ensemble sizes.

These conclusions are supported by the quantitative assessment of the loss of variability as shown in Figure 1.9 which plots the total variance versus time during the assimilation sequence. The total variance of the model parameter covariance matrix is easily computed as the sum of its eigenvalues.

The difficulties associated with the implementation of the EnKF are directly related to the accuracy of the sample covariance matrices. The constituent elements of the EnKF formulation are functions of the sample or ensemble covariance or cross-covariance matrices. As with any statistical measure computed using a finite number of samples, the accuracy of the computed quantity is strongly dependent on ensemble size. Consequently, increasing ensemble size implies better covariance estimation and in the limit of an infinite number of realizations, the covariance matrix may be assumed to be close to the truth. However, this goes against the need for smaller ensemble sizes and the associated computational efficiency.

1.6 Conclusions and Research Objectives

The primary purpose of this research work is to present a robust technique to improve EnKF performance that addresses the difficulties associated with the EnKF implementation and that is applicable to large field scale problems. The proposed technique is dependent on the use of streamline-derived information. Although, the technique is illustrated in the streamline domain, it is equally effective using finite-difference simulation models.

In section 2, the discussion in section 1 is extended to techniques that attempt to counter the adverse effects of small sample sizes. Section 3 presents the proposed solution to improve EnKF performance and illustrates the strength and utility of the approach using synthetic and field examples. These techniques exploit streamline-specific information to address issues related to sampling variability, sampling error and spurious correlations attributed primarily to the use of a small number of model replicates.

A significant advantage of the EnKF is its ability to assimilate diverse data types. This makes the EnKF a particularly appealing option for multi-scale data assimilation. The formalism of the multi-scale EnKF is presented in section 4 and is extended to the interpretation of partitioning interwell tracer tests (PITT) to identify the spatial distribution of bypassed oil targets which is of paramount significance to the successful implementation of enhanced recovery processes.

2. DATA ASSIMILATION: IMPROVED VARIANTS OF THE EnKF

The previous section provided the mathematical details related to the EnKF and the associated complications for application to large-scale problems. The discussion in this section forms the background to section 3 where we cover in detail the approach that forms the focus of this study. This section is divided into three sections discussing different approaches to address the computational and sampling-based difficulties with ensemble filtering. Each approach and the associated drawbacks or benefits are examined primarily to assess the use of the technique for detailed reservoir description.

2.1 Distance Dependant Covariance Localization

A key component of the Kalman gain is the cross-covariance calculation between the data and model variables as shown in Eq. 1.18 that quantifies the corresponding strength of association. This estimate is critically dependant on the number of random model samples. The EnKF uses these sample-based statistics to weight and distribute information from observations (well data) onto the underlying grid. Since the EnKF updates are highly sensitive to the Kalman gain, it stands to reason that a poor estimate of the cross-covariance would severely compromise EnKF performance (Anderson and Anderson 1999; Furrer and Bengtsson 2004). For small ensemble sizes with noisy and possibly erroneous cross-covariances, unrealistic updates and degraded EnKF performance is often observed (Naevdal et al. 2005; Gu and Oliver 2006; Arroyo et al. 2006; Devegowda et al. 2007). The complications associated with sampling variability can be severe when the value of the true correlation is small; in such scenarios, the noise significantly overwhelms the signal. This is illustrated in Figure 2.1. The cross-covariance between water-cut data and the permeability is shown for one well for ensemble sizes of 50 and 1000. The 2-dimensional model is discretized into 50x50 grid cells. The permeability in each cell is independent and identically distributed and

sampled from a normal distribution with a pre-specified mean and variance. The impact of increased ensemble size on the cross-covariance is clearly seen in terms of reduced noise and a more coherent cross-covariance structure.

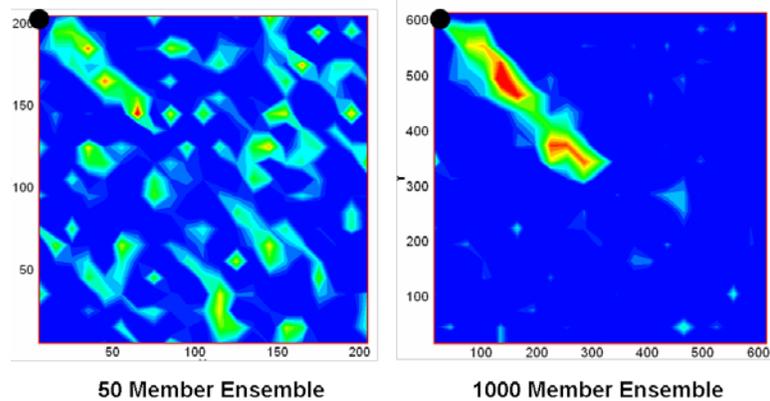


Figure 2.1 Cross-covariance between water-cut at one well (black dot) and the permeability for two different ensemble sizes.

Other studies (Furrer and Bengtsson 2004; Oke et al. 2007) have presented arguments to show that for bounded error in model covariance estimation, the size of the ensemble should vary as the square of the number of variables to be estimated ($N_e \sim N_M^2$). However, this may not be feasible or practical for most applications. Consequently, the need to accurately estimate covariance matrices for large scale systems has motivated the search for effective covariance localization methods. The aim of most covariance localization schemes is to

- Eliminate spurious terms in the cross-covariance matrix arising due to sampling errors caused by finite and small ensemble sizes and,
- Increase the effective number of ensemble members (Hamill and Whitaker 2001).

Distance based covariance localization schemes (Houtekamer and Mitchell 2001; Hamill and Whitaker 2001) are predicated on the assumption that the correlation between model

grid cells and well data is associated with certain length scales beyond which the correlation can be assumed to be zero. Within this characteristic length-scale centered at the location of the well, the cross-covariance is artificially damped with a localizing function that decreases monotonically from a value of 1.0 at the well location to zero in the region outside of the cut-off radius. This has the effect of spatially restricting and smoothing the statistical information from an observation to grid cells close to the measurement location. A sample localizing function is shown in Figure 2.2.

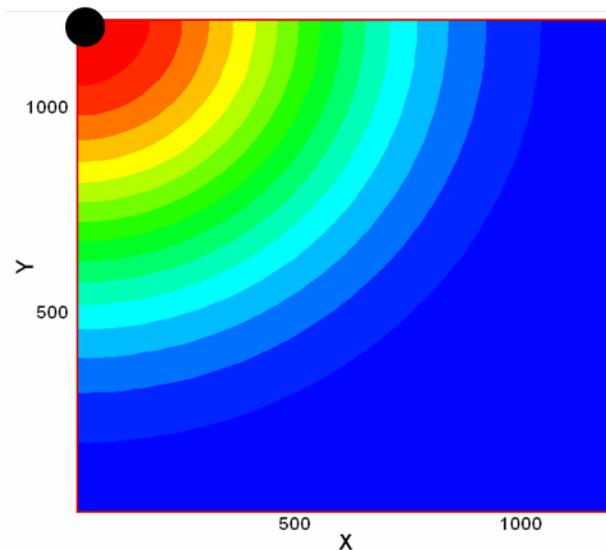


Figure 2.2 Distance based localizing function centered at the well shown in a black dot (Gaspari and Cohn 1996).

Mathematically, the localized EnKF can be expressed as

$$\tilde{\mathbf{y}} = \mathbf{y}_{prior} + \boldsymbol{\rho} \circ \mathbf{C}_y \mathbf{H}^T (\mathbf{H} \mathbf{C}_y \mathbf{H}^T + \mathbf{C}_D)^{-1} (\mathbf{d}_{obs} - \mathbf{H} \mathbf{y}_{prior}) \dots\dots\dots 2.1$$

where the localizing function $\boldsymbol{\rho}$ operates on the Kalman gain matrix. The operator \circ is an element-by-element multiplication also called a Schur product. Various valid

correlation functions are discussed further in Gaspari and Cohn (1996).

The following discussion demonstrates the application of distance based localization to the synthetic example introduced in section 1.2. A preliminary guess for the cut-off radius is 200 feet. In comparison, the synthetic model has dimensions 500 ft x 500 ft. As in the previous section, water-cut at each of the 3 production wells is used to condition the permeability field in each of the 50 realizations. The water-cut data is assimilated for a period of 3000 days. The initial water-cut predictions and the matched water-cut data from the updated realizations are shown in Figure 2.3.

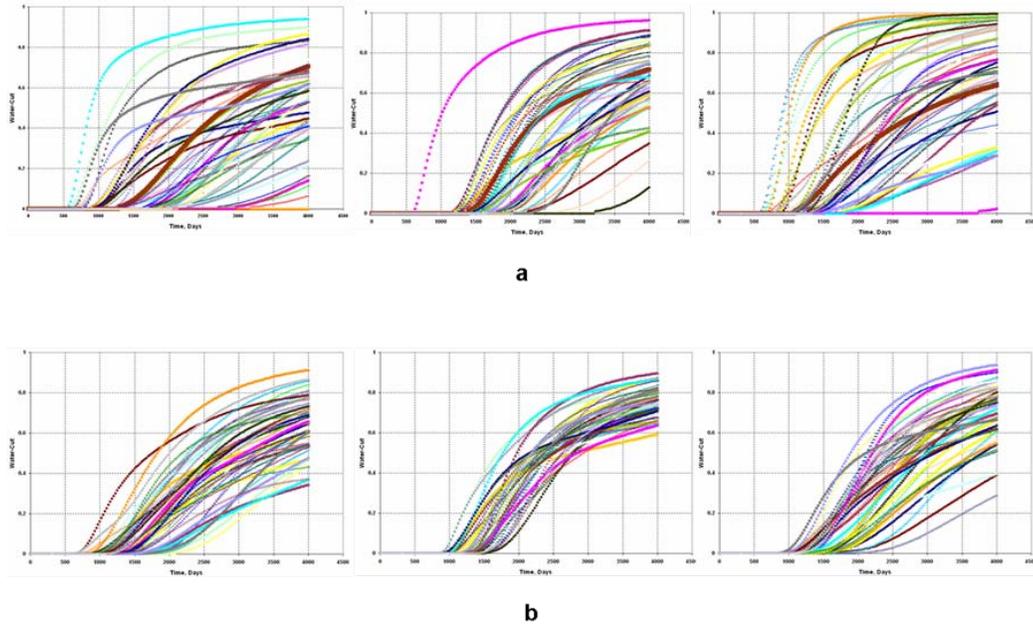


Figure 2.3 (a) is the initial water-cut at each of the 3 wells and (b) is the matched water-cut. The results do not appear to be satisfactory.

It is apparent from Figure 2.3 that the resulting water-cut predictions are not satisfactory and not within reasonable bounds. The poor match underscores the need for a proper choice of cut-off radius. The experiment is repeated with a cut-off radius of 400 feet and the results are shown in Figure 2.4. This choice seems to be more appropriate.

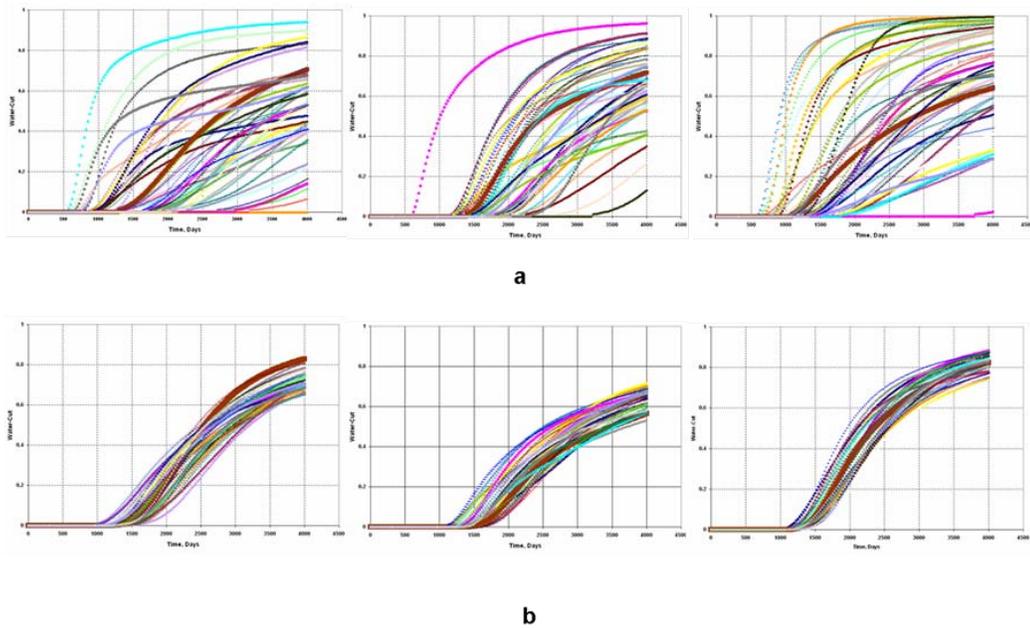


Figure 2.4 (a) is the initial water-cut at each of the 3 wells and (b) is the matched water-cut. The matches with a cut-off radius of 400 feet are considerably better.

In general, however, the selection of a characteristic length-scale is subjective and requires trial-and-error. Furthermore, the localization may not be consistent with the underlying heterogeneity that might include high permeability channels with long-range contributions (Arroyo et al. 2006). Furthermore, Hamill and Whitaker (2001) observe that the length scale chosen to achieve localization is a function of the ensemble size. When ensemble replicates are small in number, the cut-off radius is estimated to be smaller because short-range correlations are significantly corrupted by noise. As ensemble size increases, however, the effect of noise on the distant observations is diminished and calls for a larger radius of influence.

For 3-dimensional problems, an additional concern is the choice of a localizing radius in the horizontal and vertical directions, where the correlation lengths are considerably different. The optimal length-scale may likely be case-dependant and may be difficult to determine apriori. Consequently, for detailed reservoir characterization applications, distance based localization may not be robust nor efficient and may have

limited applicability. This underscores the requirement for more satisfactory and reliable localization schemes that are tied to the underlying geology.

2.2 Covariance Inflation

Covariance inflation (Anderson 2001) is an ad-hoc tuning method specifically employed to target the loss of variability as demonstrated in section 1 which results in overconfident priors. This critically impacts the assimilation process because observations have progressively lesser weighting in the EnKF. The method relies on artificially broadening the prior distribution of the state variables to counter the problem of variance-deficiency. The individual state vectors can be generated by linearly inflating the ensemble around the mean,

$$\mathbf{y}_{i,\text{inf}} = \theta(\mathbf{y}_i - \bar{\mathbf{y}}) + \bar{\mathbf{y}} \dots\dots\dots 2.2$$

where θ is referred to as the *covariance inflation factor* and has values larger than one.

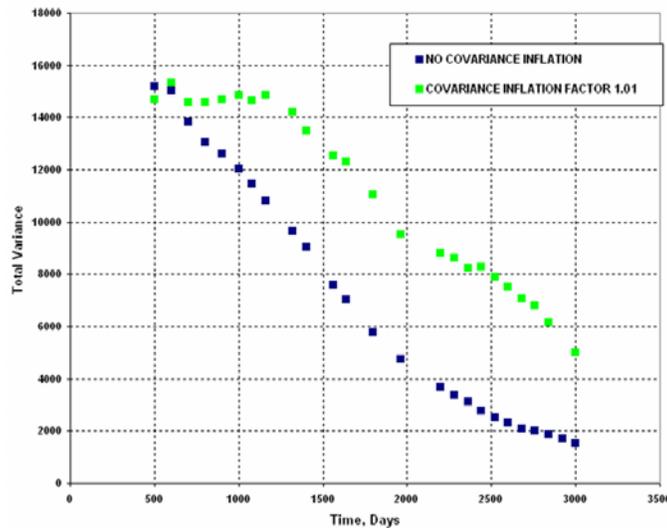


Figure 2.5 Plot of total variance versus time during assimilation comparing the covariance-inflated EnKF versus the standard EnKF for an ensemble size of 50.

By reducing the confidence in the prior using Eq. 2.2, covariance inflation also preserves the mean and the relative values of the prior model variable correlations. A suitable value of the inflation factor is chosen based on trial-and-error and can range from 1 to 5% amplification of the variance. To investigate the possible benefits of covariance inflation, this section presents results from an experiment using 50 ensemble members and a covariance inflation factor of 1.01 and utilizes the same example presented in section 1.2. The variability of the ensemble models is improved with covariance inflation and is illustrated in Figure 2.5 which plots total variance versus time during the assimilation process.

Since the method only targets the loss in ensemble variability, it does not appear to be a promising alternative especially when seen in the light of Figure 2.6 which plots the log-permeability histogram for one arbitrarily chosen realization. The transformation from a bimodal histogram to a more normal histogram is particularly evident.

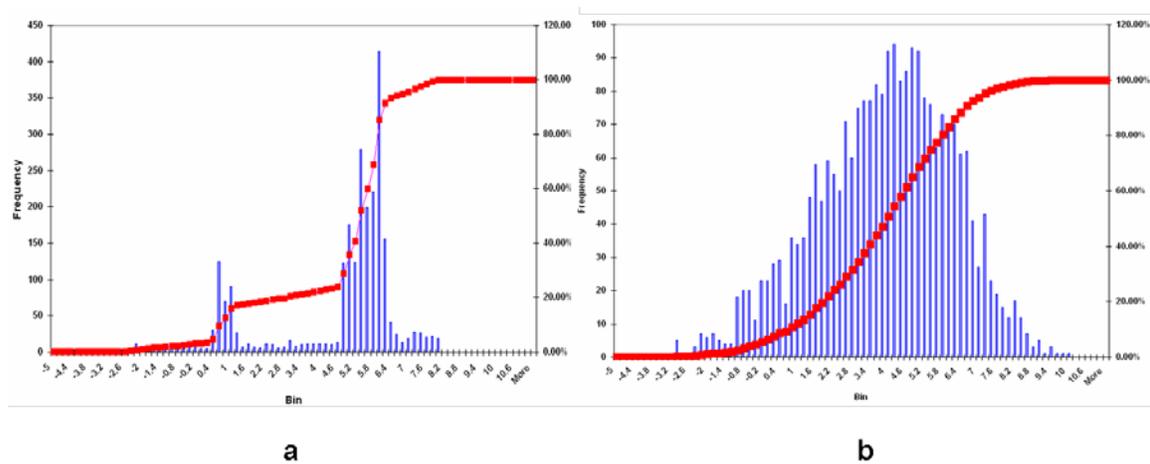


Figure 2.6 The initial log-permeability histogram on the left compared to that from an updated realization clearly shows the transformation from bimodal to Gaussian.

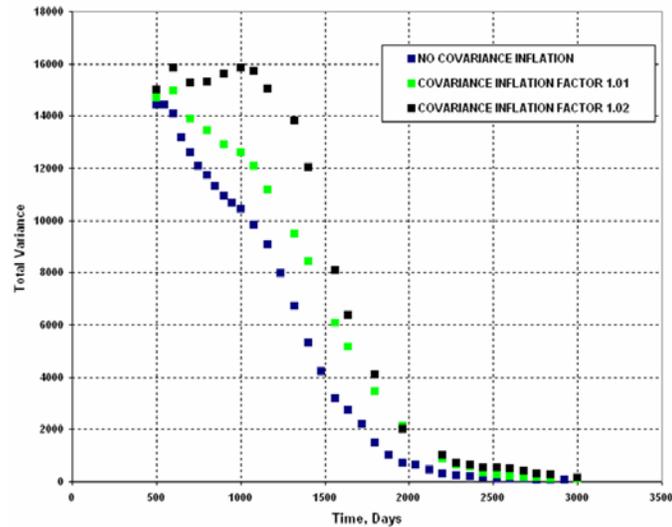


Figure 2.7 Covariance inflation may not be suitable for a small ensemble size (~ 30 model realizations).

The individual realizations also exhibit problems related to overshooting and this may be exacerbated by the covariance inflation scheme. The limited applicability of covariance inflation is compounded by the heuristics involved in the selection of the inflation factor. In our experience, the success of this technique is closely linked to the ensemble size. For smaller ensemble sizes (~ 30 members), the loss of variability is as rapid as with the standard implementation of the EnKF (Figure 2.7). This is likely a result of the inability of the methodology to suppress spurious and noisy correlations. However, it is likely that this technique might provide more satisfactory results when used in conjunction with other improved variants of the EnKF in conjunction with the use of a reasonable number of model replicates.

2.3 Smoothness-Constrained 2-Stage EnKF

The set of plausible and geologically consistent models is a subset of all possible solutions that will be able to reproduce historical production. The non-uniqueness of the solution therefore might warrant the use of additional constraints to restrict the possible

class of solutions. Johns and Mandel (2005) proposed the use of a 2-stage EnKF that imposes a measure of smoothness upon the EnKF update for wildfire modeling. The method is motivated by the fact that point observations (for e.g. well data) can resolve only large-scale features and resolution of the fine scale variability is limited. Spatial averaging is inherent in reservoir description from production data. Furthermore, the updates are tied to the prior model by the smoothness constraint providing an additional control to preserve geologic realism. This approach is analogous to the use of the regularization term in deterministic inverse modeling (Oyerinde 2008).

As shown in Eq. 1.17 the EnKF update can be derived from the minimization of an objective function comprising a prior term and a term related to the data misfit. The objective function is provided here for completeness and can be expressed as,

$$F(\mathbf{y}) = (\mathbf{y} - \mathbf{y}_{prior})^T \mathbf{C}_y^{-1} (\mathbf{y} - \mathbf{y}_{prior}) + \dots \dots \dots$$

$$\dots \dots \dots (\mathbf{d}_{obs} - \mathbf{H}\mathbf{y}_{prior})^T \mathbf{C}_D^{-1} (\mathbf{d}_{obs} - \mathbf{H}\mathbf{y}_{prior}) \dots \dots \dots 2.3$$

To specify a smoothness constraint, the objective function is augmented by another misfit term related to the model smoothness by the use of a spatial difference operator and the 2-stage EnKF is derived by a minimization of

$$F(\mathbf{y}) = (\mathbf{y} - \mathbf{y}_{prior})^T \mathbf{C}_y^{-1} (\mathbf{y} - \mathbf{y}_{prior}) + \dots \dots \dots$$

$$\dots \dots \dots (\mathbf{d}_{obs} - \mathbf{H}\mathbf{y}_{prior})^T \mathbf{C}_D^{-1} (\mathbf{d}_{obs} - \mathbf{H}\mathbf{y}_{prior}) + \dots \dots \dots 2.4$$

$$\dots \dots \dots (\mathbf{y} - \mathbf{y}_{prior})^T \mathbf{A}^T \mathbf{S}^{-1} \mathbf{A} (\mathbf{y} - \mathbf{y}_{prior})$$

where the matrix \mathbf{A} is a suitably chosen spatial first derivative and \mathbf{S} is a pre-specified diagonal positive definite covariance matrix.

Deviations from the prior smoothness are penalized by the diagonal covariance matrix \mathbf{S} . In this study, $\mathbf{S} = |\mathbf{A}\mathbf{y}_{mean}|$. When specifically applied to the permeability distribution, smaller terms refer to spatial patterns where the permeability variations are

small and larger terms imply a larger absolute permeability difference between adjacent grid cells. This particular choice of penalization ensures that smoother regions in the prior are preserved in the updated model and there is lesser control in areas where the initial model has large local spatial variations (for e.g. next to faults, channel boundaries etc.)

To implement this particular form of the EnKF, the problem is split in to two stages. The first stage incorporates the data misfit to condition the model states and the second stage imposes the smoothness constraint. Consequently, the procedure can also be viewed as the assimilation of two independent observations sequentially, the data specified by d_{obs} followed by the additional constraint. This approach has the advantage of being able to utilize the same EnKF analysis code for both the stages without any modifications.

While there are various options to specify a measure of spatial smoothness, for this study, we use a 5 point difference operator in the x- and y- directions (for a 2-D example) and a 7 point difference operator in 3 dimensions. Specifically, the matrix \mathbf{A} is defined by applying the following difference operator at each grid cell which is represented for 2 dimensional cases as a difference kernel shown in Eq. 2.5.

$$\text{Kernel} \equiv \begin{bmatrix} & -1 & \\ -1 & +4 & -1 \\ & -1 & \end{bmatrix} \dots\dots\dots 2.5$$

In other words, for any location (i,j), the difference kernel computes the smoothness as

$$(\mathbf{A} \mathbf{y}_{\text{mean}})_{i,j} = 4y_{i,j} - y_{i,j-1} - y_{i,j+1} - y_{i-1,j} - y_{i+1,j} \dots\dots\dots 2.6$$

To examine the effectiveness of this technique, it is applied to the same synthetic example introduced in previous sections. In Figure 2.8, we show the matches to the

water-cut data from the updated realizations in comparison with the initial ensemble. The method is able to reduce the considerable variability in the initial predictions; however, the profiles of the final match may not constitute a reasonable prediction envelope. An optimal choice of the smoothness matrix, \mathbf{A} , with a directional preference in constructing a spatial difference kernel based on prior geological considerations might considerably improve the quality of the results.

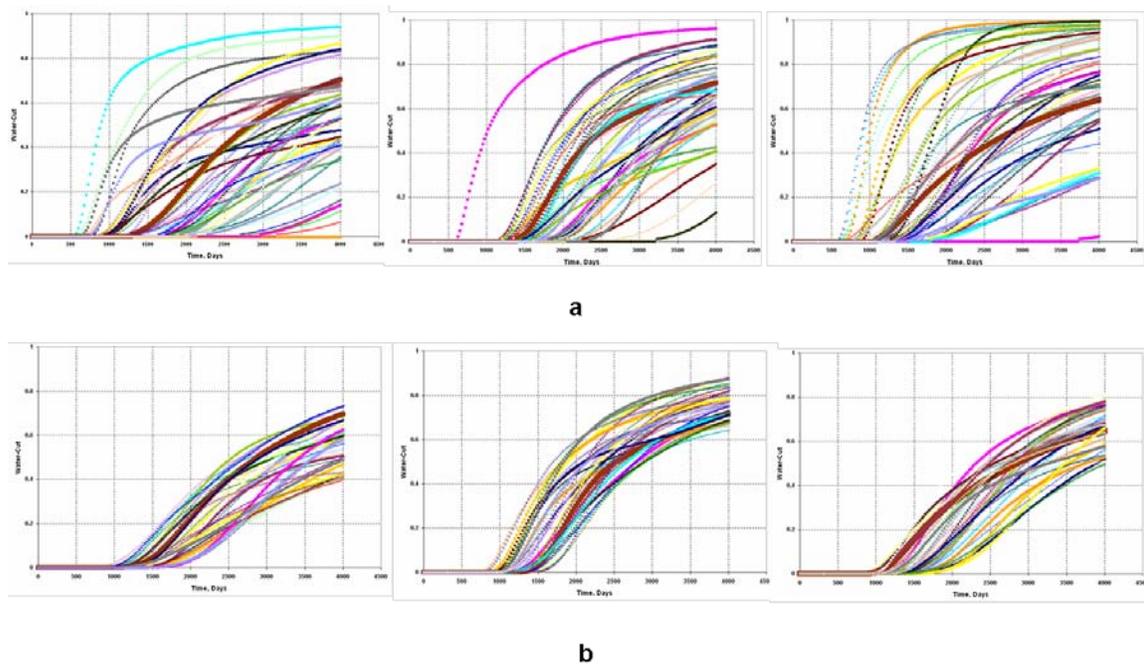


Figure 2.8 (a) is the water-cut prediction from the ensemble of 50 realizations and (b) shows the matches to the data from the updated models. The final model predictions are not within reasonable bounds.

Figure 2.9 displays a log-permeability histogram from one arbitrarily selected member from the set of updated realizations. A smoothness constraint imposed on the EnKF solution appears to be a significant enhancement over the standard EnKF.

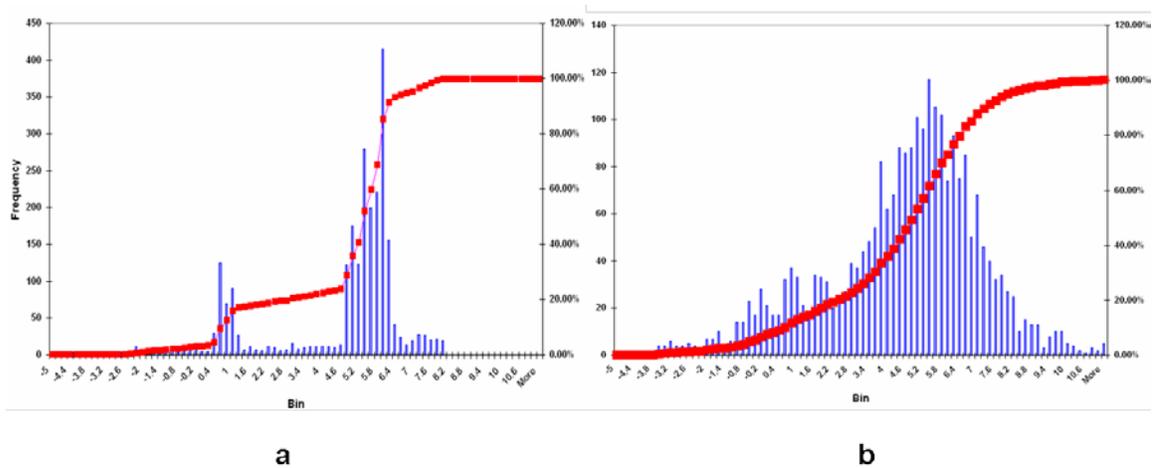


Figure 2.9 Updated log-permeability histogram in (b) in comparison with the initial histogram in (a).

For the hypothetical test study introduced in this section, this method appears most promising in comparison with distance-based localization and covariance inflation. This might be case dependent, however, principally; the approach of constraining the models to honor prior spatial smoothness is consistent with our goal of preserving geologic realism and the low resolution of the production data.

2.4 Section Summary

In this section, we briefly reviewed established techniques to improve EnKF performance. Assessment of each methodology was based on a synthetic example using an ensemble size of 50. The motivation for this part of the study was to investigate the possible benefits of each approach.

Distance-based isotropic localization assumes that long-range correlations beyond a pre-specified characteristic distance are very small. Selection of the length-scale is subjective, case-dependant and might be difficult to determine in advance. Therefore, its applicability is limited especially in highly heterogeneous 3-dimensional models primarily due to the subjective nature of the approach and because it is not tied

to the underlying heterogeneity.

Covariance inflation reduces the severity of variance-deficiency especially with modest ensemble sizes. The efficacy of the approach is possibly limited as a stand-alone fix to counter the adverse effects of sampling error in the EnKF. Variance-inflation may exaggerate the effect of spurious correlations especially for a very small ensemble sizes. The ad-hoc selection of the covariance inflation factor is another primary drawback of the approach. However, it might be a promising approach when used in conjunction with other methodologies.

The 2-stage EnKF reviewed in this section appears to be an attractive alternative without additional computational effort and has the added advantage of being able to utilize existing code with small modifications. The successful application to wildfire modeling where sharp fire fronts are a consistent feature of the ensemble (Johns and Mandel 2005) makes an extension to detailed reservoir description rather appealing and this is demonstrated with a synthetic example.

3. STREAMLINE BASED COVARIANCE LOCALIZATION

In an operational context, the standard implementation of the EnKF suffers from many drawbacks primarily because sample statistics (cross-covariance estimates) are critically dependant on the number of model realizations. Random noise can overwhelm the cross-covariance estimates especially when the ensemble size is modest and/or if the true correlation is small. The previous section focused on some currently suggested modifications to the EnKF algorithm to address these issues. However, these methods may not be compatible with coherent 3-D reservoir characterization and uncertainty assessment particularly for highly heterogeneous reservoirs.

In order to adequately address many of the currently reported difficulties in the use of the EnKF for detailed reservoir modeling, there is a strong need for a methodology that is tied to the physics of the problem. Streamlines and the information derived thereof offer particularly appealing alternatives with a sound physical basis to significantly improve the quality of the results. The use of streamline-based methods is physically intuitive and natural. The primary emphasis is to keep changes targeted to areas that have the largest influence on the solution. In the subsequent sections, the proposed approaches are discussed in detail. There are several advantages to the proposed approaches. The difficulties in the standard EnKF implementation (for e.g. overshooting, geological inconsistency, and loss of variability) are largely eliminated or mitigated in terms of their severity. All of this is achieved at significant cost savings using a limited set of realizations.

3.1 Streamline Trajectory Based Covariance Localization

In the previous section, we reviewed distance-based covariance localization as a means to improve EnKF performance specifically by modifying the cross-covariance calculations. However, as demonstrated, distance-dependant localization may not be

consistent with the underlying heterogeneity and in effect, the cut-off radius is a tuning parameter dependant on the size of the problem and number of model realizations. Potential filter divergence and poor state/parameter estimates are further drawbacks of the methodology particularly for reservoir modeling.

This section provides an introduction to streamline trajectory based covariance localization. Further details about the technique are available in Arroyo et al. (2006). The fundamental goal is to eliminate erroneous and noisy terms in the cross-covariance matrix thereby allowing for parameter updates that are consistent with the underlying geology and the production history. This can be achieved by demarcating areas within the reservoir that influence the observations strongly and then making the corresponding changes to the cross-covariance matrix. A robust and physically intuitive technique to identify zones that have significant impact on the observations is the use of streamline-derived information. Emanuel and Milliken (1998) make use of streamlines effectively for assisted history matching and the resulting methodology is shown to be superior to manual history matching and results in geologically relevant solutions.

To effectively localize regions corresponding to each observation, the only requirement is to establish whether a grid cell within the region is intersected by a streamline originating from the location of the observation. The column in the cross-covariance matrix corresponding to this particular observation will consequently only retain those terms referring to grid cells that are intersected by streamlines. The rest of the terms are set to zero. Within this framework, noisy terms arising due to sampling errors are eliminated if they are not compatible with the underlying flow field.

To implement this form of covariance localization using streamline trajectories, a zone common to the ensemble is selected by stacking the zones from each realization. An additional constraint may also be applied to ensure that the grid cell is common to at least a certain minimum number of realizations. The sequential nature of the EnKF implies that this area of influence is redrawn at every assimilation step to account for changing field conditions, additional infill drilling or simply to reflect changing well configurations. The basic definition developed here will be equally applicable to finite

difference models and are not restricted to streamline models alone. When using finite difference models, the streamlines are traced using Pollock's algorithm (Pollock 1998) or modified extensions to the tracing algorithm (Jimenez et al. 2007).

A visual representation of the spatial localization for all the wells in the synthetic example introduced in section 1 and 2 is shown in Figure 3.1 at a certain assimilation step. Clearly, the information provided by the streamline trajectories cannot be replicated by a naïve notion of distance. Some segments of the reservoir adjacent to the wells are not covered by streamlines and on the other hand, distant locations are included in the area of influence as dictated by the flow pattern.

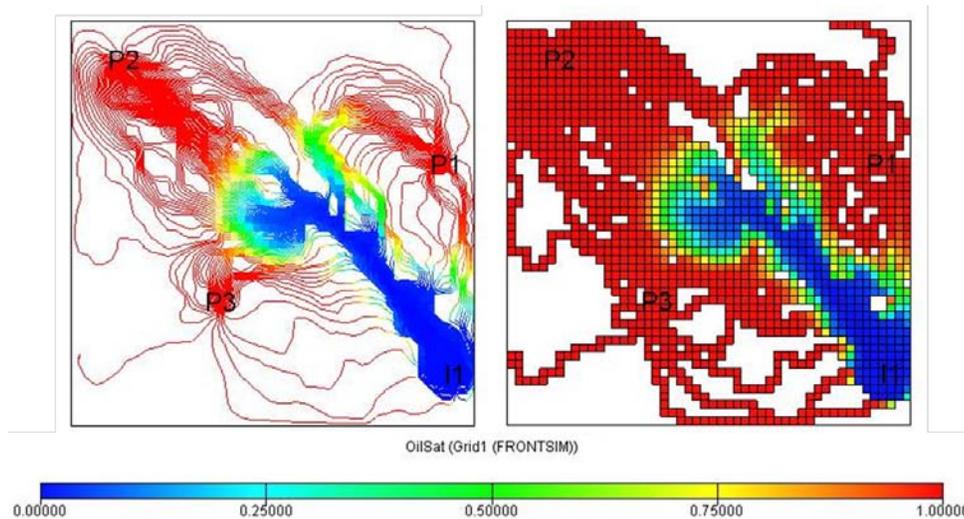


Figure 3.1 The area of influence for each observation on the right selected by the streamline pattern on the left.

Mathematically the localized covariance matrix is similar to Eq. 2.1, however, with the idea of utilizing streamline trajectories to retain relevant correlations and to eliminate noisy artifacts.

$$\tilde{\mathbf{y}} = \mathbf{y}_{prior} + \boldsymbol{\rho} \circ \mathbf{C}_y \mathbf{H}^T (\mathbf{H} \mathbf{C}_y \mathbf{H}^T + \mathbf{C}_D)^{-1} (\mathbf{d}_{obs} - \mathbf{H} \mathbf{y}_{prior}) \dots\dots\dots 3.1$$

The localizing function $\boldsymbol{\rho}$ is simply a matrix of 0's and 1's designed to select the appropriate grid cells relevant to the EnKF update.

To examine the benefits of this approach, streamline trajectory based localization is implemented using the above mentioned test case and compared to the traditional form of the EnKF. Figure 3.2 shows 4 randomly selected realizations updated using this technique. A simple visual inspection of the results shows no overshooting of the permeabilities, a strong coherent link between the pre and post assimilation models in terms of the underlying heterogeneity and reasonable inherent model variability. The bimodality of the updated histogram is also evident in Figure 3.3.

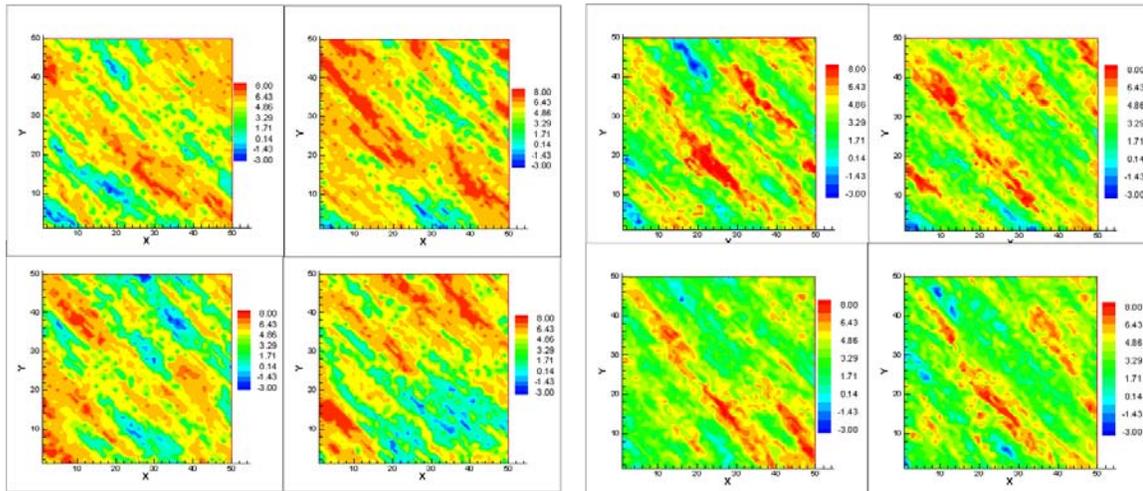


Figure 3.2 Four arbitrarily selected realizations on the left are updated using the streamline trajectory based covariance localization and the results are displayed on the right.

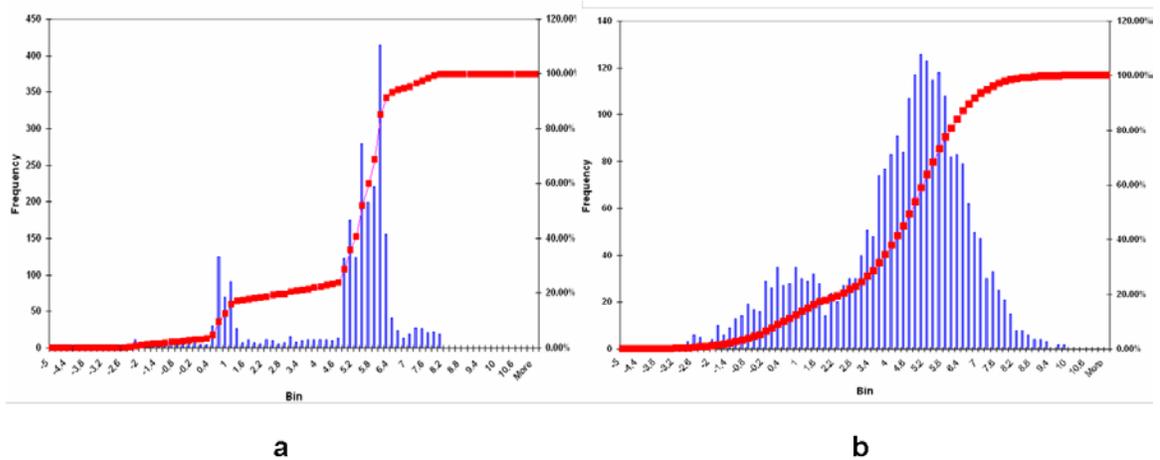


Figure 3.3 The initial log-permeability histogram on the left compared with the updated log-permeability histogram on the right for one realization.

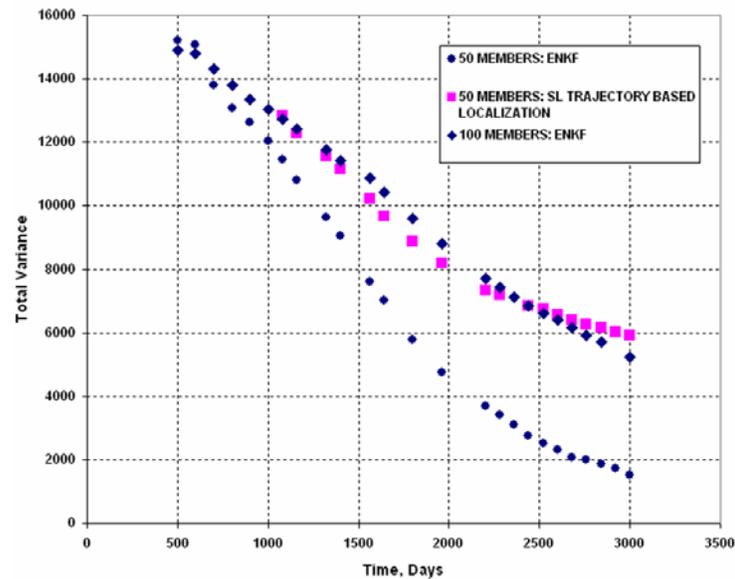


Figure 3.4 Comparison of the total variance as a function of time during the assimilation process.

To further appreciate the benefits of flow relevant covariance localization, the problem associated with the loss of variability is mitigated to a large extent as seen in

Figure 3.4 and in comparison with the standard EnKF, an equivalent accuracy is obtained with only half the ensemble members. This particular form of covariance localization is applied to field cases in Arroyo et al. (2006) and in all applications, a significant reduction in the severity of the problems associated with the EnKF is observed.

From a practical point of view, the construction of the modified cross-covariance matrix does not entail additional computational effort and consequently, forms the basis of an efficient means to enhance state and parameter estimation using the EnKF. Furthermore, for modest ensemble sizes, the method is able to better preserve geological continuities and the computational cost of localizing the cross-covariance is significantly less compared to the cost of implementing the EnKF with a large enough ensemble to achieve the same level of accuracy.

3.2 Streamline-Sensitivity Based Covariance Localization

When measurements become available, the EnKF updates a set of realizations to reflect the new information. The updates are derived from sample-based estimates of the cross-covariance matrix as seen in Eq. 1.15. Since the EnKF updates are sensitive to estimates of the cross-covariance, various modifications to the EnKF were introduced in section 2. Distance-dependant localization attempts to taper the cross-covariance on the basis of distance by assuming that the correlation is a function of distance. A more appealing alternative however retains physically relevant correlations while eliminating others by spatial localization through the use of streamline trajectories. In using streamline-sensitivity based covariance localization, we attempt to achieve the goal of tapering the cross-covariance calculations within the area of influence defined by streamline trajectories by the use of a non-isotropic localizing function.

Parameter sensitivities are often used in data integration using inverse modeling algorithms (Yoon et al. 1999; He et al. 2002; Cheng et al. 2004, Oyerinde 2008) to estimate influence of the spatial distribution of the unknown parameters, for e.g. porosity and permeability on the model response. Sensitivities relate changes in the production

response to variations in the parameters, thereby quantifying a linearized relationship between the model and the data. In this section, we review the basic methodology behind applying streamline-derived sensitivities for covariance localization in the EnKF. Using a test example, we demonstrate the use of this technique to obtain improved estimates of the cross-covariance. The second part of this discussion provides the mathematical details behind sensitivity computation and the subsequent section describes the application to the EnKF for the purposes of covariance localization. This is followed by application to synthetic test cases and field-scale examples.

3.2.1 Motivating Example

To illustrate the effects of finite sample size on the cross-covariance estimation and the benefits of streamline-sensitivity based covariance localization, we consider the hypothetical example introduced in section 2.1. The individual grid cell permeabilities were generated by sampling from a normal distribution of specified mean and variance and are independent and identically distributed. Essentially, they represent samples from a multi-Gaussian distribution with zero spatial correlation. The 2-D reservoir is completed in an inverted 5-spot pattern and is discretized into 50x50 grid cells. An illustration of two realizations and the corresponding streamline patterns for two wells at a certain time is shown in Figure 3.5. The color contours in the streamline plots indicate oil saturation values. Barring minor deviations due to the random nature of the underlying permeability values, the streamline patterns are analogous to a homogeneous permeability field. This is, of course, a consequence of the uncorrelated nature of the permeability distribution.

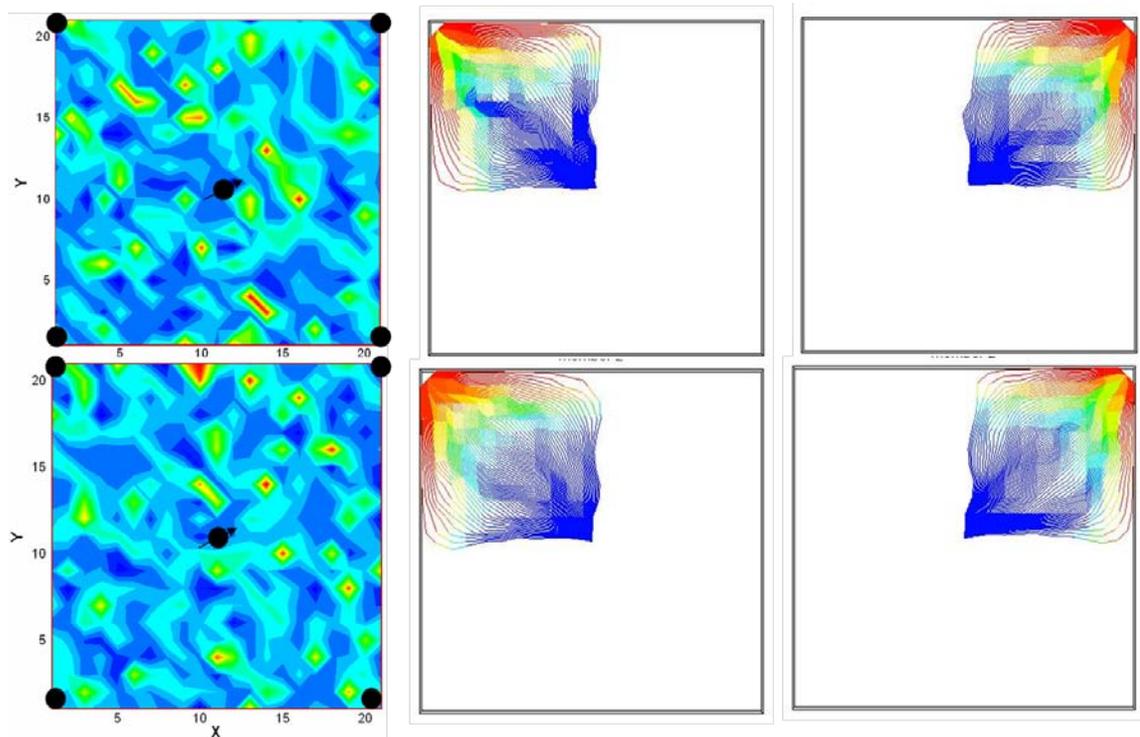


Figure 3.5 This plot shows two realizations of a completely random permeability field and the corresponding streamline patterns for two wells.

For this arrangement, a comparison of the cross-covariance between water-cut data at one well and the permeabilities at a particular time is shown in Figure 3.6. It is well known that increasing sample sizes provides better estimates of various sample-based statistical measures, for e.g. the mean and the covariance. The rationale behind increasing the ensemble size is seen in Figure 3.6. The 1000-member ensemble has a sample size roughly on the order of the model dimensions ($= 2500$) and consequently, will be closer to the true value of the cross-covariance and appears to be better suited to eliminate noisy artifacts compared to the smaller 50-member ensemble. Additionally, the relation between the well data and the permeabilities is seen to have a coherent structure with the highs located along the diagonal and the lows further away from the diagonal. This is physically intuitive but not evident with the 50-member ensemble. However, using the proposed approach for covariance localization based on streamline-derived

sensitivities, which is described in the next section, for a 50-member ensemble, the cross-covariance structure shows strong resemblance to the estimate from the larger ensemble. In contrast with the un-localized cross-covariance, the localized estimate is clearly free of noisy artifacts and spurious long-range correlations.

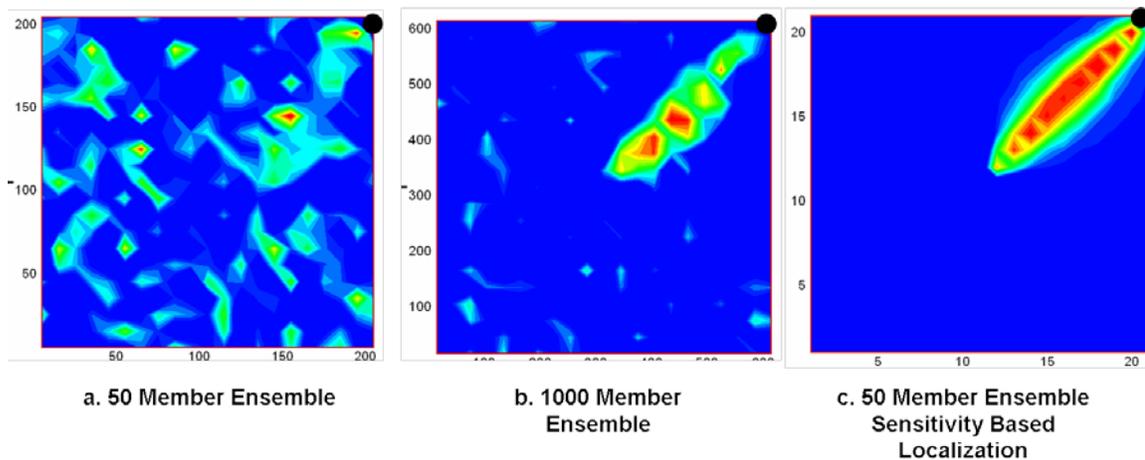


Figure 3.6 A comparison of the cross-covariance estimation between water-cut at one well (black dot) and the permeability for various ensemble sizes.

This test study serves as a motivating example to further describe the methodology of the proposed approach and to provide further examples to demonstrate its efficacy.

3.2.2 Sensitivity Computation

Our approach relies on inexpensive sensitivity calculations that can be used to modify the ensemble-derived cross-covariance estimates. Sensitivities quantify changes in production response at a well due to small perturbations in reservoir parameters. The use of sensitivities to identify regions in the reservoir where changes in the parameters will have a significant influence on model prediction enables physically consistent updates to

these key parameters to achieve a history match. While there are several approaches to compute sensitivity coefficients (Yoon et al. 1999; He et al. 2002; Cheng et al. 2004, Oyerinde 2008) streamline models are particularly well suited for sensitivity computation. By taking advantage of the analytic relationship between streamline characteristics like the time-of-flight and reservoir parameters like permeability and porosity, sensitivity estimation requires minimal computational effort and can be obtained simultaneously with the forward run of the simulator. Specifically, in our studies, we analytically derive the relationship between perturbations in model parameters and changes in production response at the wells, for example, water-cut and GOR. Previous history matching literature provides a greater insight into sensitivity computation using streamline models (Vasco et al. 1999; He et al. 2002; Cheng et al. 2004). Following these approaches, the time-of-flight sensitivities can be described analytically as simple integrals along streamlines. The time-of-flight is defined as the time it takes for a neutral tracer particle to move from one position to another location. Given the interstitial velocity, $v(\mathbf{x})$, it can be defined as integrals along the streamlines and is given by

$$\tau = \int_{\Sigma} \frac{1}{v(\mathbf{x})} dx = \int_{\Sigma} s(x) dx \dots\dots\dots 3.2$$

The time-of-flight sensitivity with respect to permeability can be expressed as

$$\frac{\partial \tau}{\partial k(\mathbf{x})} = \int_{\Sigma} \frac{\partial s(\mathbf{x})}{\partial k(\mathbf{x})} dx = - \int_{\Sigma} \frac{s(\mathbf{x})}{k(\mathbf{x})} dx \dots\dots\dots 3.3$$

where the integrals are evaluated along streamlines and the ‘slowness’ which is the reciprocal of interstitial velocity is defined as

$$s(\mathbf{x}) = \frac{\phi(\mathbf{x})}{\lambda_{rt} k(\mathbf{x}) |\nabla P(\mathbf{x})|} \dots\dots\dots 3.4$$

Streamline simulators estimate these quantities simultaneously during the forward run. For finite-difference simulators, the total inter-block fluid fluxes are used to trace streamlines and compute the time-of-flight and slowness (Jimenez et al. 2007).

Water-Cut Sensitivity. The sensitivity of the arrival time t of water cut to model parameter m , for e.g. permeability, can be analytically defined using the compressible formulation for streamlines (Cheng et al. 2007) as

$$\frac{\partial t}{\partial m} = \frac{\frac{\partial \tau}{\partial m} \frac{\partial}{\partial \tau} \left(\frac{S_w}{B_w} \right)}{\frac{\partial}{\partial \tau} \left(\frac{f_w}{B_w} \right) + \frac{f_w}{B_w} \frac{c}{\phi}} \dots\dots\dots 3.5$$

where $\nabla \cdot u_t = c$ representing the divergence of the flux can be considered as a source term due to compressibility effects and is computed along the streamline at each grid block. Using the chain-rule of differentiation, we can then relate the travel-time sensitivity to the sensitivity of water-cut at a particular time for each update step.

Gas-Oil Ratio Sensitivity. The analytic approach using the compressible formulation for streamlines (Cheng et al. 2007) leads to the expression for the sensitivity of the arrival time t of a particular value of gas-oil ratio to model parameter m as defined below.

$$\frac{\partial t}{\partial m} = \frac{\frac{\partial \tau}{\partial m} \frac{\partial}{\partial \tau} \left(\frac{S_g}{B_g} + \frac{S_o R_s}{B_o} \right)}{\frac{\partial}{\partial \tau} \left(\frac{f_g}{B_g} + \frac{f_o R_s}{B_o} \right) + \left(\frac{f_g}{B_g} + \frac{f_o R_s}{B_o} \right) \frac{c}{\phi}} \dots\dots\dots 3.6$$

where \mathbf{c} , the divergence of total flux, retains the same physical significance as previously described for water-cut sensitivities. The chain-rule for differentiation is again used to define the sensitivity of gas-oil ratio to model parameter \mathbf{m} .

3.2.3 Construction of the Localizing Function

The localizing function, $\boldsymbol{\rho}$, in the context of trajectory based localization was defined to include grid cells selected by streamlines. The area of influence for the entire ensemble was defined by stacking the selected areas of each realization. Further modifications to this selected area can be achieved by selecting only those cells common to a certain minimum number of model realizations.

Streamline-sensitivity based localization attempts to capture the strength of the correlation between model variables and the data by conditioning the cross-covariance with an appropriate tapering function or localizing function within this area of influence. For the purpose of quantifying the relative strengths of the correlation, sensitivities are particularly appealing since they directly provide a relation between the model variables and production data. For the purpose of covariance localization using sensitivities in the EnKF, the procedure is identical to describing the area of influence using streamline paths, with the additional step of weighting with a localizing function using streamline-derived sensitivities. As previously outlined, these sensitivities to model parameters are computed along streamlines quite efficiently in a single forward run of the simulator. The procedure of estimating $\boldsymbol{\rho}$ starts with stacking the sensitivities from all realizations and then normalizing these derived quantities individually for each observation on a scale of 0 to 1. An appropriate cut-off may be chosen to ensure proper scaling of these

normalized sensitivities. In all cases studied here, the sensitivity to permeability variation is used to update the parameter field (the permeability).

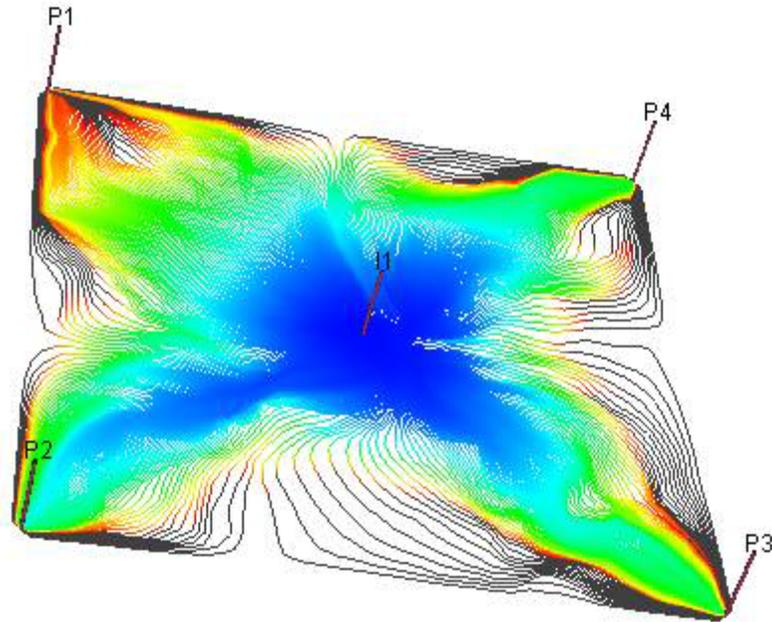


Figure 3.7 The streamline pattern and time-of-flight contours for a inverted 5-spot pattern in a synthetic permeability field.

Since ρ is a matrix of size $N_m \times N_d$ where N_m is the number of model variables and N_d is the number of observation, we can write each element of ρ as

$$\rho_{i,j} = \left| \frac{S_{i,j}}{\max_j(S_{i,j})} \right| \dots\dots\dots 3.7$$

where \mathbf{S} refers to the matrix formed by summing the sensitivities above a threshold over all the realizations. In this manner, the localizing function ρ acting on the sample cross-covariance achieves the dual objective of retaining physically relevant correlations and

weighting them in relation to their relative significance. The demarcation of an area of influence is inherent in the definition of the sensitivities which are defined along streamlines. Figure 3.7 shows the streamline pattern at a particular time while Figure 3.8 highlights the grid blocks corresponding to each measurement and the normalized sensitivities.

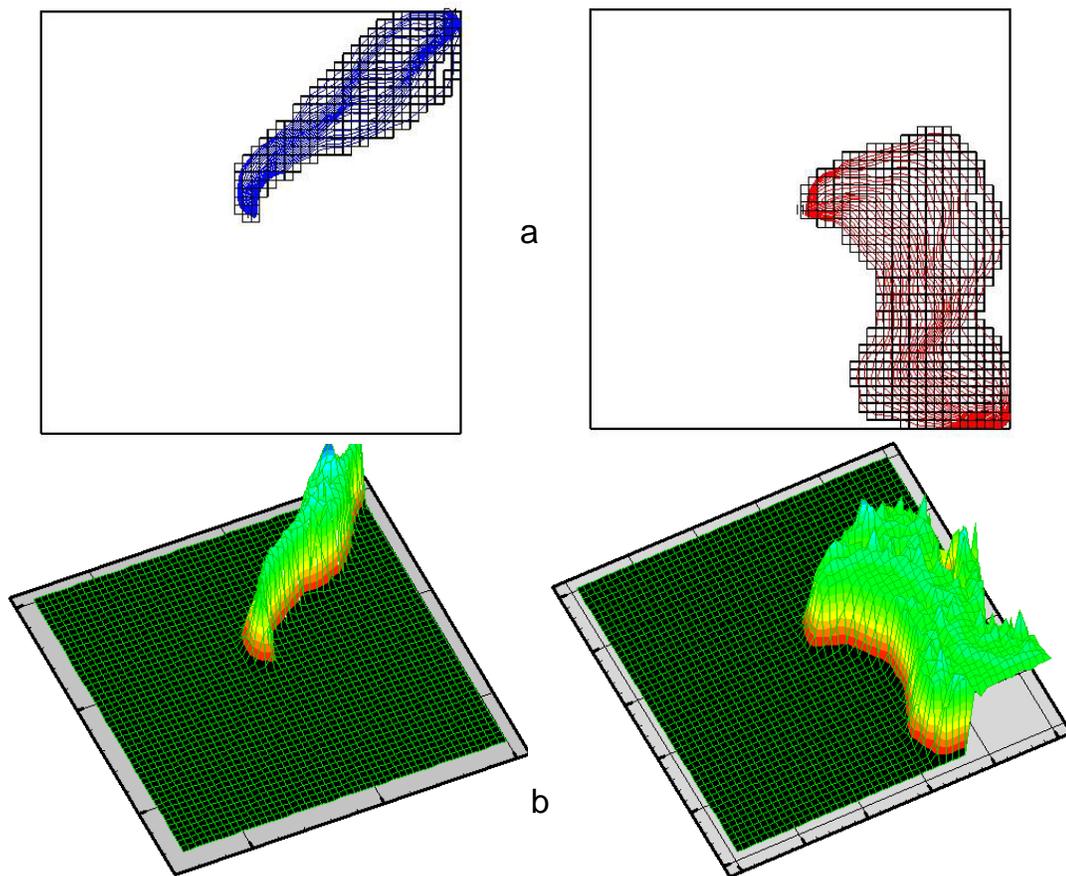


Figure 3.8 The streamline trajectory and the normalized streamline-derived sensitivities for 2 selected wells for the field shown in Figure 3.7.

The Schur product of these two matrices (term-by-term multiplication) reduces spurious noise in the covariances and the resulting tendency toward introducing

unrealistically large increments in distant grid locations. We will show that this technique will amplify the signal-to-noise ratio where the signal is the cross-covariance to be estimated. As specified in section 3.1, the EnKF equation becomes

$$\tilde{\mathbf{y}} = \mathbf{y}_{prior} + \boldsymbol{\rho} \circ \mathbf{C}_y \mathbf{H}^T (\mathbf{H} \mathbf{C}_y \mathbf{H}^T + \mathbf{C}_D)^{-1} (\mathbf{d}_{obs} - \mathbf{H} \mathbf{y}_{prior}) \dots\dots\dots 3.8$$

In this framework, the computation of the matrix $\boldsymbol{\rho}$ is performed at each update step to reflect changing field conditions or updated well configurations. The computation is not limited to streamline models and is easily extended to the finite-difference simulation scheme by using the underlying velocity field to trace streamlines (Jimenez et al. 2007) and computing the sensitivities along these trajectories.

3.3 Application: Synthetic Example

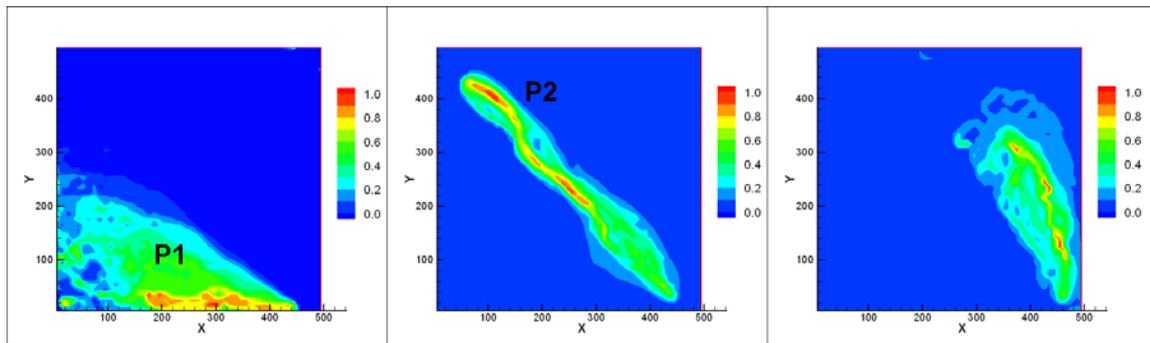


Figure 3.9 The normalized sensitivities computed for each well from an ensemble of 50 realizations at $t = 1000$ days. The significant areas impacting each well’s water-cut data are highlighted.

To illustrate the methodology and the benefits of using streamline-derived sensitivities, we test our proposed approach with the synthetic example used to demonstrate distance-dependant and streamline-trajectory based covariance localization. The cross-covariance

localizing function for each of the wells is plotted in Figure 3.9 at $t = 1000$ days for an ensemble size of 50 members.

The matches to the production data at each well are shown in Figure 3.10. Satisfactory matches to the observed production data are observed without any apparent difficulty.

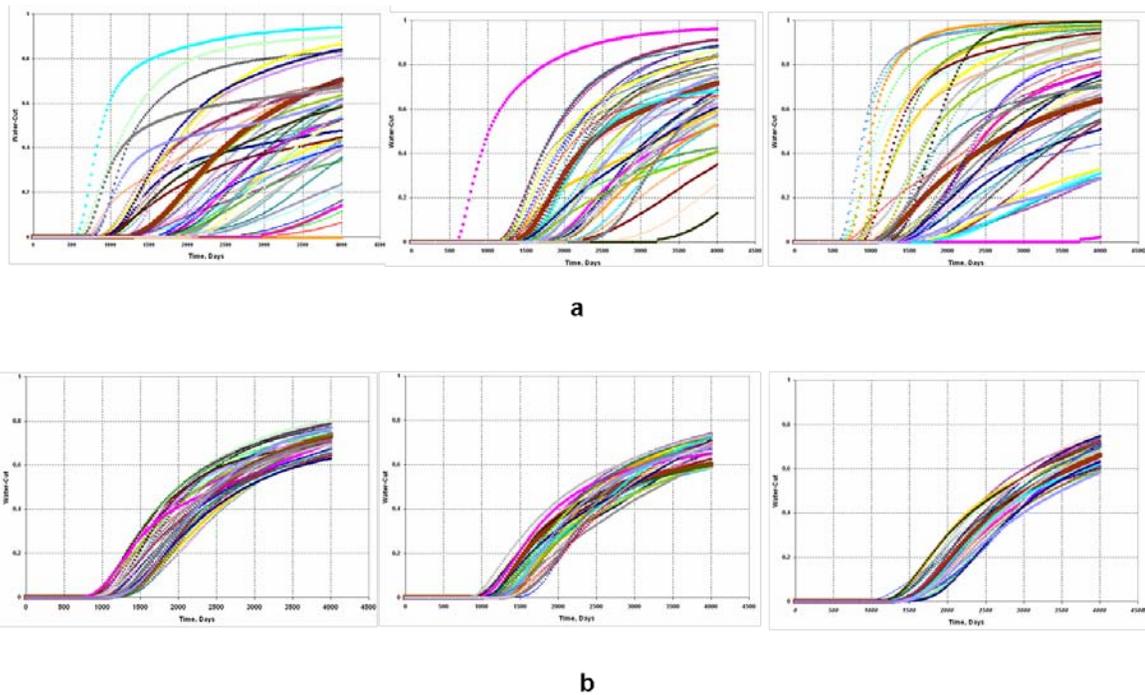


Figure 3.10 (a) is the initial spread in the predicted water-cut from the ensemble. (b) is the match to water-cut data achieved using EnKF with streamline-derived sensitivities for covariance localization.

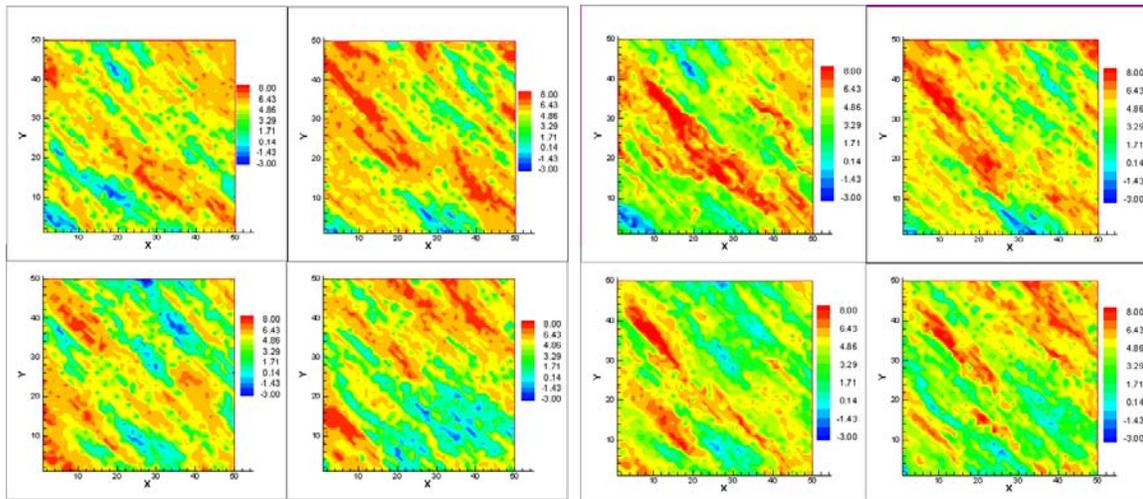


Figure 3.11 The updated permeability fields for 4 realizations on the right compared to the initial members on the left. Notice that changes to the initial models are minimal.

An indicator of the benefits achieved through flow relevant conditioning of the cross-covariance matrix is given in Figure 3.11 where we compare the updated realizations to the initial permeability fields. By using streamline-derived information, the changes are kept minimal and specifically targeted at areas that can be resolved by the production data. The tendency of the EnKF to create variance-deficient members is also mitigated leading to more meaningful results from uncertainty analyses using simulation forecasts. This benefit can also be seen in Figure 3.12 also, which compares total variance versus assimilation time for the proposed approach in comparison with EnKF using different ensemble sizes. Clearly, the proposed approach outperforms the EnKF with a much larger ensemble in terms of preserving variability.

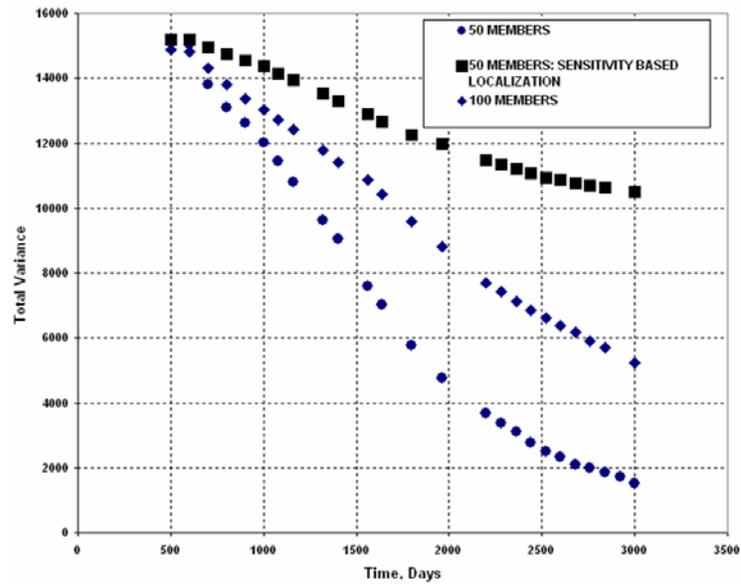


Figure 3.12 The total variance amongst the members as a function of time during the assimilation process. Flow relevant conditioning of the cross-covariance clearly provides substantial benefits by not artificially reducing the variability.

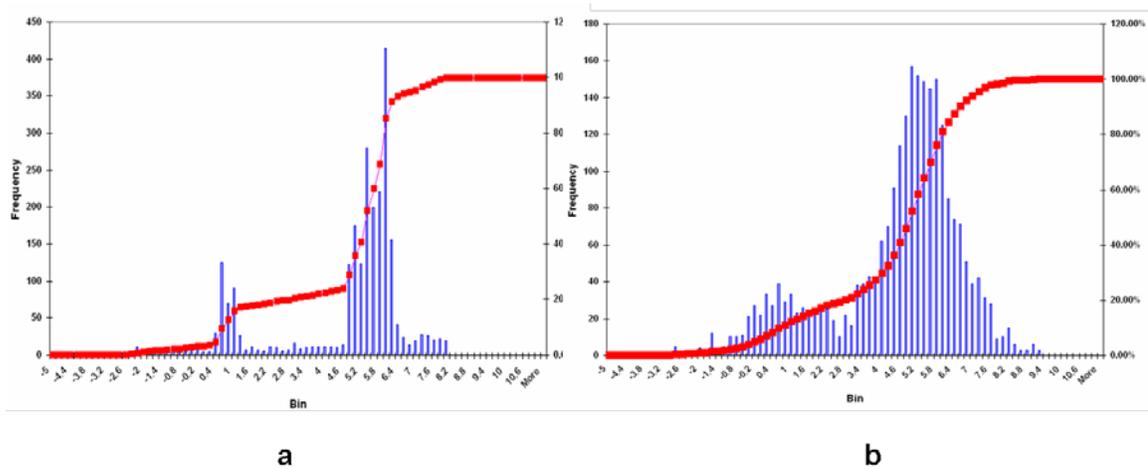


Figure 3.13 (b) is the histogram of log-permeability for one updated realization compared to (a) which is the initial histogram. We maintain the multi-modal nature of the histograms without overshooting.

Without covariance localization, the histograms of the posterior models acquire strong Gaussian characteristics. A comparison of one pre and post assimilation histogram of the log-permeability is shown in Figure 3.13 and the method tends to preserve prior model statistics. A further examination of the results indicates that parameter overshooting is largely eliminated. There are no inconsistent localized zones of high and low permeability in Figure 3.10 and this is also reflected in the respective histograms.

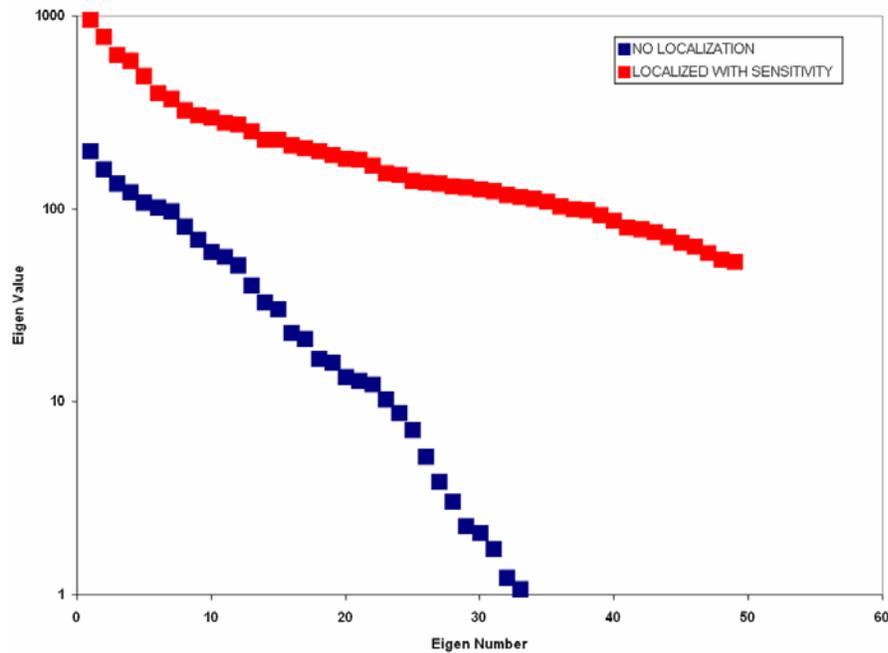


Figure 3.14 Eigen-spectrum of the covariance matrix using sensitivity based localization (in red) and without any form of localization (in blue).

Variance deficiency in the EnKF can also be addressed by examining the eigen-spectrum of the covariance matrix of the updated members. Recall that in Eq. 1.3, a sample-based estimate of the covariance estimate can be obtained whenever necessary. Figure 3.14 shows the eigen-spectrum for the permeability covariance matrix based on

this calculation. A comparison is made between the proposed approach of localization and the standard EnKF. The eigen-spectrum for the un-localized version of the EnKF shows a steep drop and consequently, there is insufficient projection of the variance in the trailing eigen-directions. The degraded covariance structure associated with these lower energy trailing eigen-values implies a smaller effective ensemble size. This is reinforced by the observation in section 1 in Figure 1.3, where the effective ensemble size is successively reduced till the model realizations appear similar.

3.4 Field Application: The Goldsmith Field Case

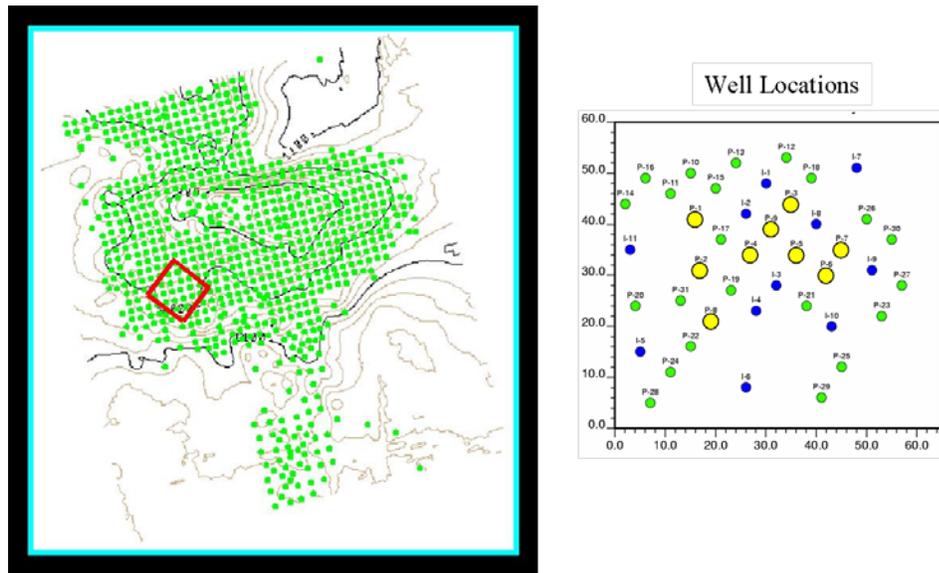


Figure 3.15 The location of the Goldsmith pilot area in the San Andreas Unit in West Texas. The location of the producers (in yellow) and the injectors (in blue) is on the right.

In this section, we validate the efficacy of streamline-derived-sensitivity based conditioning of the EnKF to a 2 phase field example, the Goldsmith CO₂ pilot project study (He et al. 2002; Cheng et al. 2005). We used an ensemble size of 50 realizations of

porosity and permeability fields conditioned to well data and secondary seismic attributes. The pilot area comprises 9 inverted 5-spot patterns covering around 320 acres and the average thickness of the formation is 100 ft.

Significant water-cut data for 20 years of production prior to the start of the CO₂ flood is available at 9 production wells and these were used to condition the permeability fields. By virtue of the location of these wells in the center of the field, it is expected that most of the changes to the permeability distribution should be concentrated in this region. Figure 3.15 is an areal plot of the location of the Goldsmith CO₂ pilot study within the extended study area showing the location of the producers (in yellow) and the injectors (in blue).

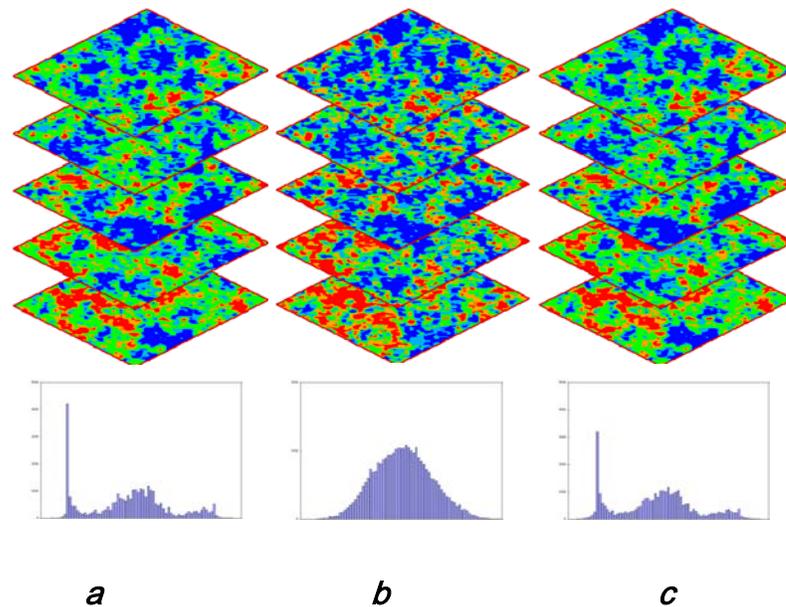


Figure 3.16 (a) is the initial permeability field showing the multi-modal permeability histogram. (b) illustrates the transformation to a normal histogram with the standard EnKF, while (c) shows the results from the streamline-sensitivity based covariance localization.

Figure 3.16a depicts the permeability distribution for one of the initial realizations. The significant peak at lower permeability values reflects a fairly large proportion of poor reservoir quality rock.

Using the standard EnKF, ensemble realizations are conditioned to water-cut history and Figure 3.16 shows one of the ensemble realizations after a sequence of updates. As demonstrated in the previous application of the standard EnKF to a hypothetical 2-d example with non-Gaussian permeability fields, the reconstructed log-permeability histogram tends to acquire strongly Gaussian features. Although the producers are completed in the center of the field, the largely indiscriminate updates contribute to the loss of structure in the permeability field.

However, a qualitative inspection of the updated permeability field in Figure 3.16c using localization derived from streamline-derived sensitivities shows permeability variations that are consistent with the initial model. This is the preferred mode of comparison to check the validity of the updates in the absence of a reference or ‘truth’ model. The preferential and targeted changes are reflected in the geologic consistency between the pre and post assimilation models. Additionally, the log-permeability histogram of the updated model clearly indicates that the relative proportion of various reservoir rock types is preserved. This is of considerable significance to obtain meaningful results from production forecast analysis.

The ability of the proposed approach to satisfactorily match production history for 3 selected wells is illustrated in Figure 3.17. The prediction envelope from the updated ensemble is within the specified noise tolerance.

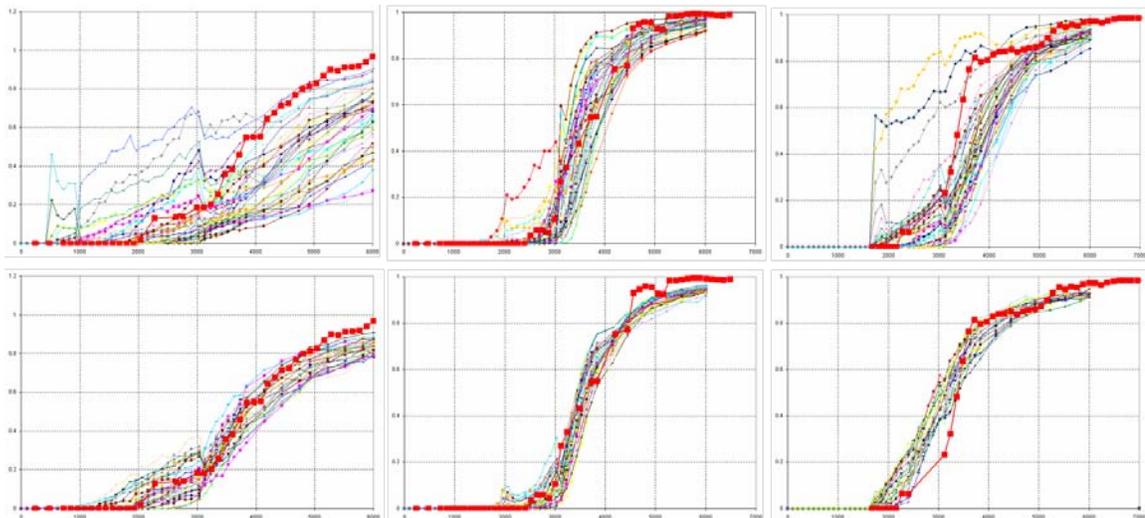


Figure 3.17 Matches to production history at 3 selected wells in the bottom row.
The initial predictions are plotted in the top row.

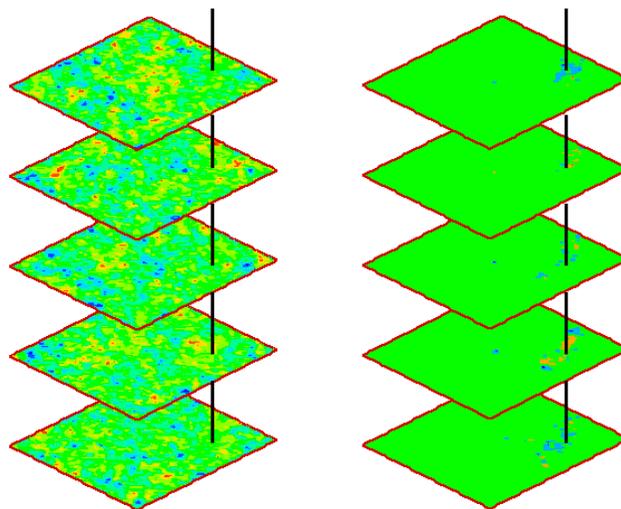


Figure 3.18 On the left is the cross-covariance between water-cut data at the well (vertical line) and permeabilities in the field. Using streamline-derived sensitivities, we condition the cross-covariance matrix and on the right is the plot of the localized cross-covariance.

Figure 3.18 plots the cross-covariance at a particular time between water-cut at one of the wells and the permeabilities in the field. The standard EnKF shows large values even for distant locations where it is reasonable to expect that there is no correlation between the permeability and well performance because of the well configuration. However, with streamline-derived sensitivity based localization, distant data points are not included in the analysis and the support is more localized in the region around the well.

3.5 Section Summary

The sampling-based errors arising due to the use of a modest set of model replicates has motivated the development of modifications to the EnKF to address the issue for large-scale problems. Among these approaches is covariance localization that relies on specifying better estimates of correlations between variables at different locations. Streamline-derived information forms the basis of a physically intuitive, natural and computationally tractable scheme for covariance localization. In this section, we present two approaches for covariance localization formulated using streamlines.

Streamline-trajectory based localization achieves spatial localization in the cross-covariance matrix by defining regions within the reservoir that directly influence well responses. Sensitivity based localization, on the other hand, defines a taper function within the previously described area of influence to eliminate spurious sampling errors in the cross-covariance estimates and to dampen the relevant correlations based on the linearized relation between the data and the model variables. In all test studies, significant time savings and reduction in computational effort is achieved.

However, streamline-trajectory based localization is better suited for bottom-hole pressure (BHP) integration because of the difficulty in estimating the corresponding sensitivities in a computationally efficient manner. Other studies have implemented successful localization schemes for assimilation of BHP data based on the concept of pressure time-of-flight.

4. A HIERARCHICAL MULTI-SCALE EnKF

Reservoir characterization problems tend to be large with the number of unknowns to be estimated often orders of magnitude larger than the number of observations (Tarantola 2004). The inherent ill-posed nature of these problems is the source of difficulties associated with non-uniqueness and stability of the solutions. This is further compounded by the non-linear nature of multiphase flow problems which leads to the existence of multiple local minima in the objective function. It is widely recognized that multi-scale data integration can reduce the severity of these problems and can assist in the search for a global solution (Yoon et al. 1999).

In this section, we examine the performance of a hierarchical multi-scale EnKF. One of the motivating factors is the ease with which the EnKF can assimilate diverse data types. Secondly, measurements, either static or dynamic, are often associated with different length-scales. Derived static measurements like permeability estimates from cores have a resolution of a few inches while seismic data has a significantly larger areal coverage with reduced precision. The proposed approach is applied to the problem of interpreting Partitioning Inter-well Tracer Tests (PITT) specifically to identify the distribution and location of residual un-swept oil.

4.1 Introduction

In fine-scale inverse modeling approaches, difficulties related to stability, existence and non-uniqueness of the solution are addressed through the use of regularization functions (Yoon et al. 1999; He et al. 2002; Tarantola 2004; Cheng et al. 2004; Oyerinde 2008). The forms and the relative strengths of the terms that constitute the regularization function are additional sources that contribute to the non-uniqueness of the solution. The search for a global minimum of a suitably formulated objective function may be confounded by the existence of multiple local minima especially for highly non-linear

multiphase flow phenomena. Yoon et al. (1999) address the issues of non-uniqueness and stability and enhance the ability of the inversion process to reach a global minimum through the use of a multi-scale inverse modeling approach. Their formulation proceeds by integrating data at the coarsest level and sequentially assimilating observations at successively finer scales.

The multi-scale estimation approach described in this section exploits the ability of the EnKF to efficiently and recursively integrate data associated with different scales. By constraining the high resolution variations in model properties to coarse-scale information, the estimates of the unknowns are better approximated by reducing the non-linearity of the underlying problem. Akella et al. (2008) derive the basic methodology describing the approach. They use a secondary constraint derived from coarse-scale estimates of the model variables like permeability or saturation. In our approach, a deterministic inverse modeling algorithm provides coarse-scale spatial distribution of the reservoir unknowns.

4.2 Methodology

The basic equations describing the EnKF are provided in section 1. An extension to the EnKF to include smoothness constraints was introduced in section 2. Our algorithm proceeds along the same lines; however, instead of imposing smoothness on the solution by explicitly specifying an ad-hoc regularization term, we use observations on a coarse grid to enhance estimates of the model variables. The procedure is implemented as follows:

- Update the high-resolution ensemble of models using data at the fine-scale.
- An inverse modeling approach is separately implemented to resolve coarse-scale features of the ensemble mean using previously obtained measurements.
- The low resolution information from the ensemble mean modifies the micro-scale variations in reservoir properties to preserve the large-scale structure and continuity.
- Subsequent updates are performed in the same manner as outlined in the

preceding steps.

In data assimilation literature, the procedure of updating the ensemble mean using inversion is not new. Zhang et al. (2007) use a feedback loop that provides ensemble-based estimates of the model covariance to a 4DVAR algorithm (inverse algorithm) operating on the ensemble mean. The coupled approach is shown to aid in preventing filter divergence and is seen to be superior to the standard form of the EnKF. However, the basis of our approach is to superimpose coarse-scale structure of the ensemble mean on the fine-scale variations of the individual ensemble realizations. The coupling does not entail explicit covariance information exchange between the EnKF and the inverse approach. However, by utilizing low resolution variations in reservoir properties, the formulation implicitly provides tuning of the sample-based model covariance.

Eq 1.15 is the basis of the first step in the multi-scale EnKF approach proposed here.

$$\tilde{\mathbf{y}} = \mathbf{y}_{prior} + \mathbf{C}_y \mathbf{H}^T (\mathbf{H} \mathbf{C}_y \mathbf{H}^T + \mathbf{C}_D)^{-1} (\mathbf{d}_{obs} - \mathbf{H} \mathbf{y}_{prior}) \quad \dots\dots\dots 4.1$$

The ensemble mean is specified by

$$\mathbf{y}_{mean} = \sum_{i=1}^{N_e} \mathbf{y}_{prior,i} \quad \dots\dots\dots 4.2$$

However for the purposes of parameter estimation, we select the vector of model parameters, \mathbf{m}_{mean} from the state vector. A deterministic formulation is used to efficiently integrate production data via the minimization of an augmented misfit function to find the elements of the change vector, $\delta \mathbf{m}$. Further details of the procedure can be found elsewhere (Yoon et al. 1999; He et al. 2002; Cheng et al. 2004; Tarantola 2004; Cheng et al. 2005; Oyerinde 2008)

$$F(\delta\mathbf{m}) = \|\delta\mathbf{d} - \mathbf{S}\delta\mathbf{m}\| + \beta_1\|\delta\mathbf{m}\| + \beta_2\|\mathbf{L}\delta\mathbf{m}\| \quad \dots\dots\dots 4.3$$

To obtain macro-scale estimates of the underlying structure, the inversion may be carried out on an upscaled version of \mathbf{m}_{mean} or the results from a fine-scale inversion may be reparametrized to lower resolution to provide a $\mathbf{m}_{\text{mean, posterior}}$. It is commonly recognized that the former approach is computationally more efficient because the model simulation is carried out on a coarse grid with fewer parameters and may converge faster (Yoon et al. 1999).

A secondary assimilation sequence then updates the finer resolution model by minimizing a misfit function described on coarse-scales.

$$\hat{\mathbf{y}} = \tilde{\mathbf{y}} + \mathbf{C}_{y,c} \mathbf{U}^T (\mathbf{U} \mathbf{C}_{y,c} \mathbf{U}^T + \mathbf{R})^{-1} (\mathbf{y}_{\text{mean},c} - \tilde{\mathbf{y}}_c) \quad \dots\dots\dots 4.4$$

where the subscript ‘c’ refers to the coarse grid model. The precision of the coarse-scale observations is contained in the covariance matrix \mathbf{R} . The upscaling operator \mathbf{U} is the link between the low and high resolution models.

$$\tilde{\mathbf{y}}_c = \mathbf{U}\tilde{\mathbf{y}} \quad \dots\dots\dots 4.5$$

The upscaling operator is general and the same procedure can be implemented to assimilate data at different length scales by the appropriate choice of \mathbf{U} . This hierarchical approach to multi-scale data assimilation is formulated based on the independence of the data at different levels of resolution. Although the condition of strict independence may not be exactly honored, it is recognized that this is only an approximation.

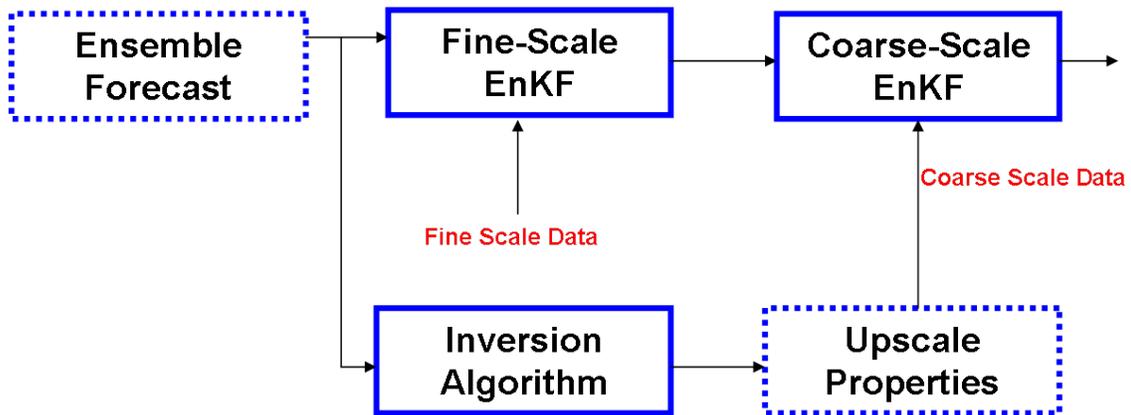


Figure 4.1 The multi-scale EnKF procedure at any given assimilation step.

Figure 4.1 gives the details of the procedure. The application of the proposed methodology will be provided in a later section in this section. First, we discuss composite tracer tests and their role in reservoir characterization.

4.3 Composite Tracers and Partitioning Interwell Tracer Tests (PITT)

Enhanced Oil Recovery (EOR) methods are often implemented to recover bypassed oil following primary and secondary depletion. Critical inputs to a successful EOR technique include accurate spatial distribution of target oil and a reliable description of preferential flow paths in the reservoir. Over the years, advances in the design and implementation of tracer tests have increased our confidence in their potential to provide detailed reservoir description.

During tracer tests, a suite of tracers are injected into the subsurface and are recovered at the observation wells. These tracers can be broadly categorized based on application into conservative and partitioning tracers. Conservative or aqueous tracers move with the water phase and do not interact with other phases in the reservoir. On the other hand, partitioning tracers are soluble in the water phase as well as the oil or gas phases. Consequently, the conservative tracer response is primarily sensitive to the underlying reservoir heterogeneities while the partitioning tracer profile is affected by both the reservoir fluid flow paths and the distribution of the phases it comes in contact

with.

The partitioning is driven by the slow diffusion of tracer molecules from the tracer slug in the injected phase (typically, water) into the oil or gas phases. After the passage of the tracer slug, the concentration gradient reversal causes tracer molecules to slowly diffuse back into the injected phase. It is this phenomenon that contributes to a chromatographic delay in the observed partitioning tracer response (Oyerinde 2004). Figure 4.2 shows a typical composite tracer response with the associated chromatographic delay. When the tracer is sampled at multiple areally and vertically distributed locations, the resulting tracer profiles can be used to infer target oil saturations and provides a reliable means for detailed 3-D reservoir characterization.

Previous efforts towards estimating average in-situ volumes and oil saturations in the reservoir relied on analytic tools like the method of moments approach. The formulation of the approach is rigorous and these analytic tools are simple and easy to implement only bulk average volumes can be estimated. The reformulation of the problem in a streamline-based inverse modeling framework has led to an increased interest and enhanced capabilities to measure reservoir heterogeneity and dominant flow paths in the reservoir in addition to being able to resolve the spatial distribution of the un-swept oil (Yoon et al. 1999; Oyerinde 2004). The two-step procedure determines the spatial distribution of subsurface permeability and then utilizes the chromatographic delay between the tracer responses to indirectly infer fluid saturations. However, in the presence of mobile non-aqueous phase saturations, interpretation of the results becomes considerably more complicated because of the dependence of the retardation on the individual phase saturations, the partitioning coefficients and the local fractional flow profile.

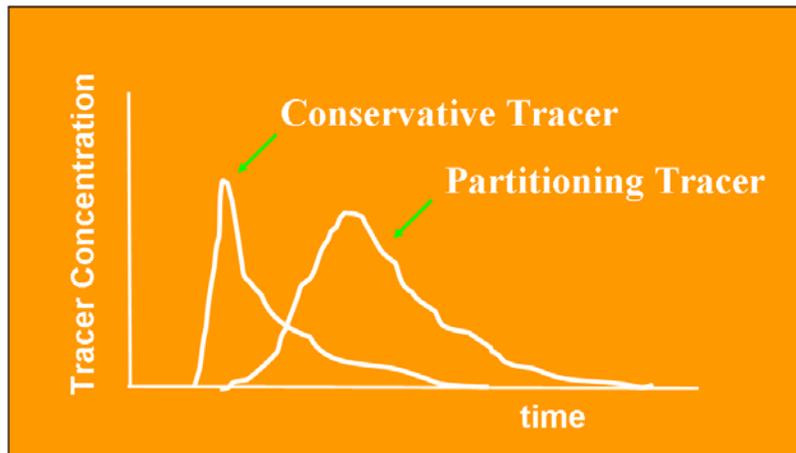


Figure 4.2 Typical composite tracer responses showing the delay of the partitioning tracer (Oyerinde 2004).

In the context of data assimilation using ensemble-based filtering, Valestrand et al. (2008) utilize tracer responses to provide improved estimates of reservoir model parameters. Using a synthetic test example, they demonstrate better convergence characteristics of the EnKF when simultaneously assimilating water-cut and tracer information. When water-cut measurements alone are assimilated, the estimates of the reservoir parameters are less representative of the truth.

In the proposed approach using the multi-scale EnKF, the goal is to use the additional information from tracer responses to provide better resolution of subsurface rock properties and spatial distribution of phase saturations. The coupling of the inverse modeling approach to provide valuable low resolution information to the EnKF estimates is the basis of the approach. The algorithm is general and can account for the effects of mobile non-aqueous phase saturations.

4.4 Application: Synthetic 2-D Example

A 2-D test reservoir is used to demonstrate the efficacy and feasibility of the proposed approach. The reservoir is completed in an inverted 5-spot pattern and the domain is discretized into 24x24 grid cells. The reservoir is produced for a period of 1500 days

with the tracer slugs injected 500 days into the waterflood. A suite of 100 members conditioned to well and geostatistical information (Deutsche and Journel 1992) constitutes the initial ensemble. The partitioning coefficient, which is a measure of the equilibrium concentration in the oil to that in the water phase, is 0.6. Figure 4.3 shows the individual tracer elution profiles with the reference permeability field.

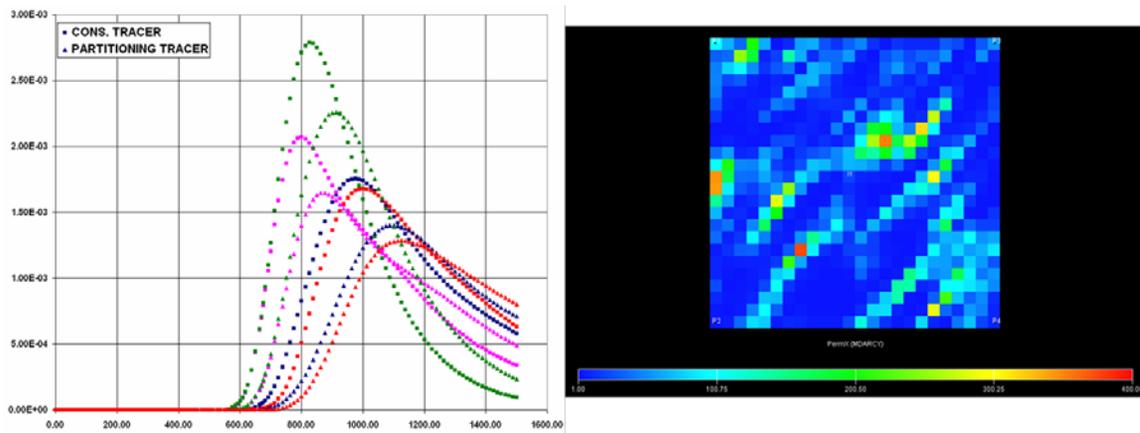


Figure 4.3 The tracer profiles at each of the 4 producing wells showing the chromatographic delay of the partitioning tracer. The reference field is on the right.

The measurements at each of the 4 producing wells include water-cut data and tracer responses and the reservoir state vector comprising permeability, phase saturations and fluid pressures is conditioned to the available data using the EnKF. The updated permeability fields are used in a production forecast from initial time to assess the quality of the matches. The match to water-cut data at each of the 4 wells is shown in Figure 4.4 and the tracer data matches at two selected wells are illustrated in Figure 4.5.

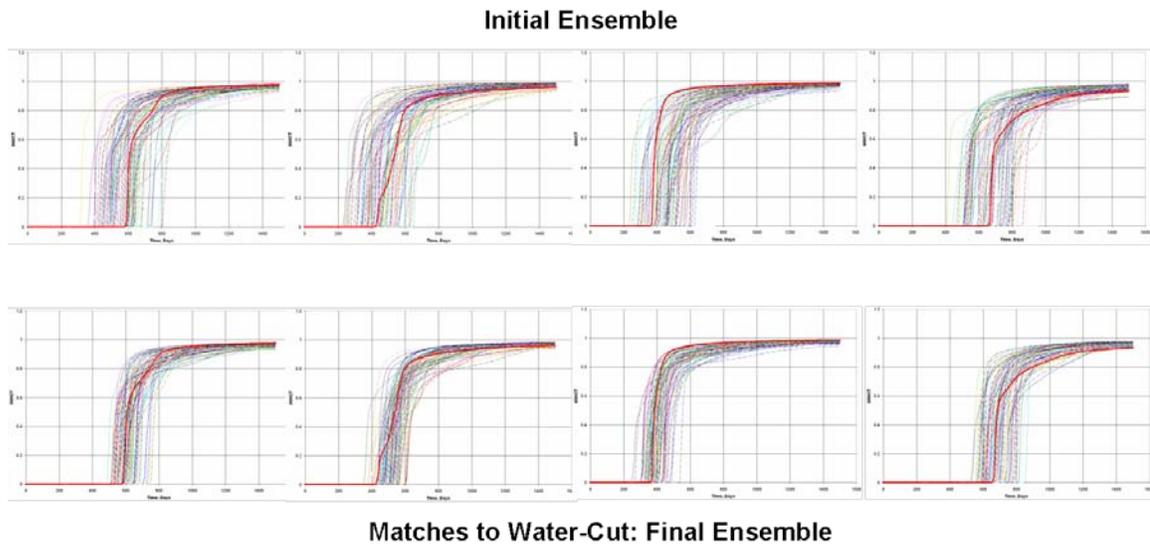


Figure 4.4 The top row shows the initial water-cut predictions and the bottom row shows the matches to the data from the updated realizations.

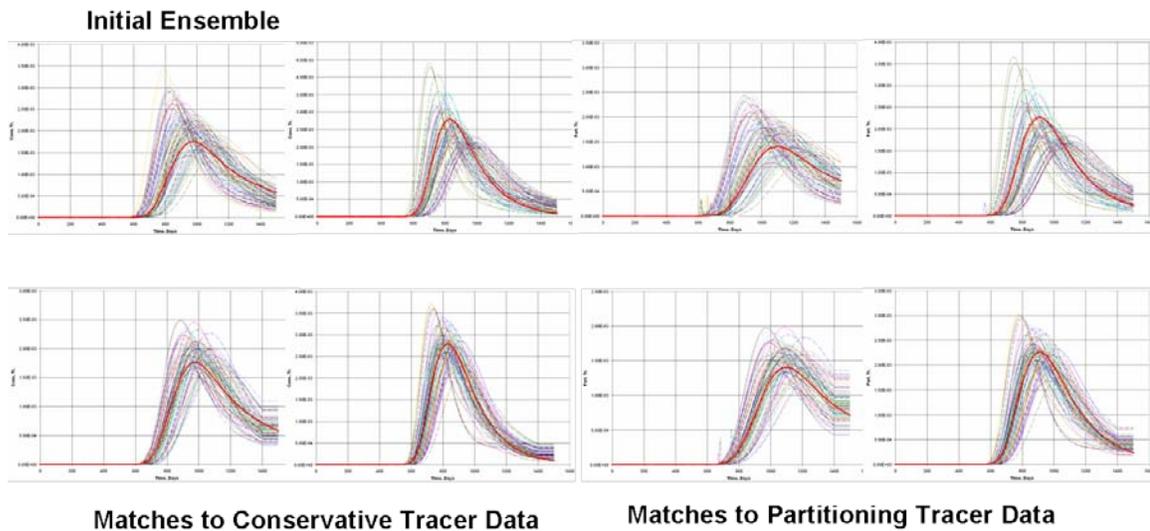


Figure 4.5 Plot of composite tracer responses from the initial ensemble for two selected wells in the top row and the corresponding matches in the bottom row.

From the above figures, it is evident that the EnKF can match production history

satisfactorily. However, the EnKF cannot capture key preferential fluid migration pathways as seen in Figure 4.6.

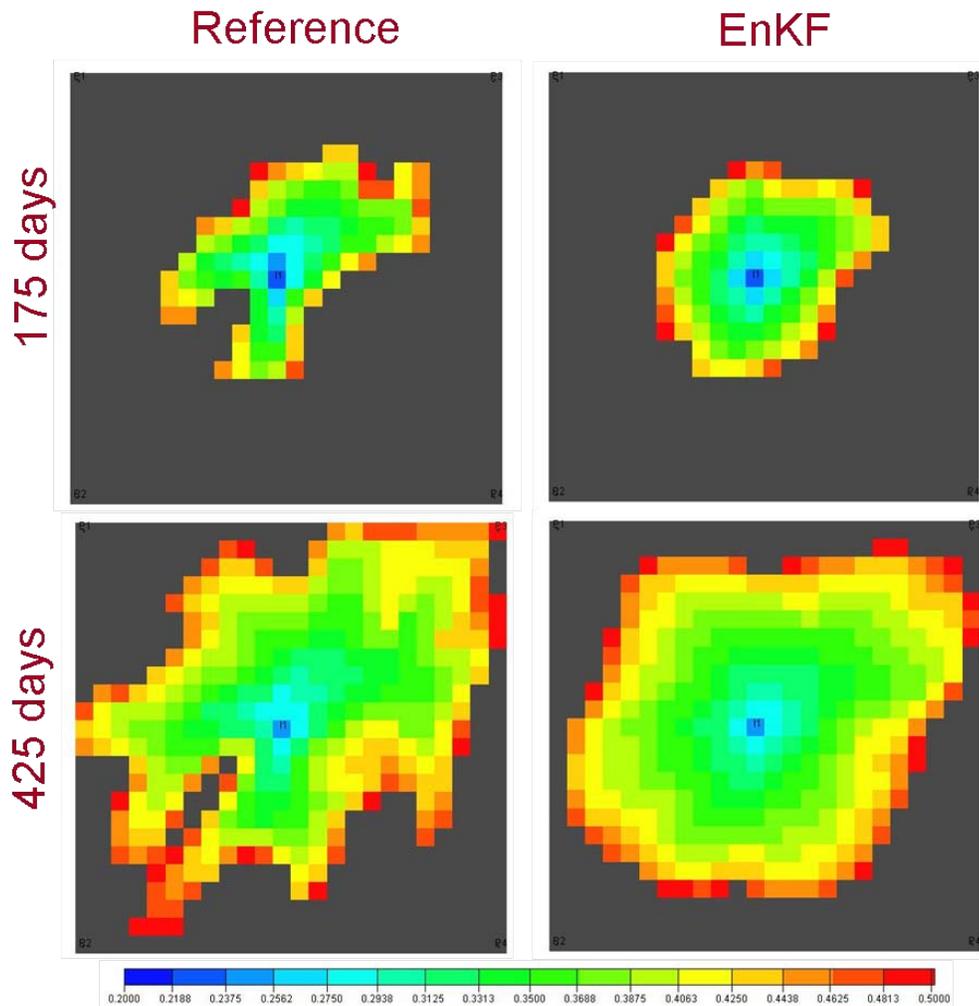


Figure 4.6 Comparison of the saturation contours of the reference field on the left with the mean of the ensemble on the right at two different times.

The ensemble estimates are seen to be smooth and lack the features of the reference model. The distribution of the bypassed oil saturation is not satisfactorily captured by simply assimilating well data using the EnKF.

We then apply the proposed methodology to the above described 2-D example. The objective is to capture key large-scale features of the spatial distribution of target oil. The approach proceeds as follows: First, we assimilate the available well production data using the EnKF. At a time of 800 days, the permeability distribution of the ensemble mean is updated using a streamline-based inversion approach (Yoon et al 1999; He et al. 2002; Cheng et al. 2004; Cheng et al. 2005; Cheng et al. 2007). Coarse-scale fluid saturation distributions are then obtained from a simple forward run of a numerical reservoir simulator and then coarsened to a 6x6 grid using volumetric averaging. The resulting phase saturation serves as a constraint in the second stage of the multi-scale EnKF as shown in Eq. 4.4. The ensemble of derived coarse-scale saturations is conditioned to this observation with a pre-specified precision. Following this step, the assimilation process proceeds with the standard form of the EnKF and at 1300 days, a second coarse-scale measurement is provided as an additional measurement.

The coarse-scale measurement at 800 days is shown in Figure 4.7.

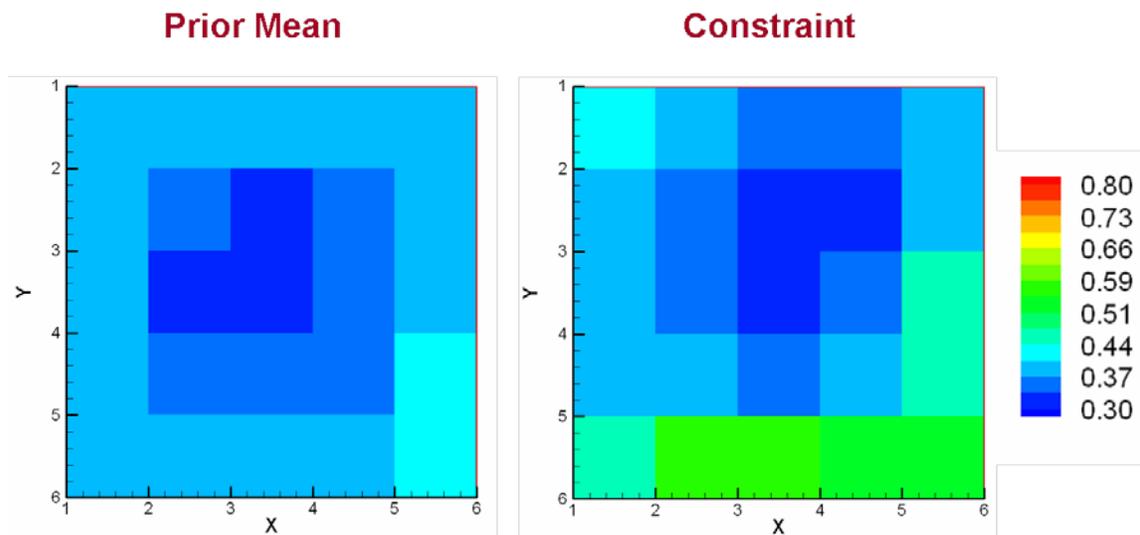


Figure 4.7 The coarse-scale constraint on the right serves as an observation to improve the estimates of the fluid saturation distribution in the ensemble prior on the left.

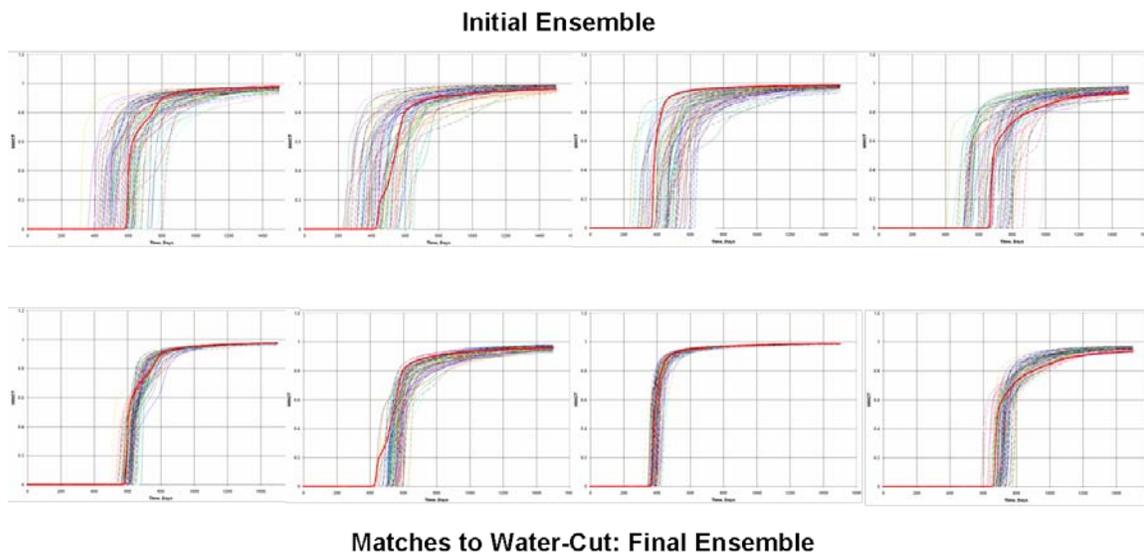


Figure 4.8 The top row shows the initial water-cut predictions and the bottom row shows the matches to the data from the updated realizations.

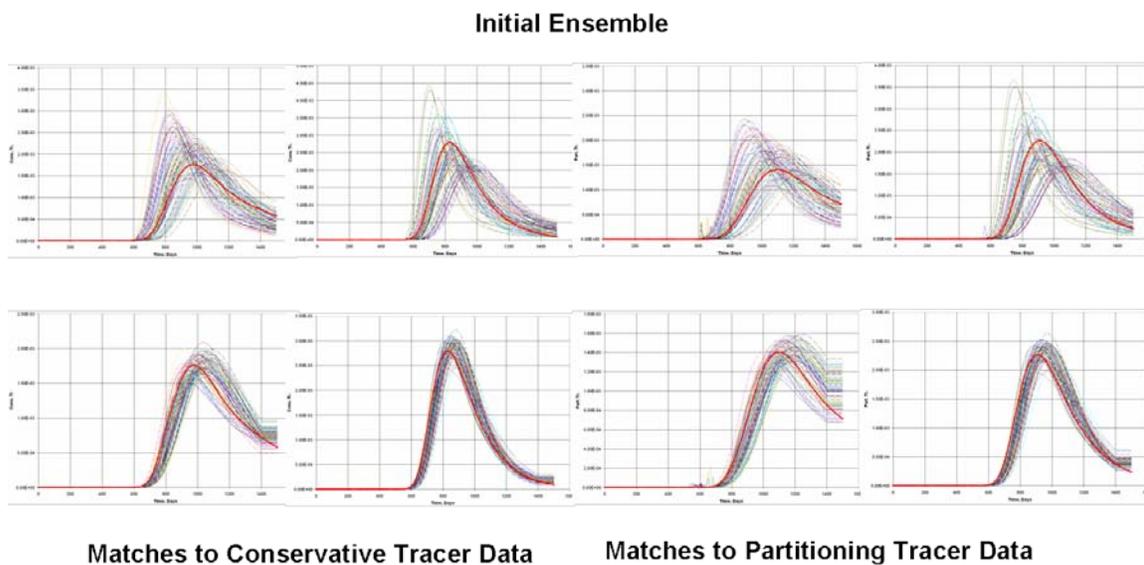


Figure 4.9 The top row shows the conservative and partitioning tracer responses from the initial ensemble for two selected wells and the bottom row shows the corresponding matches.

The matches to the data are shown in Figures 4.8 and 4.9. The resulting profiles are satisfactory and within reasonable bounds.

The objective of the proposed approach is to obtain better estimates of the subsurface measurements of oil saturations. In Figure 4.10, a comparison is made between the reference saturation, the saturation obtained from the single-stage standard EnKF and the reconstructed saturation profile sourced from the updates using the 2 stage multi-scale EnKF indicating significantly better estimates of bypassed oil saturations. These saturation profiles are obtained during the assimilation process.

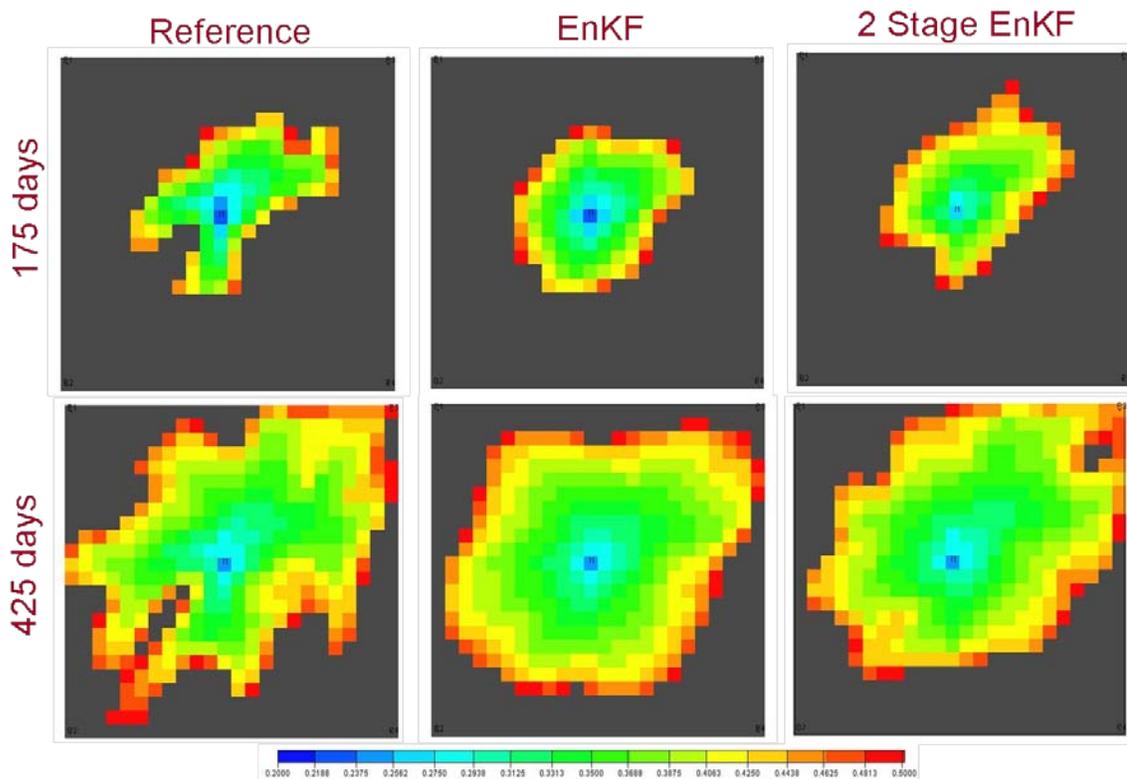


Figure 4.10 Comparison of the target oil saturations from the standard EnKF (center) with those obtained from the 2 stage EnKF (right) at two different times. The profile from the ‘true’ model is provided as a reference (left).

At a later time, the contrast between the standard EnKF and the 2-stage multi-scale EnKF is shown in Figure 4.11. The EnKF estimates are smooth due to the averaging properties when computing the ensemble mean. The secondary constraint comprising coarse-scale saturations obtained from a streamline-based inversion of production data is clearly able to better resolve the regions of high and the low saturations, while the estimates from the standard EnKF implementation are relatively smooth and do not adequately reflect the underlying heterogeneity. This can severely limit the efficacy of an EOR process designed using the results from the single-stage EnKF.

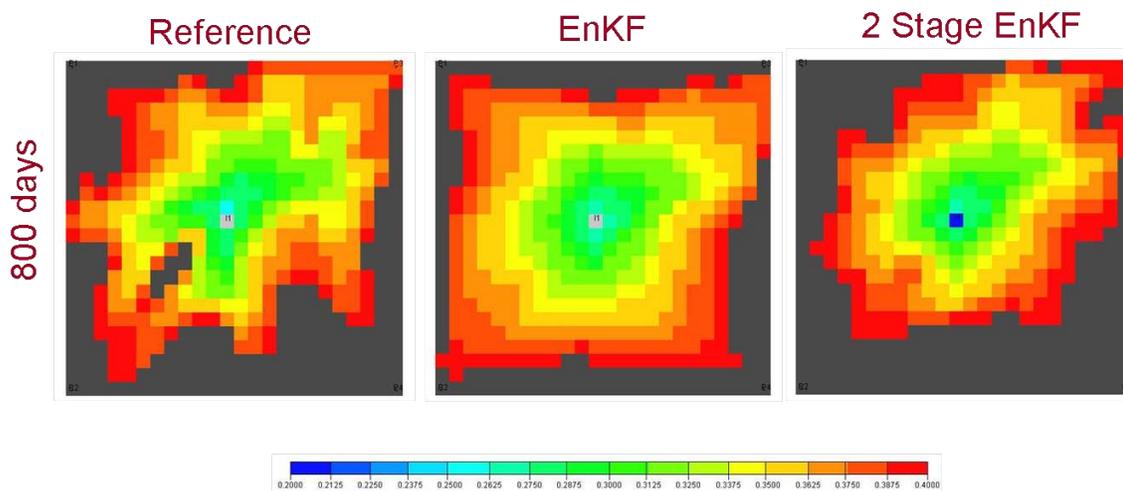


Figure 4.11 The 2-stage EnKF (right) outperforms the standard EnKF implementation (center) when compared to the reference saturation.

In the next section, we focus on an application to the 3-D hypothetical reservoir to demonstrate the utility of the approach for more complex problems.

4.5 Application: 3-D Example

In this section, the use of the multi-scale EnKF is demonstrated on a 3-dimensional

example with considerable underlying heterogeneity. The location of the wells in the reference model that is used to generate ‘observed’ data is shown in Figure 4.12. The reservoir is completed with 8 producing wells and 2 injectors. The computational domain is a 80x40x5 system. The initial set of 100 model realizations are conditioned to well data and conditional spatial statistics (SGSIM) (Deutsche and Journel 1992) is used to fully characterize the permeability in the remaining areas of the model. The total production time is 4000 days and the injected water is tagged with a conservative and partitioning tracer at 600 days into the waterflood.

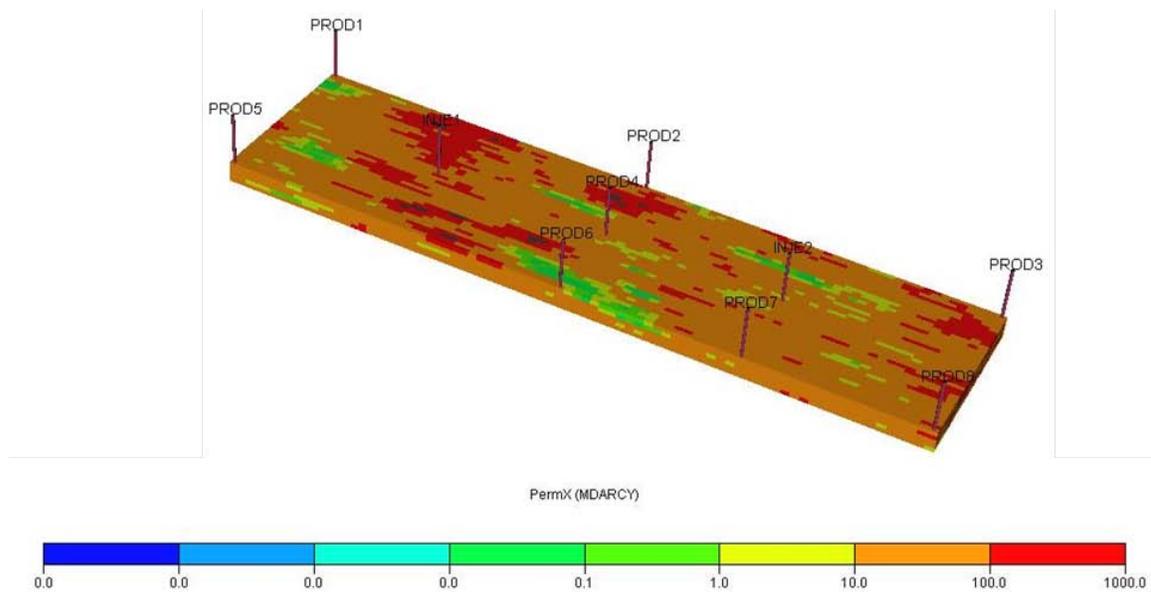


Figure 4.12 The reference log-permeability field showing the location of the producers and injection wells.

The standard form of the EnKF is first implemented to assimilate water-cut data and tracer measurements at each of the 8 producers. The sequence comprises 14 assimilation steps starting at 600 days when there is an observable water breakthrough and is spread over the next 2200 days in steps of 200 days. For the sake of brevity in this

dissertation, the matches to the observed history are plotted for only 3 selected wells in Figures 4.13a, 4.13b and 4.13c. The results from the other 5 producers are similar in the quality of the match. The considerable variability in the initial predictions is reduced to obtain a reasonable match.

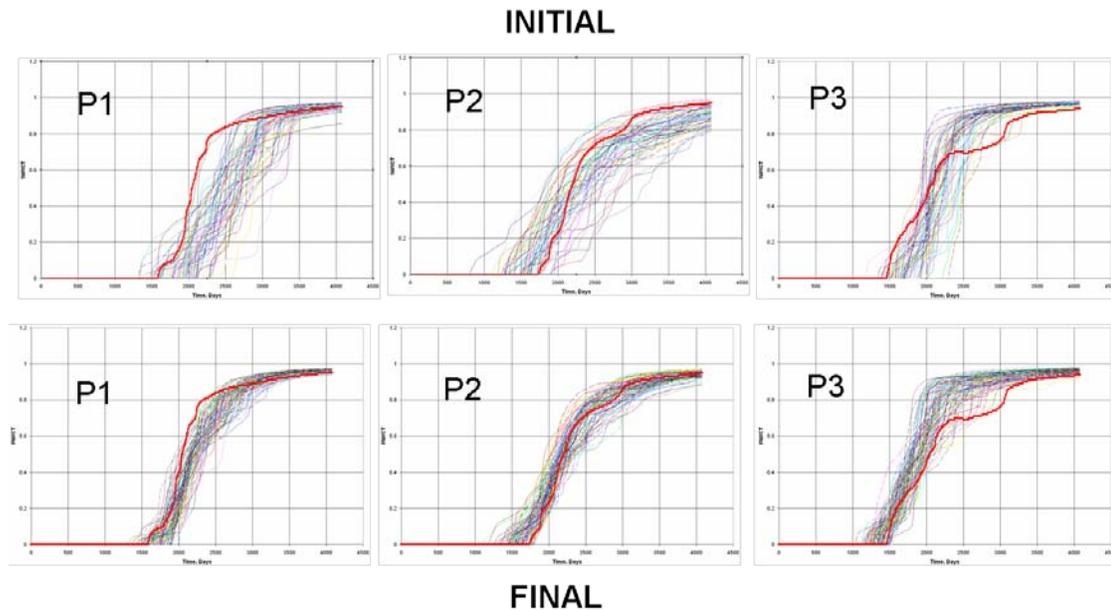


Figure 4.13a Matches to the water-cut profile in the bottom row. The initial ensemble predictions are in the top row. The only wells shown are P1, P2 and P3.

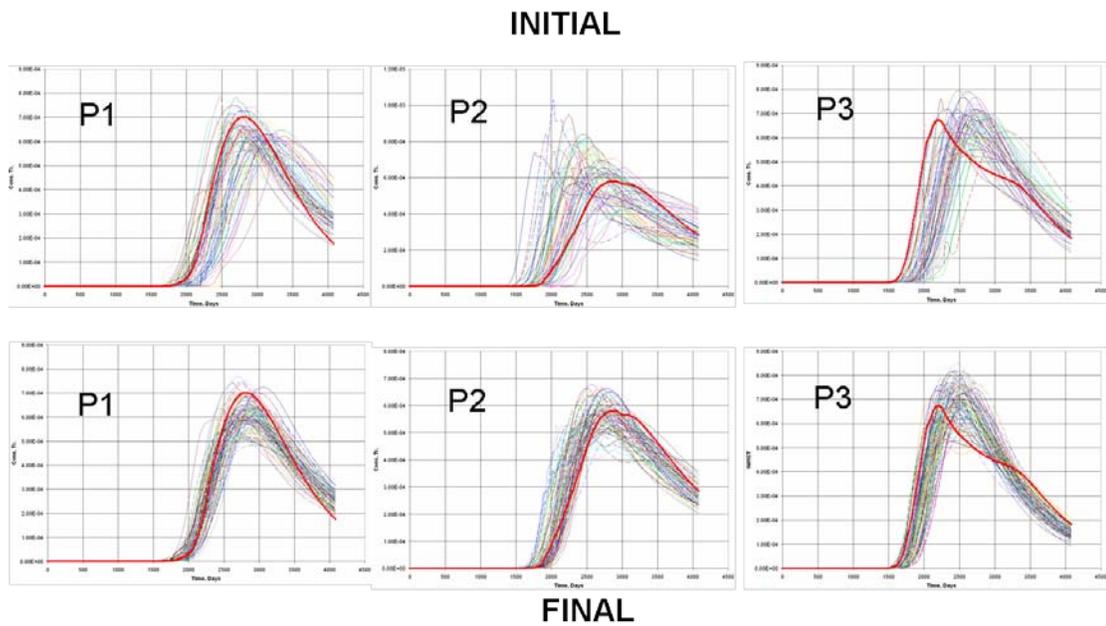


Figure 4.13b Matches to the conservative tracer data in the bottom row. The initial predictions are in the top row. The only wells shown are P1, P2 and P3.

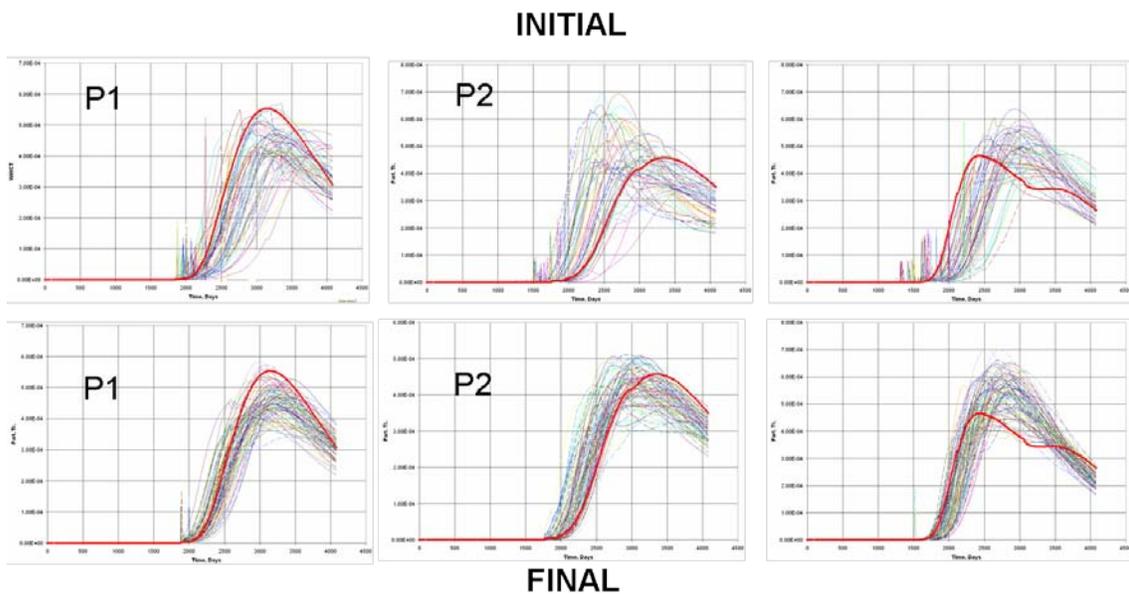


Figure 4.13c Matches to the partitioning tracer profile in the bottom row. The initial predictions are in the top row. The only wells shown are P1, P2 and P3.

However, a qualitative inspection, in Figures 4.14a through 4.14f, of the unswept non-aqueous phase obtained from the updated realizations reveal that the standard EnKF may not be capable of providing a realistic representation of the reference model.

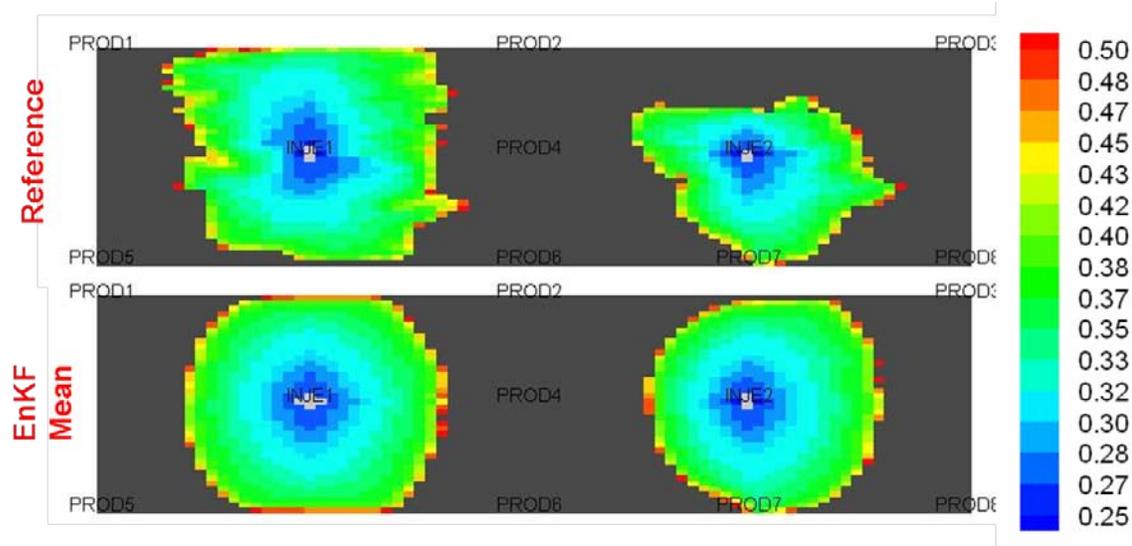


Figure 4.14a Saturation profile in Layer 1 at 1350 days.

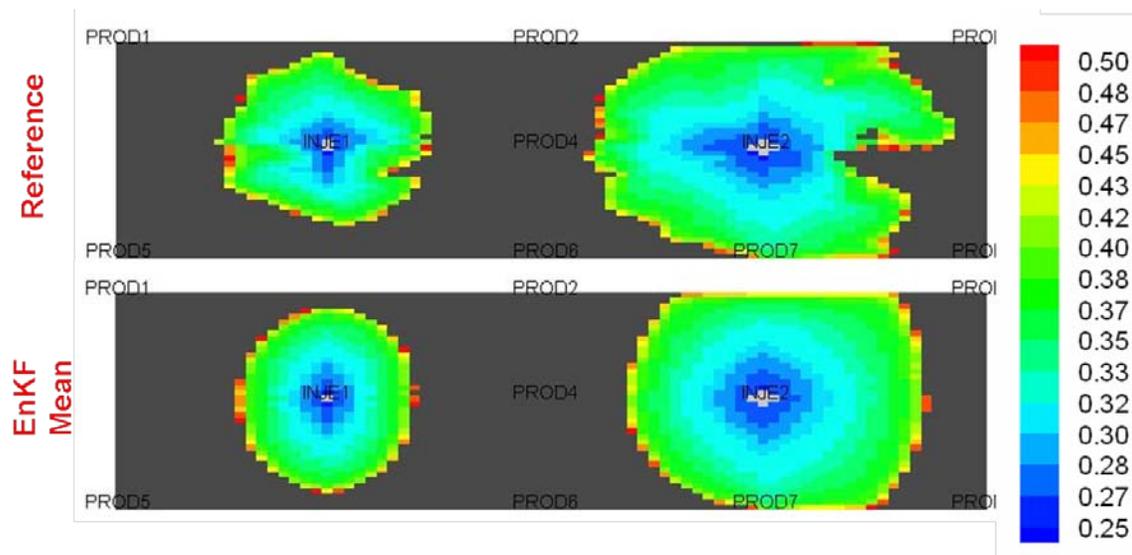


Figure 4.14b Saturation profile in Layer 5 at 1350 days.

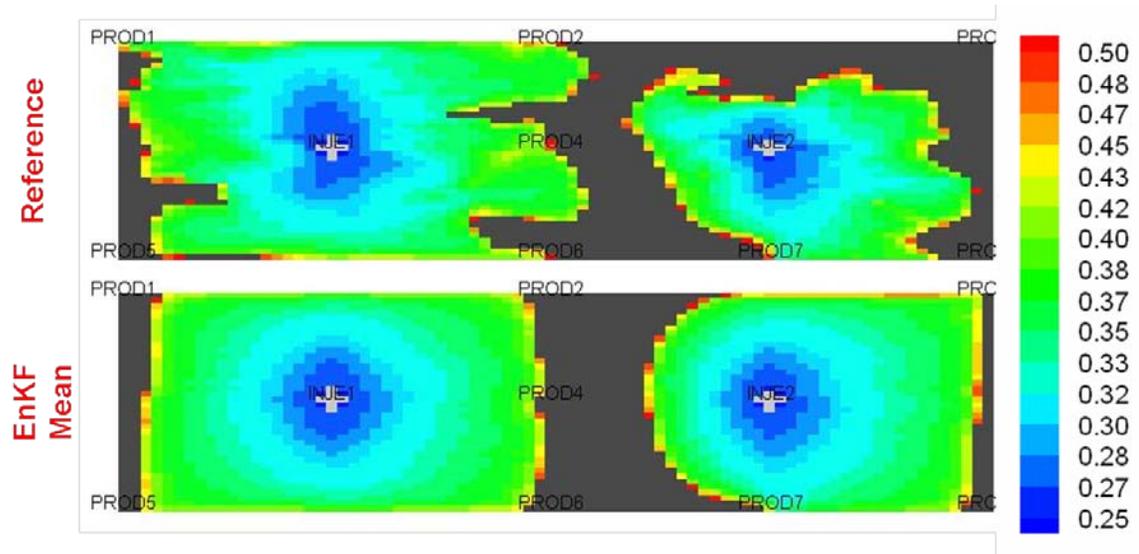


Figure 4.14c Saturation profile in Layer 1 at 2250 days.

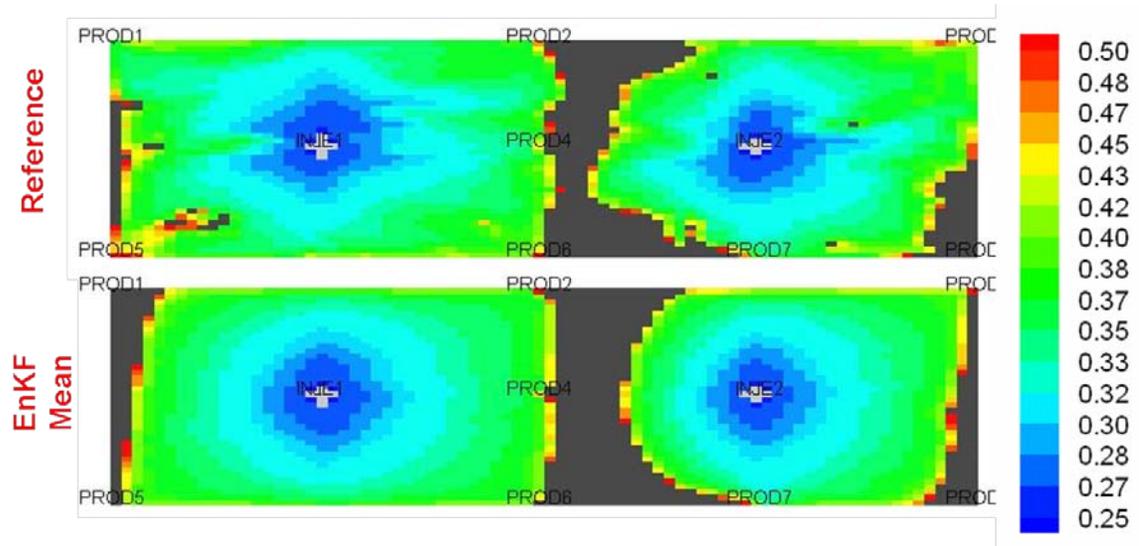


Figure 4.14d Saturation profile in Layer 5 at 2250 days.

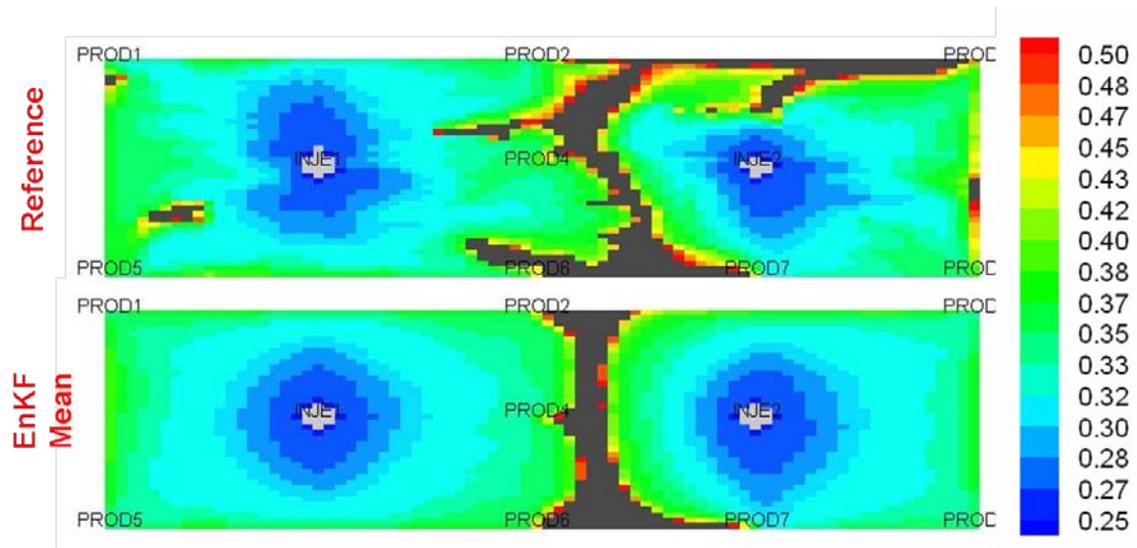


Figure 4.14e Saturation profile in Layer 1 at 4000 days.

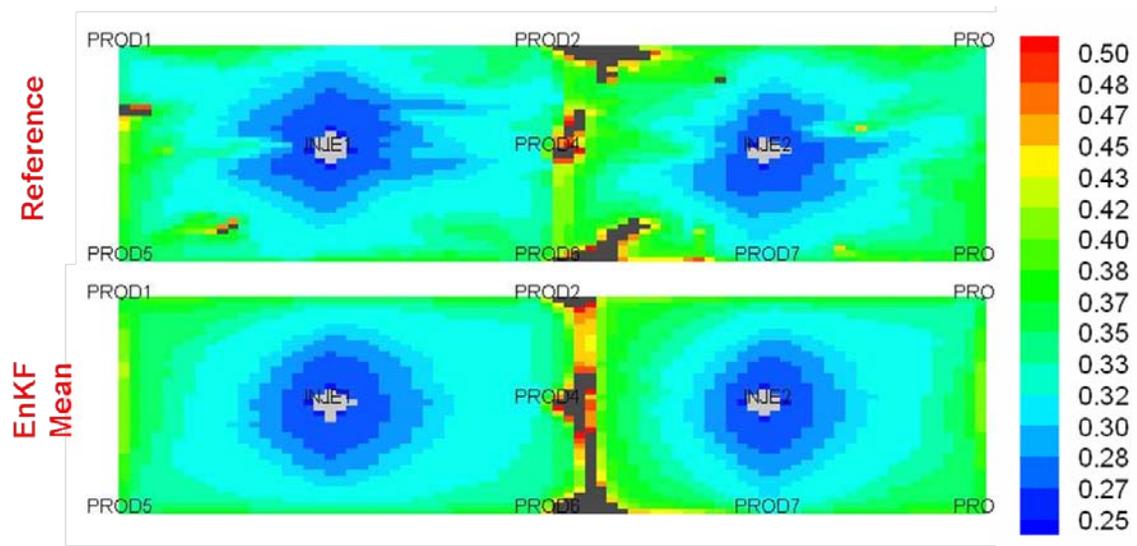


Figure 4.14f Saturation profile in Layer 5 at 4000 days.

At earlier times (1350 days and 2250 days) the differences between the predicted and observed bypassed oil saturation is considerable and at later times (4000 days) the

differences are less pronounced. However, the reconstructed saturations from the EnKF predictions remain relatively smooth and are not consistent with the reference model. Designing an EOR process based on these results may reduce the effectiveness of the recovery mechanism due to lack of resolution in the assessment of bypassed oil targets.

To further validate the utility of the 2-stage multi-scale implementation of the EnKF, we use coarse-scale observations obtained from a streamline-specific data integration framework during the assimilation process at 1800 days and 2400 days. The low-resolution saturation is utilized to tune the permeability and saturation of the individual model replicates. The procedure to obtain these low resolution measurements has been discussed in a previous section. The coarse-scale measurements are computed on a $20 \times 10 \times 5$ domain to preserve layering information.

At the assimilation time of 1800 days, the individual realizations are perturbed to honor the constraint shown in Figure 4.15 which also shows the smoother prior variations in the mean saturation.

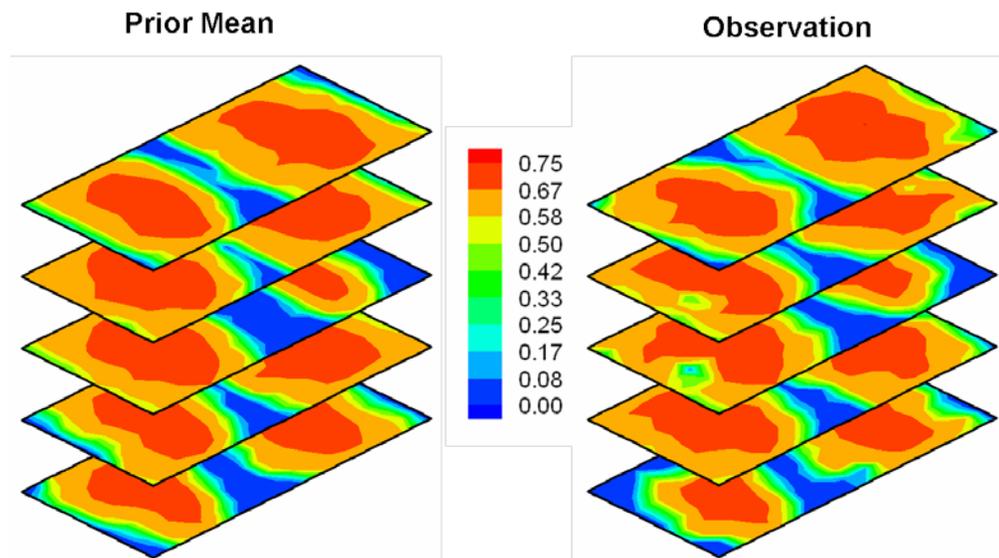


Figure 4.15 A comparison of the prior mean saturation on the left with the imposed coarse-scale constraint on the right.

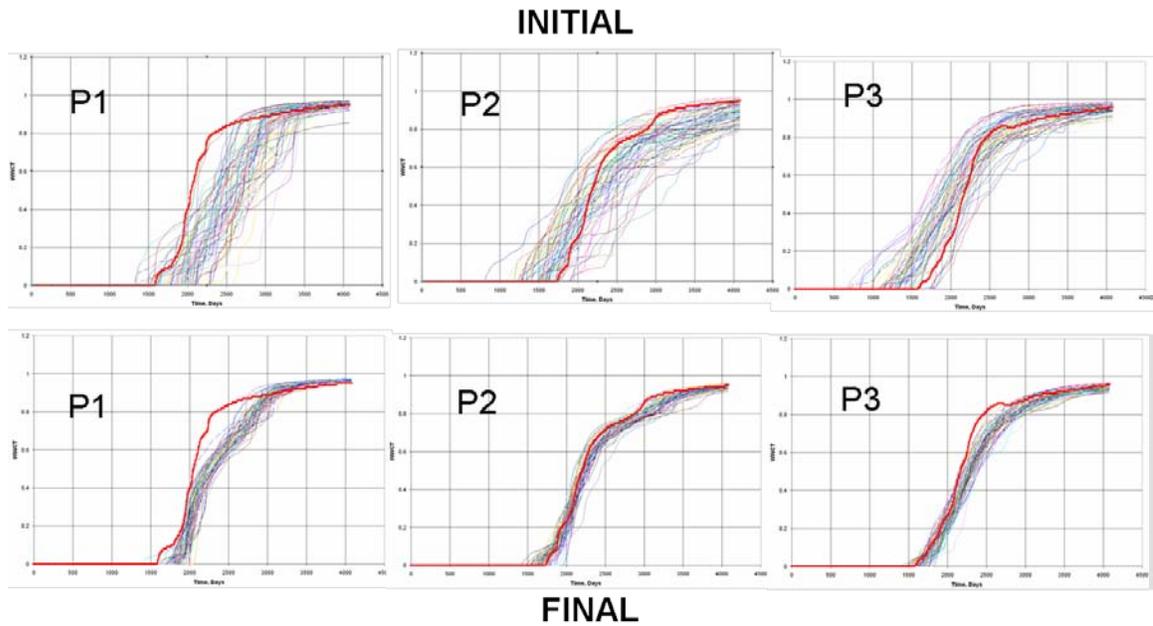


Figure 4.16a Matches to water-cut data in the bottom row for wells P1, P2 and P3 using the 2-stage EnKF. The initial predictions are in the top row.

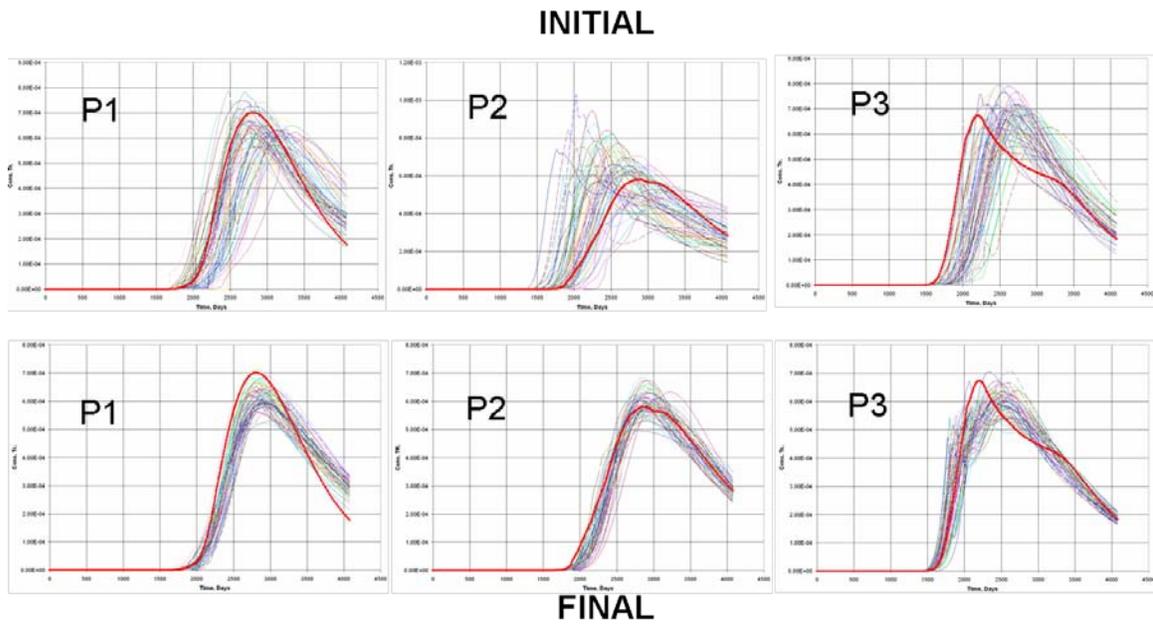


Figure 4.16b Matches to conservative tracer data in the bottom row for wells P1, P2 and P3 using the 2-stage EnKF. The initial predictions are in the top row.

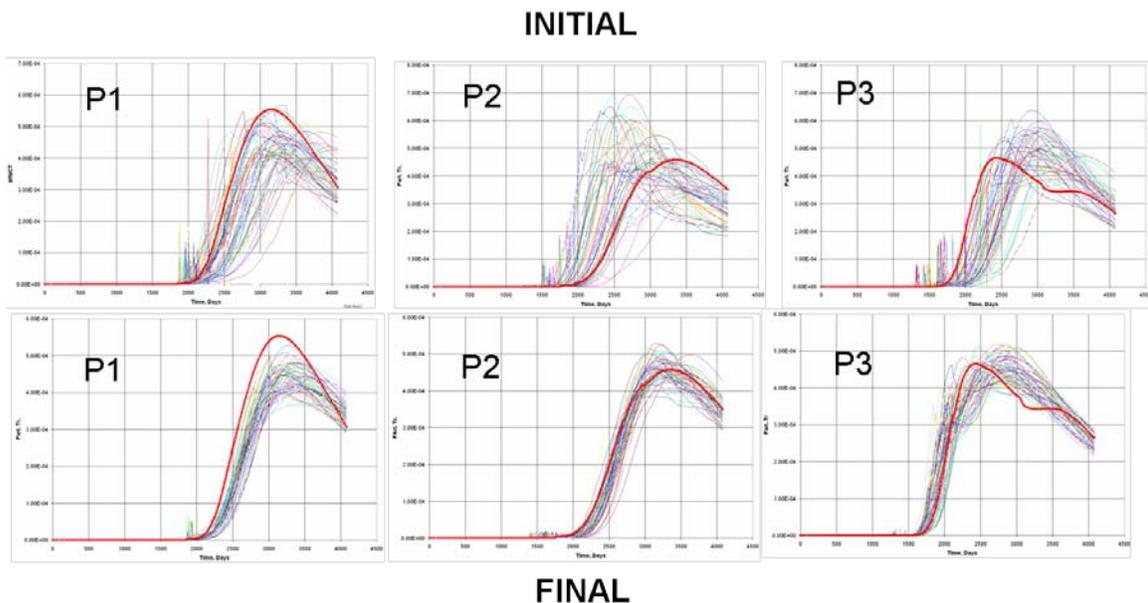


Figure 4.16c Matches to partitioning tracer data in the bottom row for wells P1, P2 and P3 using the 2-stage EnKF. The initial predictions are in the top row.

Figures 4.16a, b and c illustrate the satisfactory quality of the results in terms of matching available data. The primary benefit of the 2-stage approach can be seen in Figures 4.17a through 4.17f which plots saturation profiles at various times for different layers which show reasonably good portrayal of the underlying bypassed oil targets. The saturation variations are estimated by using the parameter set (permeability vector) at the end of 2400 days of assimilation. This involves a re-run of the simulator from initial time and is used only for comparison purposes.

The following plots reinforce the validity of the approach. The layers that demonstrate the most benefit by using a coarse-scale constraint are layers 1 and 5. The contrast in target oil saturation distributions in other layers was less dramatic. However, the comparison plots indicate that re-utilization of the data in an inversion framework to generate low resolution macro-scale observations is a particularly effective scheme for effective and reliable reservoir characterization.

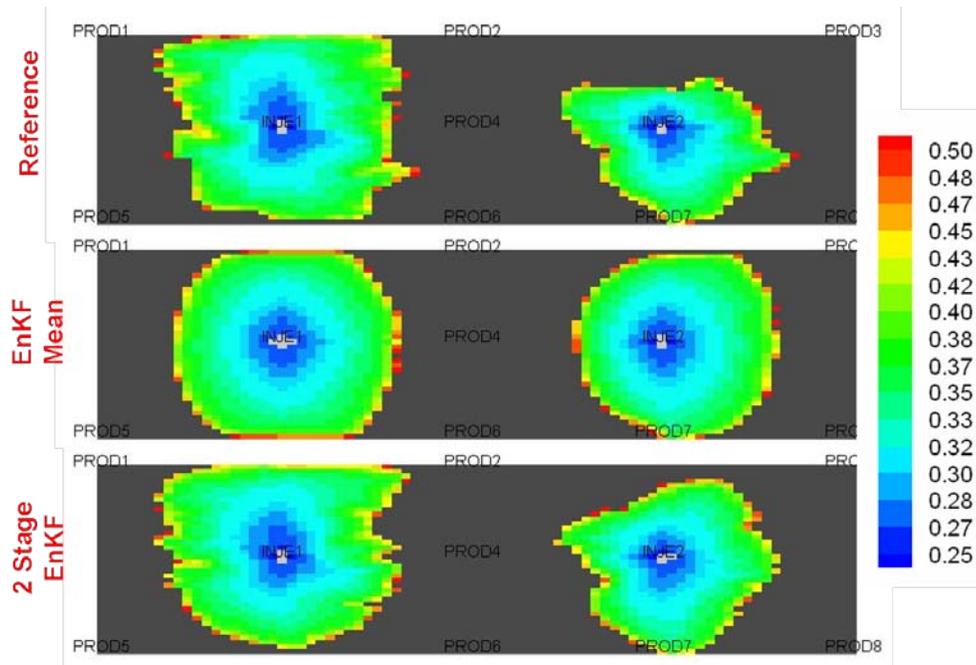


Figure 4.17a Contour plot of the bypassed oil saturation at 1350 days in Layer 1.

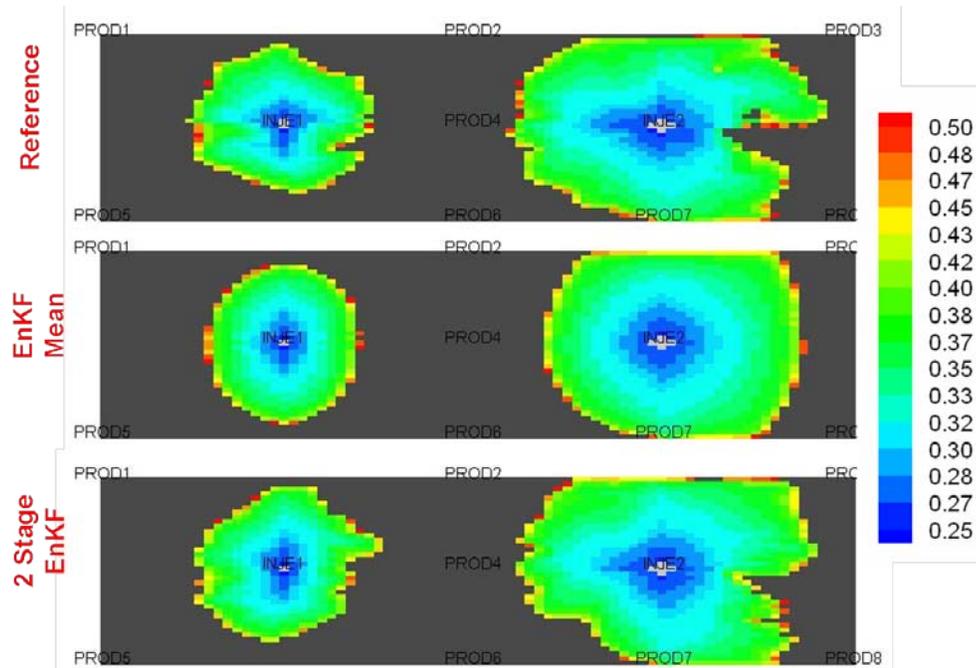


Figure 4.17b Contour plot of the bypassed oil saturation at 1350 days in Layer 5.

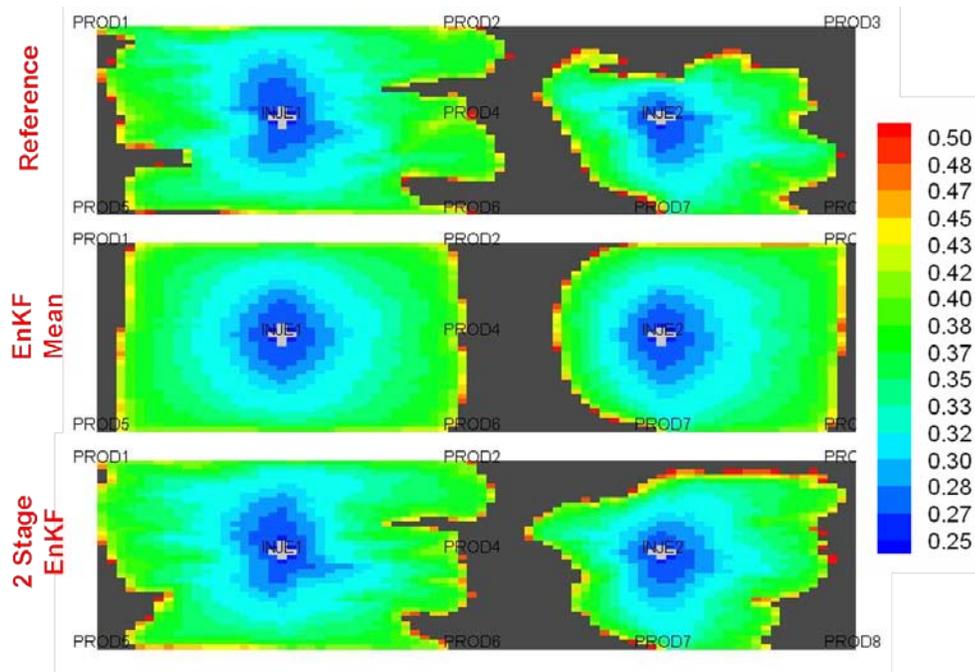


Figure 4.17c Contour plot of the bypassed oil saturation at 2250 days in Layer 1.

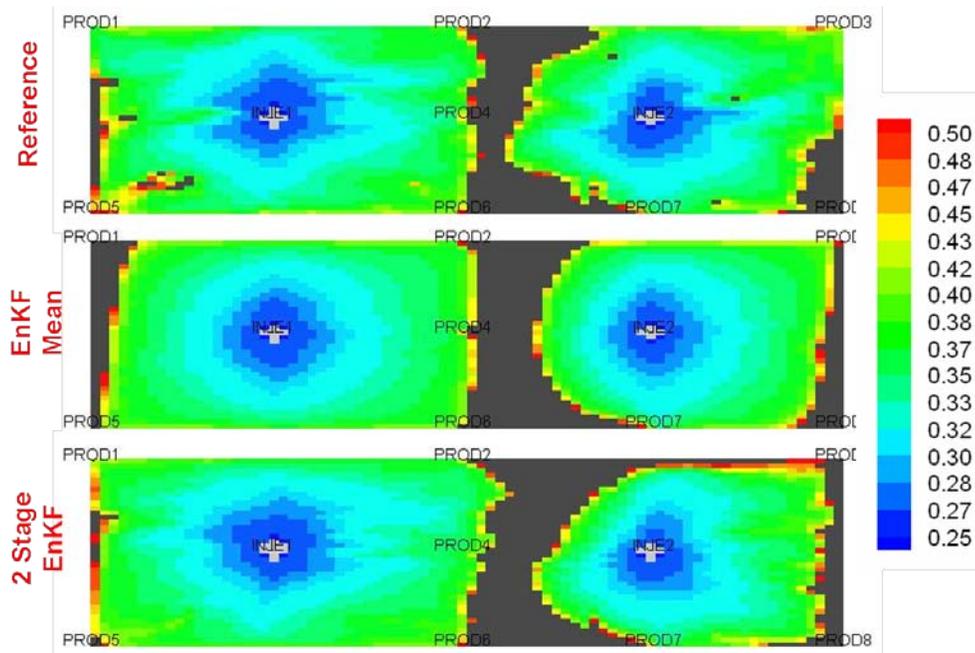


Figure 4.17d Contour plot of the bypassed oil saturation at 2250 days in Layer 5.

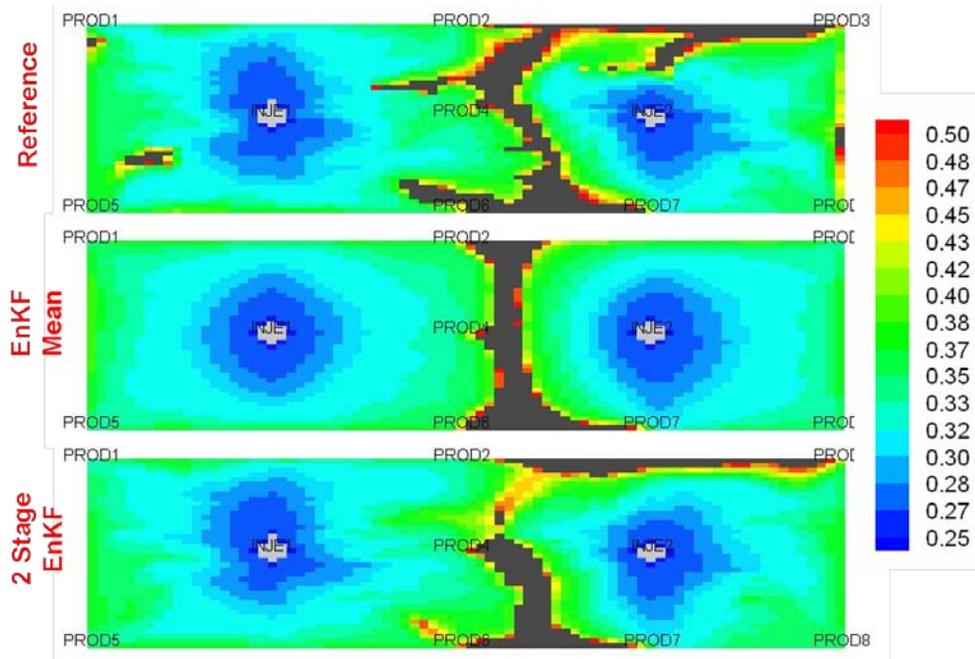


Figure 4.17e Contour plot of the bypassed oil saturation at 4000 days in Layer 1.

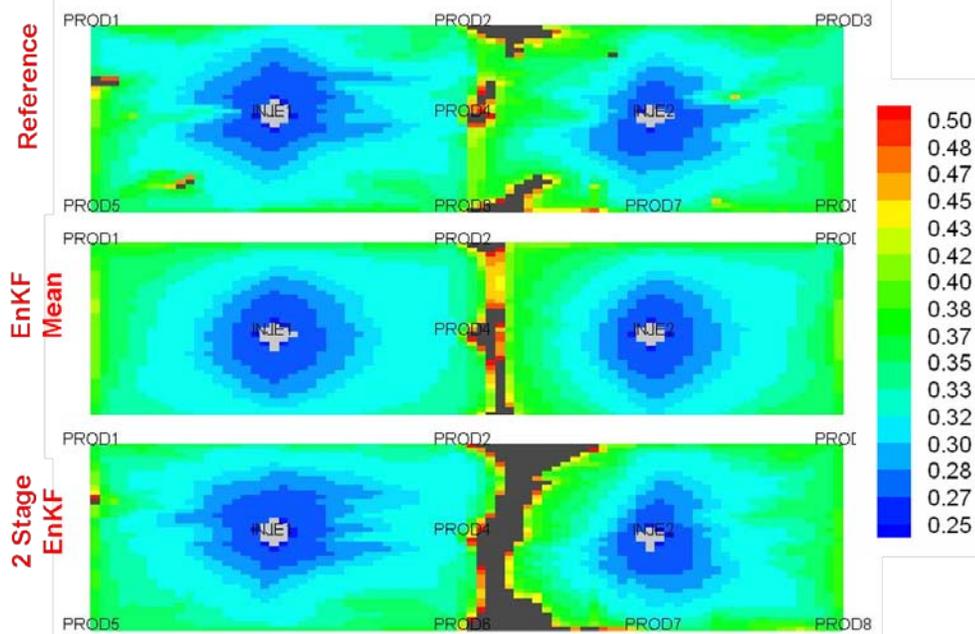


Figure 4.17f Contour plot of the bypassed oil saturation at 4000 days in Layer 5.

4.6 Section Summary

Identification of the distribution and location of target oil is a vital step in reservoir management strategy, specifically to design tertiary recovery processes or to plan infill drilling locations. The capability of tracers to contact large inter-well fluid volumes make them particularly well suited for reservoir modeling studies.

In this section, we investigate the use of the EnKF for PITT interpretation. In spite of the reasonable matches to well data, the resulting subsurface distribution of non-aqueous phase saturations is not consistent with the observations. This motivates the use of the 2-stage multi-scale EnKF to refine the resulting saturations to be reliable estimates of the truth. The operating principle is the coupling of a streamline-specific inverse modeling approach with the basic EnKF through coarse-scale constraints. In all examples studied, the proposed method indicates better resolution of the in-situ bypassed oil.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

Operational data assimilation problems for the purposes of reservoir characterization and uncertainty quantification tend to be very large and computationally challenging. The increased deployment of downhole monitors, permanent sensors and intelligent well systems intensifies the need for efficient algorithms for continual model updating via sequential data assimilation. The Ensemble Kalman Filter (EnKF) is a promising approach combining computational tractability and the capability of maintaining ‘live’ models conditioned to previously observed history. As opposed to traditional history matching approaches where the calibration of the reservoir model is achieved using all previously recorded historical data, the EnKF updated a suite of models whenever data becomes available. By working with an ensemble or suite of model replicates, the EnKF provides quantifying the underlying uncertainties, both in the model space and in future performance predictions

The EnKF utilizes sample-based statistical measures to tune the model realizations towards honoring production history. These statistical computations are necessary to specify cross-covariances between data and the model variables and are critically dependant on ensemble size for accuracy. Noisy and spurious terms in these estimates can severely degrade the EnKF analysis. The adverse affects of sampling errors has led to the development of many schemes specifically to target the deficiencies in the EnKF approach. Of these many approaches, the technique of covariance localization is particularly appealing for a wide variety of applications. For the purposes of reservoir model calibration, the use of streamlines is especially effective because the spatial localization achieved is tied to the underlying flow field and is consistent with the prior geologic information.

In this work, we examine two techniques for covariance localization using

streamline-specific information. The first approach utilizes streamline trajectories to delineate regions that will have the largest influence on the solution. Sample-based cross-covariance estimates are limited to these areas where there is significant streamline coverage. The second approach using streamline-derived sensitivities for covariance localization dampens the cross-covariance estimates within this area of influence through the use of normalized sensitivities. The primary advantages of the presented techniques is the ability to adaptively localize the cross-covariance to account for existing field conditions and well configurations thereby eliminating erroneous long and short range correlations. The proposed methods of covariance localization are shown to exhibit superior performance vis-à-vis the traditional form of the EnKF and this is successfully demonstrated with synthetic examples and real field cases.

A favorable aspect of the EnKF is its capability to integrate data originating from various sources and corresponding to different length-scales. This may be achieved either through recursive or simultaneous assimilation. In either case, the problem can be addressed using one single EnKF analysis subroutine contributing to the elegance of the algorithm. We exploit this feature of the EnKF to propose a 2-stage multi-scale EnKF and extend the application to interpretation of Partitioning Inter-well Tracer Tests (PITT). We show that the multi-stage EnKF using coarse-scale constraints outperforms the traditional implementation of the EnKF in identifying the spatial variation of the non-aqueous phase saturations.

Based on the findings from the results obtained in the different sections of this dissertation, the major conclusions from this work are summarized as follows:

- Sampling error is a significant source of noise in ensemble-based estimation of model covariance matrices. Variance-deficient ensembles, geologically inconsistent realizations, parameter overshooting and loss of spatial continuity are some problems attributed to a degraded EnKF analysis. These can be largely mitigated by the use of larger sized ensembles which can impose a large computation burden for field-scale problems.
- Streamlines provide unique advantages for covariance localization for ensemble-

based data assimilation. The particular forms of the proposed streamline-derived approaches outlined in section 2 enhance the EnKF analysis by manipulation of the cross-covariance matrices to retain physically relevant terms.

- The presented approaches to spatial localization tend to preserve geologic realism, avoid the problems associated with the loss of variability and maintain the connectivities of the extreme values of permeability; all of this is achieved at a reduced cost when compared to a larger ensemble providing equivalent performance.
- A coupled 2-stage multi-scale EnKF forms the basis for an efficient scheme to utilize information from PITT for estimation of the underlying fluid saturation distribution and the reservoir heterogeneities. The coupling superimposes information from a streamline-based data integration methodology operating with the ensemble mean onto the individual realizations of the ensemble. This is of significant benefit prior to the initiation of EOR processes.

5.2 Recommendations

Several recommendations that could improve the performance of the EnKF algorithm or extend the applications of presented concepts are listed below:

- Streamline-specific covariance localization is strictly valid for those types of production data directly related to fluid transport (for e.g. oil and water rates, water-cut or gas-oil ratio information). The extension of the methodology to incorporate bottom-hole pressure information is only an approximation. For the integration of these types of data, a more consistent reformulation of the localization scheme may be recommended, based on the trajectories of the evolution of the pressure front.
- The theory of covariance localization is extended using hierarchical EnKF's (Khare and Anderson 2006; Anderson 2007) where an ensemble of ensembles is used to provide a fairly general, application-independent means to detect sample error and apply corrections to the small-sample cross-covariance estimates. This

method was not investigated in the course of this dissertation, but the central idea of a hierarchical EnKF is appealing and could be the focus of a subsequent study.

- The 2-stage smoothness constrained EnKF appears to be a promising alternative, but a deeper investigation into the methodology might give greater insight into its relative merits and demerits. The basic formulation is predicated on the assimilation of the regularizing smoothness ‘observation’. Potentially, this could lead to significant benefits in terms of preserving geologic consistency and allow for smooth updates consistent with the low resolution and averaging properties of production data.
- In proposing the multi-scale EnKF, we essentially constrained the ensemble to coarse-scale information obtained from integration of production data to calibrate the ensemble mean. Another promising approach is to exploit the speed-up achievable with running a large set of coarsened realizations. This would have a few advantages: First, running with a large set of ensemble members implies that the associated problems with spurious correlations might be eliminated. Second, there is a significant computational advantage to using a coarse-scale model over a fine-scale model realization. Third, due to the use of low resolution models with fewer parameters, we can eliminate problems related to the existence of multiple minima with fine-scale model estimation; therefore it is more likely that the EnKF would converge to a global minimum.

REFERENCES

- Akella, S., Efendiev, Y. and Datta-Gupta, A. 2008. A Coarse-Scale Constrained Ensemble Kalman Filter for Subsurface Characterization. Paper submitted to *Advances in Water Resources* for review.
- Anderson, J.L and Anderson, S.L. 1999. A Monte-Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. *Monthly Weather Review* **127**(12): 2741-2758.
- Anderson, J.L. 2001. An Ensemble Adjustment Filter for Data Assimilation. *Monthly Weather Review* **129**(12): 2994-2903.
- Anderson, J.L. 2007. An Adaptive Covariance Inflation Error Correction Algorithm for Ensemble Filters. *Tellus* **59**(2): 210-224.
- Arroyo, E., Devegowda, D., Datta-Gupta, A. and Choe, J. 2006. Streamline Assisted Ensemble Kalman Filter for Rapid and Continuous Reservoir Model Updating. *SPE* **11** (6): 1046-1060. SPE-114255-PA.DOI:
- Bishop, C.H., Etherton, B.J. and Majumdar, S.J. 2001. Adaptive Sampling With The Ensemble Transform Filter. Part I: Theoretical Aspects. *Monthly Weather Review* **129**(3): 420-436.
- Cheng, H., Wen, X.H., Milliken, W.J. and Datta-Gupta, A. 2004. Field Experiences With Assisted and Automatic History Matching Using Streamline Models. Paper SPE 89857 presented at the SPE Annual Technical Conference and Exhibition, Houston, 24-27 September. DOI: 10.2118/89857-MS
- Cheng, H., Khargoria, A., He, Z. and Datta-Gupta, A. 2005. Fast History-Matching of Finite Difference Models Using Streamline-Derived Sensitivities. *SPE* **8**(5): 426-436. SPE-89447-PA. DOI: 10.2118/89447-PA.
- Cheng, H., Oyerinde, A.S. and Datta-Gupta, A. 2007. Compressible Streamlines and Three-Phase History Matching. *SPEJ* **12**(4):375-385. SPE-99465-PA. DOI:

- 10.2118/99465-PA.
- Deutsche, C.V and Journel, A. 1992. *GSLIB Geostatistical Software Library and User's Guide*. USA. Oxford University Press.
- Devegowda, D., Arroyo, E., Datta-Gupta, A. and Douma, S.G. 2007. Efficient and Robust Reservoir Model Updating Using Ensemble Kalman Filter With Sensitivity Based Covariance Localization. Paper SPE 106144 presented at the SPE Reservoir Simulation Symposium, The Woodlands, Texas, 24-26 February. DOI: 10.2118/106144-MS.
- Emanuel, A.S. and Milliken, W.J. 1998. History Matching Finite-Difference Models with 3D Streamlines. Paper SPE 49000 presented at the SPE Annual Technical Conference and Exhibition, New Orleans, 27-30 September. DOI: 10.2118/49000-MS.
- Evensen, G. 1994. Sequential Data Assimilation With a Nonlinear Quasi-Geostrophic Model Using Monte-Carlo Methods to Forecast Error Statistics. *J. of Geophysical Research* **99**(C5):10143-10162.
- Evensen, G. 2003. The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation. *Ocean Dynamics* **53**:353-367.
- Furrer, R. and Bengtsson, T. 2004. Estimation of High-Dimensional Prior and Posterior Covariance Matrices in Kalman Filter Variants. *J. of Multivariate Analysis* **98**: 227-255.
- Gaspari, G. and Cohn, S.E. 1996. Construction of Correlation Functions in Two and Three Dimensions. Office Note Series on Global Modeling and Data Assimilation, DAO Office Note 96-03R1.
- Gu, Y. and Oliver, D.S. 2006. The Ensemble Kalman Filter for Continuous Reservoir Model Updating. *Journal of Energy Resources Technology* **128**: 79-91.
- Hamill, T.M and Whitaker, J.S. 2001. Distance-Dependant Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Monthly Weather Review* **129**: 2776-2793.
- He, Z., Yoon, S. and Datta-Gupta, A. 2002. Streamline Based Production Data

- Integration Under Changing Field Conditions. *SPEJ* **7**(4):423-436. SPE-81208-PA. DOI: 10.2118/81208-PA.
- Houtekamer, P.L and Mitchell, H.L. 1998. Data Assimilation Using Ensemble Kalman Technique. *Monthly Weather Review* **126**(3):796-811.
- Houtekamer, P.L and Mitchell, H.L. 2001. A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review* **129**(1):123-137.
- Jimenez, E., Sabir, K., Datta-Gupta, A. and King, M. J. 2007. Spatial Error and Convergence in Streamline Simulation. *SPEE* **10**(3): 221-232. SPE-92873-PA. DOI: 10.2118/92873-PA.
- Johns C.J. and Mandel, J. 2005. A Two-stage Ensemble Kalman Filter for Smooth Data Assimilation. *Environmental and Ecological Statistics*, in print.
- Khare, S.P. and Anderson, J.L. 2006. An Examination of Ensemble Kalman Filter Based Adaptive Observation Methodologies. *Tellus* **58**(2): 179-195.
- Nævdal, G., Johnsen, L.M., Aanonsen, N.I. and Vefring, E.H. 2005. Reservoir Monitoring and Continuous Model Updating Using Ensemble Kalman Filter. *SPEJ* **10**(1):66-74. SPE-84372-PA. DOI: 10.2118/84372-PA.
- Oke, P.R., Sakov, P. and Corney, S.P. 2007. Impacts of Localization in the EnKF and EnOI: Experiments With a Small Model. *Ocean Dynamics* **57**(1):32-45.
- Oyerinde, A.S. 2004. A Composite Tracer Analysis Approach to Reservoir Characterization. MS thesis, Texas A&M University, College Station.
- Oyerinde, A.S. 2008. Streamline Based Three Phase History Matching. PhD dissertation, Texas A&M University, College Station.
- Pollock, D.W. 1998. Semi-analytical Computation of Path Lines for Finite Difference Models. *Ground Water* **26**(6): 743-752.
- Tarantola, A. 2004. Inverse Problem Theory and Methods for Model Parameter Estimation. USA. Society for Industrial and Applied Mathematics.
- Valestrand, R., Sagen, J., Nævdal, G., Huseby, O. and Forghany, S. 2008. The Effect of Including Tracer Data in the EnKF Approach. SPE 113440 presented at the 2008 SPE Improved Oil Recovery Symposium, Tulsa, OK, 19-23 April. DOI:

10.2118/113440.

- Vasco, D.W., Yoon, S. and Datta-Gupta, A. 1999. Integrating Dynamic Data Into High Resolution Reservoir Models Using Streamline-Based Analytic Sensitivity Coefficients. *SPEJ* 4(4):389-399. SPE-59253-PA: DOI: 10.2118/59253-PA.
- Vega, L., Rojas, D., and Datta-Gupta, A. 2004. Scalability of the Deterministic and Bayesian Approaches to Production-Data Integration Into Reservoir Models. *SPEJ* 9(3): 330-338. SPE-88961-PA. DOI: 10.2118/88961-PA.
- Whitaker, J.S. and Hamill, T.M. 2002. Ensemble Data Assimilation without Perturbed Observations. *Monthly Weather Review* **130**(7):1913-1924.
- Yoon, S., Barman, I., Datta-Gupta, A. and Pope, G.A. 1999. In-Situ Characterization of Residual NAPL Distribution Using Streamline Based Inversion of Partitioning Tracer Tests. Paper SPE 52729 presented at the SPE/EPA Exploration and Production Environmental Conference, Austin, 1-3 April. DOI: 10.2118/52729-MS.
- Zhang, F., Zhang, M. and Hansen, J.A. 2007. Coupling Ensemble Kalman Filter with Four Dimensional Variational Data Assimilation. Submitted to *Monthly Weather Review* for publication.

VITA

Deepak Devegowda completed his B.Tech. degree at the Indian Institute of Technology, Department of Electrical Engineering, Chennai, India in July 1998, and later obtained his M.S. and Ph.D. degree at the Harold Vance Department of Petroleum Engineering, Texas A&M University, in December, 2003 and August, 2009. His permanent address is Mewbourne School of Petroleum and Geological Engineering, T-301, Sarkeys Energy Center, University of Oklahoma, 100, E. Boyd St., Norman, Oklahoma 73019.