T H E U N I V E R S I T Y O F T U L S A

THE GRADUATE SCHOOL

ASSESSING THE UNCERTAINTY IN RESERVOIR DESCRIPTION AND PERFORMANCE PREDICTIONS WITH THE ENSEMBLE KALMAN FILTER

by
Mohammad Zafari

A thesis submitted in partial fulfillment of
the requirements for the degree of Master of Science
in the Discipline of Petroleum Engineering

The Graduate School
The University of Tulsa

2005

T H E U N I V E R S I T Y O F T U L S A

THE GRADUATE SCHOOL

ASSESSING THE UNCERTAINTY IN RESERVOIR DESCRIPTION AND PERFORMANCE PREDICTIONS WITH THE ENSEMBLE KALMAN FILTER

by

Mohammad Zafari

A THESIS

APPROVED FOR THE DISCIPLINE OF

PETROLEUM ENGINEERING

By Thesis Committee

_____ , Chairperson

Dr. Albert C. Reynolds

_____

Dr. Peyton Cook

_____

Dr. Gaoming Li

COPYRIGHT STATEMENT

ABSTRACT

Mohammad Zafari (Master of Science in Petroleum Engineering)

Assessing the Uncertainty in Reservoir Description and Performance Predictions With the Ensemble Kalman Filter

Directed by Dr. Albert C. Reynolds

92 pp. Chapter 6

(297 words)

Recently, the ensemble Kalman filter (EnKF) has gained popularity in atmospheric science for the assimilation of data and the assessment of uncertainty in forecasts for complex, large-scale problems. A handful of papers have discussed reservoir characterization applications of the EnKF, which can easily and quickly be coupled with any reservoir simulator. Neither adjoint code nor specific knowledge of simulator numerics is required for implementation of the EnKF. Moreover, data are assimilated (matched) as they become available; a suite of plausible reservoir models (the set of ensembles) is continuously updated to honor data without rematching data assimilated previously. Because of these features, the method is far more efficient for history matching dynamic data than automatic history matching algorithms based on optimization algorithms. Moreover, the suite of ensembles provides a way to evaluate the uncertainty in reservoir description and performance predictions. Here we establish a firm theoretical relation between randomized maximum likelihood and the ensemble Kalman filter. We show that EnKF is similar to doing one Gauss-Newton iteration for matching data at one time using an average sensitivity matrix where the average sensitivity matrix is constructed from the set of predicted data generated from the set of ensembles obtained from the previous time at which different data were assimilated. Starting from another viewpoint, we show that if predicted data are included in the random state vector along with model parameters, then

the EnKF update equation can be obtained from a single Gauss-Newton iteration based on an average sensitivity matrix. We also show that the mean of the set of ensembles is equivalent to the cokriging estimate. We present examples where its performance in terms of characterizing uncertainty is not completely satisfactory, as well as examples where the performance of EnKF provide a reliable characterization of uncertainty.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my advisor, Dr. Albert C. Reynolds, Jr. for his continuous support, guidance, inspiring discussions, patience and encouragement during my study and research in the University of Tulsa. I would like to thank my master committee members, Dr. Gaoming Li and Dr. Peyton Cook for their instructions, comments, critiques and suggestions. I would like to extend my thanks to all the other faculty members and graduate students of the Petroleum Engineering department at TU, and all my friends who helped and encouraged me in the past 2 years. At last but not least I am thankful to my family for their unconditional love, support and patience over the last few years.

This work is dedicated to my parents, Zabihollah Zafari and Akram Hamidi for their endless love.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Our main interest is in characterizing the uncertainty in reservoir description and reservoir performance prediction as the first step towards the optimization of reservoir management. To do so, we wish to generate a suite of plausible reservoir models (realizations) that are consistent with all information and data. If the set of models are obtained by correctly sampling the pdf, then the set of models give a characterization of the uncertainty in the reservoir model. By predicting future reservoir performance with each of the realizations, and calculating statistics on the set of outcomes, one can evaluate the uncertainty in reservoir performance predictions.

Uncertainty can only be formally characterized using the mathematics of probability and statistics. For data integration problems of interest in reservoir modelling and characterization, Bayesian statistics provides a convenient framework for characterizing and evaluating uncertainty and a convenient framework for assimilating data as they become available. In particular, starting with a given probability density function (pdf) for a random vector, the Bayes theorem can be applied to derive the pdf conditional to the observed data, i.e., to update the pdf. This updated pdf may be thought of as the new prior pdf, which itself can then be updated as data become available at some later time. The application of Bayes theorem to write a theoretical expression for updating a pdf by assimilating data is general; it is not necessary to assume the prior or any of the pdf's are Gaussian or assume a linear relation between data and the model.

1.1 Literature Review

Even though Bayes theorem can be applied to write a theoretical expression for the pdf of interest, sampling the pdf is much more difficult for reservoir problems that contain

thousands or several tens of thousands of model parameters. For such problems, we believe that theoretically rigorous methods of sampling such as the Markov chain Monte Carlo method (MCMC) or rejection algorithm are not computationally feasible [3, 22], although there exist references that suggest otherwise [2]. On attempt to define an algorithm that is somehow similar to MCMC, a method now commonly known as randomized maximum likelihood (RML) was introduced by [27]. A precursor of this method was given in [7] and the theoretical bases for the case where data are linearly related to the model was given by [26]. The method we refer to as RML was also suggested independently by [19].

RML is also firmly embedded in the Bayesian framework, and its application is aimed at generating a correct sampling of a pdf conditional to observed data. Although the prior pdf is arbitrary, to render RML computationally attractive, one normally assumes the "prior" pdf for the model may be reasonably approximated by a multivariate Gaussian distribution at least for the purpose of conditioning an unconditional realization of the prior model to observed data. As commonly applied [17, 21, 35, 8], RML generates a realization of a pdf that is conditional to all observed data by minimizing an appropriate objective function. However, if the total data set is partitioned into sub data sets where measurement errors for each data subset are uncorrelated with measurement errors for all other data subsets, then it is possible to condition a realization from the prior model to each data subset sequentially provided that after each data set is integrated, the prior covariance and mean are updated to the posterior covariance and mean. One can show that this Bayesian updating procedure, where uses only means and covariances, is rigorously valid only when the relation between data and model is linear [30]. Moreover, for large scale problems, standard formulas for updating the covariance matrix are impractical, and in fact, storage of this covariance matrix may not be practical. Nevertheless, this sequential conditioning of a model to data is very much in the spirit of the ensemble Kalman filter (EnKF).

Although RML can only be proven to sample correctly in the linear case, i.e., when data is linearly related to the model, several computational experiments suggest that the method does an adequate job of sampling in the nonlinear case [28, 22, 12].

When RML is implemented using the limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) [25, 35] algorithm for optimization with the gradient of the objective function computed with the implementation of the adjoint method [6, 5] given by [21], large scale history matching problems are feasible. However, our experience is that each realization (approximate sample of the pdf) generated by RML usually requires computational time that is equivalent to at least 50 to 100 forward reservoir simulation runs.

Very recently, some research groups have begun extensive work on using the ensemble Kalman filter to assimilate data and characterize the predicted performance. The ensemble Kalman filter (EnKF) was originally introduced by [9] as a sequential data assimilation algorithm. The sequential algorithm consists of two major steps, first, a forecast step which is equivalent to running the simulation to predict data at the next assimilation time step, second, an analysis and updating step which includes the implicit calculation of the Kalman gain matrix and updating the model parameters so that they are consistent with data. EnKF is a Monte Carlo approach, where an ensemble of reservoir models is used to construct the error statistics of model parameters and predicted data. The ensemble of states are effectively a set of unconditional realizations for the next assimilation time step with predictions of data. Model parameters and "state" variables generated from the ensemble members used to estimate covariance matrices for the purpose of assimilating data at the next assimilation time step. Later, [4] showed that in the implementation of the analysis scheme in the EnKF, the observations must be treated as random variables. This can be done by adding noise with the correct statistics to the observations. However, [32] has shown that the addition of noise to the observations in an ensemble data assimilation system reduces the accuracy of the analysis error covariance estimate by increasing the sampling error, and increases the probability that the analysis error covariance will be underestimated by the ensemble.

In [18] a variant of the ensemble Kalman filter was introduced, where in their new approach a double ensemble is used and the gain of each ensemble is used to update the other ensemble in the analysis step. They argued that the systematic underestimation of covariances are much less likely to happen in the new method. However, [20] has

shown that EnKF leads to systematically underestimated variances for small ensemble sizes regardless of whether the ensemble is updated with a gain calculated from that same ensemble or not. Where the variances are underestimated, the filter assumes the relative accuracy of the prior information compared to the accuracy of the data is greater than it is, and consequently the updated model parameters are insufficiently corrected toward the truth. Therefore, algorithm modifications are necessary to ensure that covariances are not systematically underestimated [16, 1]. In petroleum engineering field [24, 23] were the first to apply the EnKF method to reservoir characterization and history matching. Recently, [15, 12] applied it to the well known PUNQ-S3 problem data set.

1.2 Objectives and Research Scope

The advantage of the EnKF is that no sensitivity calculations are required, i.e., the work of developing adjoint code for a specific simulator is avoided and the EnKF method can be easily coupled with any reservoir simulator. Furthermore, dynamic data is assimilated continuously in time and covariance matrices are updated at each data assimilation time step. Another key advantage is that new production data or seismic data can be assimilated as they become available; there is no need to rerun the simulator from time zero and match the old data and no iteration is done; the idea is the old data is automatically honored because it was incorporated as covariances were updated.

In this work, we attempt to theoretically compare the EnKF update equations with the RML method. In particular, when data are linearly related to the model, we show that both methods give a correct sampling of the pdf conditional to all data up to the final assimilation time. Moreover, it shows that EnKF really uses predicted data to construct crude approximations to derivatives that would be used if the Gauss-Newton method were used to minimize the objective function when applying RML to generate a realization of the same pdf. Based on this analogy, we provide an alternate derivation and interpretation of the computational equations for EnKF. Our results suggest that EnKF implicitly assumes that the autocovariance of predicted data, the model and their cross-covariances can be approximated by Gaussian distributions for the purpose of data

assimilation. We also show that the mean of the set of ensembles is equivalent to the cokriging estimate.

In our implementation of RML, the random vector that is updated includes the model parameters, primary variables and predicted data. Where history matching or assimilating production data, predicted data refer to quantities such as wellbore pressure, producing gas-oil ratio (GOR), producing water-oil ratio (WOR) or phase rates. The model parameters refer to quantities such as gridblock permeabilities and porosities and primary variables refer to variables predicted by reservoir simulation, e.g., gridblock pressures and saturations. As EnKF updates not only model parameters but primary variables at a data assimilation step, it is possible to obtain physically unreasonable values of pressures and saturations that must be modified to physically reasonable values. In our history matching application, we simply truncate any such value back to a prescribed upper or lower bound for the variable. Intuitively, one might expect that this problem could be eliminated by rerunning the simulation time step, but our theoretical results show that this will lead to an incorrect stochastic sampling procedure. Many of the results presented here have been presented publicly in [34, 33].

CHAPTER 2

THEORETICAL FRAMEWORK

2.1 Randomized Maximum Likelihood Method

2.1.1 The RML Algorithm

Assuming a prior multivariate Gaussian pdf for the model m, that the vector of data measurement errors is Gaussian with mean zero and covariance matrix C_D, the posterior pdf conditional to an N_d-dimensional column vector of observed data, d_{obs} is given by

$$f(m|d_{\text{obs}}) = c\exp[-O(m)], \tag{2.1}$$

where c is the normalizing constant, and $O(m)$ is given by

$$\begin{aligned} O(m) =& \frac{1}{2}(m - m_{\text{prior}})^T C_M^{-1}(m - m_{\text{prior}}) \\ &+ \frac{1}{2}(g(m) - d_{\text{obs}})^T C_D^{-1}(g(m) - d_{\text{obs}}). \end{aligned} \tag{2.2}$$

Throughout, N_m denotes the dimension of the model space so m is an N_m-dimensional column vector. (Unless noted otherwise, all vectors introduced are column vectors.)

Let the N_m-dimensional random vector Z_M be normal with mean equal to the N_m-dimensional zero vector and covariance given by I_{N_m}, the $N_m \times N_m$ identity matrix. Let L_M be the lower triangular matrix such that

$$C_M = L_M L_M^T, \tag{2.3}$$

is the Cholesky decomposition of the prior covariance matrix C_M, then it is easy to show

that the random vector

$$m_{uc} = m_{\text{prior}} + L_M Z_M, \tag{2.4}$$

is Gaussian with mean m_{prior} and covariance matrix given by C_M, i.e., Eq. 2.4 provides a way to sample the prior multivariate Gaussian distribution. The N_d-dimensional random vector Z_D represents the multinormal random vector with zero mean and covariance given by the $N_d \times N_d$ identity matrix I_{N_d}. If L_D is the lower triangular matrix such that

$$C_D = L_D L_D^T, \tag{2.5}$$

is the Cholesky decomposition of the covariance matrix C_D, then

$$d_{uc} = d_{\text{obs}} + L_D Z_D, \tag{2.6}$$

is a multi-normal random vector with expectation given by d_{obs} and covariance matrix C_D.

The RML method generates a set of (m_{uc}, d_{uc}) pairs, and for each pair, calculates an associated "conditional realization" by minimizing

$$\begin{aligned} O_r(m) =& \frac{1}{2}(m - m_{\text{uc}})^T C_M^{-1}(m - m_{\text{uc}}) \\ &+ \frac{1}{2}(g(m) - d_{\text{uc}})^T C_D^{-1}(g(m) - d_{\text{uc}}). \end{aligned} \tag{2.7}$$

Some computational evidence indicates that the resulting set of conditional realizations represents an approximate sampling of the posterior pdf of Eq. 2.1 [27, 22] and that, in the context of reservoir modeling and simulation, the set of reservoir performance predictions generated from the set of conditional realizations can be used to obtain a reasonable assessment of the uncertainty in performance predictions [29, 28, 12]. As shown in the next section, RML samples the a posteriori distribution for m correctly if predicted data $g(m)$ is linearly related to m.

2.1.2 RML for the Linear Case

We consider the case where predicted data $g(m)$ is linearly related to the model, i.e.,

$$g(m) = g(m_0) + G(m - m_0) \tag{2.8}$$

where $G = G(m_0)$ is the $N_d \times N_m$ sensitivity matrix evaluated at the fixed model m_0 and is independent of m. Previous proofs that RML samples correctly assumed $g(m) = Gm$, but for the purpose of comparing RML with EnKF, it is more instructive to consider the general form of Eq. 2.8. Note that Eq. 2.8 represents the linearization of $g(m)$ about the model m_0 using a first order Taylor series. We will show that when this is true, RML provides a correct sampling of the posteriori pdf.

When Eq. 2.8 holds, the objective function given by Eq. 2.2 is quadratic. Thus, for any $\hat{m}$ the following second order Taylor series is exact:

$$O(m) = O(\hat{m}) + \big(\nabla O(\hat{m})\big)^T (m - \hat{m}) + \frac{1}{2}(m - \hat{m})^T H (m - \hat{m}). \tag{2.9}$$

In the preceding equation, H represents the Hessian matrix, which is given by

$$H = C_M^{-1} + G^T C_D^{-1} G. \tag{2.10}$$

Note that H is real-symmetric, positive-definite and independent of m. Thus, the objective function of Eq. 2.2 has a unique minimum, the MAP estimate, which can be obtained by setting $\nabla O(m) = 0$ and solving for m. Letting

$$r_0 = g(m_0) - d_{\text{obs}}, \tag{2.11}$$

the MAP estimate, denoted here by m_∞ can be written as

$$m_\infty = H^{-1}\Big(C_M^{-1} m_{\text{prior}} + G^T C_D^{-1}\big[Gm_0 - r_0\big]\Big) \tag{2.12}$$

Since the gradient of $O(m)$ evaluated at the MAP estimate is zero, the exact Taylor series

expansion of $O(m)$ about m_∞ ($\hat{m} = m_\infty$ in Eq. 2.9) gives

$$O(m) = O(m_\infty) + \frac{1}{2}(m - m_\infty)^T H(m - m_\infty). \tag{2.13}$$

Using the last result, we can write the a posteriori pdf as

$$f(m|d_{\text{obs}}) = a \exp\Big[-\frac{1}{2}(m - m_\infty)^T H(m - m_\infty)\Big], \tag{2.14}$$

where a is now the normalizing constant. Thus we have provided a very simple proof of a very known result that the posterior pdf is Gaussian with expectation (mean) m_∞ and covariance H^{-1}.

Next we show that in the linear case, RML samples correctly. Although the results of [26] show this is true, the proof given here follows one given in [29, 28]. When Eq. 2.8 holds, the conditional model which minimizes the objective function of Eq. 2.7 can be obtained by setting the gradient of $O(m)$ equal to zero and can be written as

$$m_c = H^{-1}\big(C_M^{-1} m_{\text{uc}} + G^T C_D^{-1}\big[Gm_0 - g(m_0) + d_{\text{uc}}\big]\big). \tag{2.15}$$

Note the random vector m_c represents the one defined by RML, that is, every model generated with RML is a realization of the random vector m_c . Using Eqs 2.4, 2.6 and 2.11 in Eq. 2.15 and rearranging

$$\begin{aligned} m_c &= H^{-1}\Big(C_M^{-1}\big[m_{\text{prior}} + L_M Z_M\big] + G^T C_D^{-1}\big[Gm_0 - g(m_0) + d_{\text{obs}} + L_D Z_D\big]\Big) \\ &= H^{-1}\Big(C_M^{-1}\big[m_{\text{prior}} + L_M Z_M\big] + G^T C_D^{-1}\big[Gm_0 - r_0 + L_D Z_D\big]\Big). \end{aligned} \tag{2.16}$$

Because m_c is a linear combination of the Gaussian random vectors Z_M and Z_D, m_c is Gaussian. Thus, to show RML samples correctly only requires that we show the expectation of m_c is m_∞ and its covariance matrix, denoted by C_c, is equal to H^{-1}. Using

Eq. 2.12 we can rewrite Eq. 2.16 as

$$m_c = m_\infty + H^{-1}\big(C_M^{-1} L_M Z_M + G^T C_D^{-1} L_D Z_D\big). \tag{2.17}$$

Throughout E denotes the expectation operator. Taking the expectation of Eq. 2.17 and recalling that Z_D and Z_M have zero vectors as expectations gives

$$E[m_c] = m_\infty. \tag{2.18}$$

It follows from Eqs. 2.18 and 2.17 that the covariance matrix for m_c is

$$\begin{aligned} C_c &= E\big[\big(m_c - m_\infty\big)\big(m_c - m_\infty\big)^T\big] \\ &= E\big[H^{-1}\big(C_M^{-1} L_M Z_M + G^T C_D^{-1} L_D Z_D\big)\big(H^{-1}\big(C_M^{-1} L_M Z_M + G^T C_D^{-1} L_D Z_D\big)\big)^T\big] \\ &= H^{-1} E\big[\big(C_M^{-1} L_M Z_M + G^T C_D^{-1} L_D Z_D\big)\big(Z_M^T L_M^T C_M^{-1} + Z_D^T L_D^T C_D^{-1} G\big)\big] H^{-1}. \end{aligned} \tag{2.19}$$

The N_m-dimensional vectors Y and W defined by

$$Y = C_M^{-1} L_M Z_M, \tag{2.20}$$

and

$$W = G^T C_D^{-1} L_D Z_D. \tag{2.21}$$

have zero means and are independent. Thus, the covariance matrices $E[YW^T]$ and $E[WY^T]$ are both equal to the $N_m \times N_m$ null matrix. Using this fact, together with Eqs. 2.3, 2.5, 2.20 and 2.21 and the facts that $E[Z_D Z_D^T] = I_{N_d}$ and $E[Z_M Z_M^T] = I_{N_m}$, it follows from Eq. 2.19 that

$$\begin{aligned} C_c &= H^{-1} E\big[\big(C_M^{-1} L_M Z_M Z_M^T L_M^T C_M^{-1}\big) + YW^T + WY^T + G^T C_D^{-1} L_D Z_D Z_D^T L_D^T C_D^{-1} G\big)\big] H^{-1} \\ &= H^{-1}\big[\big(C_M^{-1} L_M L_M^T C_M^{-1}\big) + \big(G^T C_D^{-1} L_D L_D^T C_D^{-1} G\big)\big] H^{-1} = H^{-1} H H^{-1} = H^{-1}. \end{aligned} \tag{2.22}$$

Thus for the linear case, we have shown that m_c has the same pdf as m, or equivalently,

for the linear case, RML provides a correct sampling of the posterior pdf.

2.1.3 Assimilation of Data by RML, Linear Case

If data measurement errors are uncorrelated in time and the relationship between data and the model is linear, one can match data sequentially in time and obtain identical results to those that would be obtained by matching all observed data simultaneously, see [30]. Suppose we have two sets of observed data d_{obs_1} and d_{obs_2}. We assume a linear relationship between data and model parameters. From Eqs. 2.10 to 2.12 for the first set of data we can write

$$H_1 = C_M^{-1} + G_1^T C_{D_1}^{-1} G_1, \tag{2.23}$$

$$r_{0_1} = g_1(m_0) - d_{\text{obs}_1}, \tag{2.24}$$

$$m_{\infty_1} = H_1^{-1}\Big(C_M^{-1} m_{\text{prior}} + G_1^T C_{D_1}^{-1}\big[G_1 m_0 - r_{0_1}\big]\Big). \tag{2.25}$$

For the conditional model after matching the first set of data, Eq. 2.17 gives

$$m_{c_1} = m_{\infty_1} + H_1^{-1}\big(C_M^{-1} L_M Z_M + G_1^T C_{D_1}^{-1} L_{D_1} Z_{D_1}\big). \tag{2.26}$$

From Eq. 2.22, the covariance of the random vector m_{c_1} is given by

$$C_{c_1} = H_1^{-1}. \tag{2.27}$$

Thus the pdf after conditioning to d_{obs_1} is

$$f(m|d_{\text{obs}_1}) = a \exp\Big[-\frac{1}{2}(m - m_{\infty_1})^T H_1 (m - m_{\infty_1})\Big], \tag{2.28}$$

where a is the normalizing factor. The pdf $f(m|d_{\text{obs}_1})$ is the prior model for the second data assimilation step, where m_{∞_1} is the prior mean and C_{c_1} is the covariance matrix. By generating a new set of unconditional realizations from the pdf of Eq. 2.28 we can rewrite

Eqs. 2.23 through 2.27, respectively for the second data assimilation step as

$$H_2 = C_{c_1}^{-1} + G_2^T C_{D_2}^{-1} G_2, \tag{2.29}$$

$$r_{0_2} = g_2(m_0) - d_{\text{obs}_2}, \tag{2.30}$$

$$m_{\infty_2} = H_2^{-1}\Big(C_{c_1}^{-1} m_{\infty_1} + G_2^T C_{D_2}^{-1}\big[G_2 m_0 - r_{0_2}\big]\Big), \tag{2.31}$$

$$m_{c_2} = m_{\infty_2} + H_2^{-1}\big(C_{c_1}^{-1} L_c Z_c + G_2^T C_{D_2}^{-1} L_{D_2} Z_{D_2}\big), \tag{2.32}$$

$$C_{c_2} = H_2^{-1}. \tag{2.33}$$

Now by substituting Eq. 2.25 into Eq. 2.31 and using Eq. 2.27, we have

$$\begin{aligned} m_{\infty_2} &= H_2^{-1}\Big(C_M^{-1} m_{\text{prior}} + G_1^T C_{D_1}^{-1}\big[G_1 m_0 - r_{0_1}\big] + G_2^T C_{D_2}^{-1}\big[G_2 m_0 - r_{0_2}\big]\Big) \\ &= H_2^{-1}\Big(C_M^{-1} m_{\text{prior}} + \begin{bmatrix} G_1^T & G_2^T \end{bmatrix} \begin{bmatrix} C_{D_1}^{-1} & O \\ O & C_{D_2}^{-1} \end{bmatrix} \Big(\begin{bmatrix} G_1 \\ G_2 \end{bmatrix} m_0 - \begin{bmatrix} r_{0_1} \\ r_{0_2} \end{bmatrix} \Big)\Big). \end{aligned} \tag{2.34}$$

By substituting Eq. 2.27 into Eq. 2.29 and using Eq. 2.23, we find that

$$\begin{aligned} H_2 &= C_M^{-1} + G_1^T C_{D_1}^{-1} G_1 + G_2^T C_{D_2}^{-1} G_2 \\ &= C_M^{-1} + \begin{bmatrix} G_1^T & G_2^T \end{bmatrix} \begin{bmatrix} C_{D_1}^{-1} & O \\ O & C_{D_2}^{-1} \end{bmatrix} \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \end{aligned} \tag{2.35}$$

We assume the observed data, d_{obs_1} and d_{obs_2}, measurement errors for the two sets are uncorrelated, therefore the covariance matrix of measurement errors is

$$C_D = \begin{bmatrix} C_{D_1} & O \\ O & C_{D_2} \end{bmatrix}, \tag{2.36}$$

and we can rewrite Eqs. 2.35 and 2.34, respectively as

$$H_2 = C_M^{-1} + G^T C_D^{-1} G = H, \tag{2.37}$$

$$m_{\infty_2} = H^{-1}\Big(C_M^{-1} m_{\text{prior}} + G^T C_D^{-1}\big[Gm_0 - r_0\big]\Big), \tag{2.38}$$

where Eqs. 2.37 and 2.38 respectively, are equivalent to Eqs. 2.10 and 2.12 obtained by matching all the data simultaneously. Here, it is important to note that after matching the first set of data it is not necessary to regenerate a new set of unconditional realizations; In other word, the m_{c_1} are automatically samples of $f(m|d_{\text{obs}_1})$ and are equivalent to $m_{\infty_1} + L_c Z_c$.

2.1.4 RML for Nonlinear Case

For the nonlinear case, the objective function of Eq. 2.7 can be minimized by the Gauss-Newton method. The Gauss-Newton method can be written in the form

$$m_j^{l+1} = m_{uc,j} + C_M G_{l,j}^T \big(C_D + G_{l,j} C_M G_{l,j}^T\big)^{-1} \big(G_{l,j}(m^l - m_{uc,j}) - g(m_j^l) + d_{uc,j}\big), \tag{2.39}$$

[30], where l is the iteration index and the subscript j represents the index of the realization. If we wish to generate N_e realizations conditional to an observed data vector, Eq. 2.39 is applied for $j = 1, 2 \cdots N_e$. Note each realization can be generated independently, the Gauss-Newton iteration process for m_j^{l+1} is independent of the iterative process for m_i^{l+1} for $i \neq j$. The matrix $G_{l,j}$ denotes the sensitivity matrix evaluated at the m_j^l and is given by

$$G_{l,j} = \begin{bmatrix} [\nabla g_1(m_j^l)]^T \\ [\nabla g_2(m_j^l)]^T \\ \vdots \\ [\nabla g_n(m_j^l)]^T \end{bmatrix}^T, \tag{2.40}$$

where g_i is the ith component of N_d-dimensional column vector of predicted data, $d = g(m)$, and ∇g_i is the gradient of g_i with respect to m.

2.2 Ensemble Kalman Filtering Method

2.2.1 EnKF from Gauss-Newton

We wish to write Eq. 2.39 in the form that is similar to the updating equation

used in the ensemble Kalman filter. For doing this, we suppose that the iterations for all ensembles are done concurrently and define the average model at iteration l by

$$\bar{m}^l = \frac{1}{N_e}\sum_{j=1}^{N_e} m_j^l, \tag{2.41}$$

where N_e is the number of ensembles.

If we do one iteration of the Gauss-Newton algorithm ($l = 0$) with the initial guess $m_j^0 = m_{uc,j}$ which we denote by m_j^p, and define $m_j^u = m_j^1$ as the updated realization, then Eq. 2.39 reduces to

$$m_j^u = m_j^p + C_M G_{0,j}^T \big(C_D + G_{0,j} C_M G_{0,j}^T\big)^{-1} \big(d_{uc,j} - g(m_j^p)\big). \tag{2.42}$$

As will be shown later, when suitable approximations are introduced, this equation can be reduced to the ensemble Kalman filter (EnKF) equation for updating the jth ensemble by assimilating data $d_{uc,j}$ for the purpose of sampling the a posteriori pdf for m conditional to data d_{obs} provided all observed data in d_{obs} corresponds to the same time. This means that the EnKF updating process is similar to doing one iteration of the Gauss-Newton method. To apply Eq. 2.42 still requires the sensitivity matrices $G_{0,j}$.

Instead of computing $G_{0,j}$ for $j = 1, 2, \cdots N_e$ by a method such as the adjoint method, we approximate all $G_{0,j}$ by the same matrix which is based on an average model. To do so, we define a mean model by

$$\bar{m}^p = \frac{1}{N_e}\sum_{j=1}^{N_e} m_j^p, \tag{2.43}$$

and let $\bar{G}$ denote the sensitivity matrix evaluated at $\bar{m}^p$. Then from Taylor's series, $g(m_j^p)$, the predicted data corresponding to m_j^p is given by

$$g(m_j^p) = g(\bar{m}^p) + \bar{G}(m_j^p - \bar{m}^p) + e_j \text{ for } j = 1, 2, \cdots N_e. \tag{2.44}$$

We estimate the average truncation error, $\bar{e}$, by

$$\bar{e} = \frac{1}{N_e} \sum_{j=1}^{N_e} e_j. \tag{2.45}$$

We estimate the mean of the predicted data by

$$\bar{g}^p = \frac{1}{N_e} \sum_{j=1}^{N_e} g(m_j^p). \tag{2.46}$$

Summing Eq. 2.44 over j, dividing the result by N_e and using Eqs. 2.45 and 2.44, we obtain

$$\bar{g}^p = g(\bar{m}^p) + \bar{e}. \tag{2.47}$$

Next, we approximate the covariance between m and $g(m)$ by using the approximate samples of m and $g(m)$, i.e., m_j^p and $g(m_j^p)$, for $j = 1, 2, \cdots N_e$. Before doing this, it is important to note that the underlying objective is to obtain a sampling of the posterior pdf for m conditional to d_{obs} and yet we are using the m_j^p's (unconditional realizations from the prior model) as an approximate sampling at least for the purpose of estimating the covariance between m and $g(m)$. The standard result for the estimation of this covariance is

$$\text{cov}(m, g(m)) = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(m_j^p - \bar{m}^p\right) \left(g(m_j^p) - \bar{g}^p\right)^T. \tag{2.48}$$

Using Eq. 2.44 in Eq. 2.48 gives

$$\text{cov}(m, g(m)) = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(m_j^p - \bar{m}^p\right) \left(\bar{G}(m_j^p - \bar{m}^p) + g(\bar{m}^p) - \bar{g}^p + e_j\right)^T, \tag{2.49}$$

or using Eq. 2.47 and simplifying,

$$\text{cov}(m, g(m)) = \left(\frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(m_j^p - \bar{m}^p\right) \left(m_j^p - \bar{m}^p\right)^T \bar{G}^T\right) + E^p, \tag{2.50}$$

where

$$E^p = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(m_j^p - \bar{m}^p\right)\left(e_j - \bar{e}\right)^T. \tag{2.51}$$

Because each e_j is a quadratic in m and the m_j^p's may be expected to oscillate about their mean $\bar{m}^p$, one might conjecture that E^p is sufficiently small so that it can be neglected but we have not been able to show this is true. If we could assume that m and e are independent random variables then the expectation of E^p would be a null matrix, but this assumption is not justified because the e_j's are directly related to the m_j^p's.

In any case, to obtain the updating formula for EnKF, we assume from this point on that E^p can be neglected so Eq. 2.50 reduces to

$$\mathrm{cov}(m, g(m)) = \Big(\frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(m_j^p - \bar{m}^p\right)\left(m_j^p - \bar{m}^p\right)^T\Big)\bar{G}^T. \tag{2.52}$$

As $N_e \to \infty$, Eq. 2.52 converges to $C_M \bar{G}^T$ so we write as an approximation

$$\mathrm{cov}(m, g(m)) \approx \Big(\frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(m_j^p - \bar{m}^p\right)\left(m_j^p - \bar{m}^p\right)^T\Big)\bar{G}^T \approx C_M \bar{G}^T. \tag{2.53}$$

Similarly, writing the standard estimator for the autocovariance of $g(m)$ and using Eqs. 2.44 and 2.47, we find that

$$\begin{aligned}
\mathrm{cov}(g(m), g(m)) &\approx \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(g(m_j^p) - \bar{g}^p\right)\left(g(m_j^p) - \bar{g}^p\right)^T \\
&= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(g(\bar{m}^p) + \bar{G}(m_j^p - \bar{m}^p) + e_j - \bar{g}^p\right)\left(g(\bar{m}^p) + \bar{G}(m_j^p - \bar{m}^p) + e_j - \bar{g}^p\right)^T \\
&= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(\bar{G}(m_j^p - \bar{m}^p) + e_j - \bar{e}\right)\left(\bar{G}(m_j^p - \bar{m}^p) + e_j - \bar{e}\right)^T \\
&= \left(\frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(\bar{G}(m_j^p - \bar{m}^p)\right)\left(m_j^p - \bar{m}^p\right)^T \bar{G}^T\right) + E_2^p,
\end{aligned} \tag{2.54}$$

where

$$E_2^p = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \Bigg[\bar{G}(m_j^p - \bar{m}^p)\big(e_j - \bar{e}\big)^T + \big(e_j - \bar{e}\big)\big(\bar{G}(m_j^p - \bar{m}^p)\big)^T + (e_j - \bar{e})(e_j - \bar{e})^T \Bigg]. \quad (2.55)$$

Similar to E^p, to proceed we must assume E_2^p is negligible even though, at this time, we can neither prove this not rigorously quantify the effect of doing so. Setting E_2^p equal to the null matrix of appropriate size in Eq. 2.54 gives

$$\mathrm{cov}(g(m), g(m)) \approx \bar{G} \left(\frac{1}{N_e - 1} \sum_{j=1}^{N_e} \big((m_j^p - \bar{m}^p)\big)\big(m_j^p - \bar{m}^p\big)^T \right) \bar{G}^T \approx \bar{G} C_M \bar{G}^T. \quad (2.56)$$

If for all j, we replace $G_{0,j}$ by $\bar{G}$ in Eq. 2.42 we obtain

$$m_j^u = m_j^p + C_M \bar{G}^T \big(C_D + \bar{G} C_M \bar{G}^T\big)^{-1} \big(d_{uc,j} - g(m_j^p)\big), \quad (2.57)$$

Now we can avoid computing $\bar{G}$ and also avoid explicit use of C_M by simply applying the approximations we have derived. Specifically, Eqs. 2.54–2.56 yield the approximation

$$\bar{G} C_M \bar{G}^T \approx \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \big(g(m_j^p) - \bar{g}^p\big)\big(g(m_j^p) - \bar{g}^p\big)^T, \quad (2.58)$$

and Eqs. 2.48–2.52 result in the following approximation:

$$C_M \bar{G}^T \approx \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \big(m_j^p - \bar{m}^p\big)\big(g(m_j^p) - \bar{g}^p\big)^T \Big). \quad (2.59)$$

Eqs. 2.56–2.59 provide the ensemble Kalman filter method for updating the ensembles or realizations m_j^p, $j = 1, 2, \cdots N_e$. Based on our derivation, if the observed data vector corresponds to a set of measurements at a single time, the EnKF is similar to doing one iteration of the Gauss-Newton method with a sensitivity coefficient equal to the sensitivity of data at the average model used to update all ensembles.

2.2.2 EnKF from RML

We consider both the model m and the predicted data, $g(m)$, as random vectors and define a random vector, y, as

$$y = \begin{bmatrix} m \\ g(m) \end{bmatrix}. \tag{2.60}$$

All vectors are column vectors with the dimension of m given by N_m and the dimension of $g(m)$ denoted by N_d. Thus, the dimension of y, denoted by N_y, is given by $N_y = N_m + N_d$. In the implementation of EnKF, $g(m)$ denotes data predicted at a specific time t_l corresponding to the vector of observed data we wish to assimilate. If measured data at all times were assimilated (matched) simultaneously, then the vector d_{obs} would include all measured data and $g(m)$ would represent the corresponding vector of predicted data. In reservoir characterization problems of interest to us, m represents reservoir variables such as gridblock porosities and permeabilities or log-permeabilities and $g(m)$ is obtained by running a simulator with the model m as input.

A critical step in the EnKF is that the assumption that the pdf for y can be approximated by the multivariate Gaussian distribution given by

$$f(y) = a \exp\Big(-\frac{1}{2}\big(y - \bar{y}\big)^T C_Y^{-1}\big(y - \bar{y}\big)\Big), \tag{2.61}$$

where a is the normalizing constant, $\bar{y}$ is the mean and C_Y is the covariance matrix for y given by

$$C_Y = \begin{bmatrix} C_M & C_{M,G} \\ C_{G,M} & C_G \end{bmatrix}. \tag{2.62}$$

Throughout, C_M is the autocovariance matrix for m, C_G is the autocovariance matrix for $g(m)$ and $C_{M,G}$ and $C_{G,M}$ are the cross covariance matrices for m and $g(m)$, i.e.,

$$C_M = E\big[\big(m - \bar{m}\big)\big(m - \bar{m}\big)^T\big], \tag{2.63}$$

$$C_G = E\big[\big(g(m) - \overline{g(m)}\big)\big(g(m) - \overline{g(m)}\big)^T\big], \tag{2.64}$$

$$C_{M,G} = E\big[\big(m - \bar{m}\big)\big(g(m) - \overline{g(m)}\big)^T\big], \tag{2.65}$$

If measurement error plus modelling error is a Gaussian random vector with mean given by the N_d-dimensional zero vector and covariance given by C_D, a standard application of Bayes' theorem gives that the posterior pdf for y conditional to d_{obs} is given by

$$\begin{aligned} f(y|d_{\text{obs}}) = a \exp\Big(-\frac{1}{2}\big(g(m) - d_{\text{obs}}\big)^T C_D^{-1}\big(g(m) - d_{\text{obs}}\big)\Big) \\ \exp\Big(-\frac{1}{2}\big(y - \bar{y}\big)^T C_Y^{-1}\big(y - \bar{y}\big)\Big). \end{aligned} \tag{2.66}$$

where a is now the normalizing constant for the conditional pdf $f(y|d_{\text{obs}})$. Recall that $g(m)$ and d_{obs} are N_d-dimensional column vectors. Let I_{N_d} denote the $N_d \times N_d$ identity matrix, O denote the $N_d \times N_m$ null matrix and define the $N_d \times N_y$ matrix H by

$$H = \begin{bmatrix} O & I_{N_d} \end{bmatrix}. \tag{2.67}$$

It follows from Eq. 2.60 that data predicted for a given y is linearly related to y since predicted data is given by

$$g(m) = Hy, \tag{2.68}$$

and Eq. 2.66 can be written as

$$\begin{aligned} f(y|d_{\text{obs}}) = a \exp\Big(-\frac{1}{2}\big(Hy - d_{\text{obs}}\big)^T C_D^{-1}\big(Hy - d_{\text{obs}}\big)\Big) \\ \exp\Big(-\frac{1}{2}\big(y - \bar{y}\big)^T C_Y^{-1}\big(y - \bar{y}\big)\Big). \end{aligned} \tag{2.69}$$

It is well known [26] that the randomized maximum likelihood method, RML, can be applied to obtain a correct sampling of the posterior pdf, $f(y|d_{\text{obs}})$ given by Eq. 2.69.

To generate one realization using RML, we generate a set of N_e unconditional realizations of y from the Gaussian pdf of Eq. 2.61. Such an unconditional realization, $y_{uc,j}$, can be obtained by generating an N_y-dimensional vector Z_j of independent standard

random normal deviates and calculating

$$y_{uc,j} = \bar{y} + L_y Z_j, \tag{2.70}$$

where

$$C_Y = L_y L_y^T, \tag{2.71}$$

is the Cholesky decomposition of C_Y. For each unconditional realization $m_{uc,j}$, $1 \leq j \leq N_e$, we generate a set of realization of the data, $d_{uc,j}$ by adding noise to the data i.e.,

$$d_{uc,j} = d_{\text{obs}} + L_D Z_{D,j}, \tag{2.72}$$

where $Z_{D,j}$ is a vector of standard random normal deviates and $L_D L_D^T$ is the Cholesky decomposition of the measurement error covariance matrix, C_D. For each j, we minimize

$$\begin{aligned} O_j(y) = &\frac{1}{2}\big(Hy - d_{uc,j}\big)^T C_D^{-1}\big(Hy - d_{uc,j}\big) \\ &+ \frac{1}{2}\big(y - y_{uc,j}\big)^T C_Y^{-1}\big(y - y_{uc,j}\big). \end{aligned} \tag{2.73}$$

Then it is well known that the set of minima, y_j^u, $1 \leq j \leq N_e$ represents a set of samples from the conditional pdf $f(y|d_{\text{obs}})$. The minimum of $O_j(y)$ can be found by setting its gradient with respect to y equal to the N_y-dimensional zero vector and solving for y. Denoting this solution by y_j^u, we find that

$$y_j^u = y_{uc,j} + C_Y H^T \Big(C_D + H C_Y H^T\Big)^{-1} \Big(d_{uc,j} - H y_{uc,j}\Big). \tag{2.74}$$

The preceding equation is effectively the basic equation for generating a suite of realizations using EnKF except $y_{uc,j}$ is generated as data are assimilated sequentially in time, the data part of each $y_{uc,j}$ is generated by using prediction from the forward model and C_Y is generated by an approximate method.

Gu and Oliver [14] were able to relate the MAP estimate to the ensemble Kalman filter. Here, we extend their result to show that if data are linearly related to the model,

the prior model is multivariate Gaussian, and data are uncorrelated in time, then the generation of ensembles using EnKF is equivalent to sampling with randomized maximum likelihood [27] as the number of ensembles becomes infinite.

Next we establish conditions under which the preceding procedure is equivalent to using RML to sample the pdf $f(m|d_{obs})$, the conditional pdf for m given d_{obs}. We consider the case where (i) the prior model for m is Gaussian with mean m_{prior} and covariance matrix C_M and (ii) predicted data is linearly related to the model by

$$g(m) = g(m_0) + G(m - m_0). \tag{2.75}$$

From Eq. 2.75, the expectation of $g(m)$ is given by $g(m_{\text{prior}})$ so

$$\bar{y} = \begin{bmatrix} m_{\text{prior}} \\ g(m_{\text{prior}}) \end{bmatrix}. \tag{2.76}$$

and Eq. 2.62 can be rewritten as

$$C_Y = \begin{bmatrix} C_M & C_M G^T \\ GC_M & GC_M G^T \end{bmatrix}. \tag{2.77}$$

Letting $C_M = L_M L_M^T$ be the Cholesky decomposition of C_M, and defining the block lower triangular matrix L_c by

$$L_c = \begin{bmatrix} L_M & O \\ GL_M & O \end{bmatrix}, \tag{2.78}$$

where the O's denote null matrices, it follows easily from Eq. 2.77 that $L_c L_c^T = C_Y$, thus, $L_c = L_y$. Using this result and Eq. 2.76 in Eq. 2.70 gives

$$y_{uc,j} = \begin{bmatrix} m_{uc,j} \\ g(m)_{uc,j} \end{bmatrix} = \begin{bmatrix} m_{\text{prior}} + L_M Z_{M,j} \\ g\big(m_{\text{prior}} + L_M Z_{M,j}\big) \end{bmatrix}, \tag{2.79}$$

where $Z_{M,j}$ denotes the N_m dimensional vector with entries identical to the corresponding

components of the first N_m dimensional entries of Z_j. The vector consisting of the first N_m components of $y_{uc,j}$ pertains to the model m and is given by

$$m_{uc,j} = m_{\text{prior}} + L_M Z_{M,j}. \tag{2.80}$$

which represents a correct sampling of the prior Gaussian distribution for m. From Eqs. 2.67, 2.77 and 2.74, it is straightforward to show that the model part of y_j^u is given by

$$m_j^u = m_{uc,j} + C_M G^T \Big(C_D + G C_M G^T \Big)^{-1} \Big(d_{uc,j} - g(m_{uc,j}) \Big). \tag{2.81}$$

Eqs. 2.80 and 2.81 imply that the m_j^u's extracted from the y_j^u's are equivalent to applying RML to sample $f(m|d_{obs})$ when predicted data is linearly related to m and the prior pdf for m is $N(m_{\text{prior}}, C_M)$, i.e., a normal distribution with expectation m_{prior} and covariance C_M. When the prior model is $N(m_{\text{prior}}, C_M)$, $f(m|d_{obs})$ is given by

$$\begin{aligned} f(m|d_{\text{obs}}) = a \exp\Big(-\frac{1}{2} \big(g(m) - d_{\text{obs}}\big)^T C_D^{-1} \big(g(m) - d_{\text{obs}}\big) \Big) \\ \exp\Big(-\frac{1}{2} \big(m - m_{\text{prior}}\big)^T C_M^{-1} \big(m - m_{\text{prior}}\big) \Big), \end{aligned} \tag{2.82}$$

where in the linear case under consideration, $g(m) = g(m_0) + G(m - m_0)$.

In the case where d_{obs} contains all dynamic data we wish to assimilate, then for the purposes of uncertainty evaluation in the model and future predictions using the forward model, our objective is to sample from $f(m|d_{\text{obs}})$. To summarize our results, we have shown that the set of y_j^u's (Eq. 2.74) generated by RML always represents a correct sampling of the pdf of Eq. 2.69 or equivalently Eq. 2.66. When the conditional pdf $f(m|d_{\text{obs}})$ is given by Eq. 2.82 and data are linearly relted to the model, then the model part of the y_j^u's represents a correct sampling of $f(m|d_{\text{obs}})$. If in addition the

observed data, d_{obs}, can be written as

$$d_{\text{obs}} = \begin{bmatrix} d_{\text{obs}}(t_1) \\ d_{\text{obs}}(t_2) \\ \vdots \\ d_{\text{obs}}(t_n) \end{bmatrix}, \tag{2.83}$$

where $d_{\text{obs}}(t_\ell)$ represents data measured at time t_ℓ and $t_1 < t_2, \cdots, < t_n$, and individual data are uncorrelated in time, then generating the conditional mean of m (the MAP estimate) by minimization is equivalent to generating a sequence of MAP estimates where at the ℓth step in the sequence, we minimize an objective function which includes only the data mismatch between $d_{obs,\ell}$ and the corresponding predicted data [30]. After each minimization, the covariance matrix must be updated to the a posteriori covariance matrix, $C_{M,\ell}$. The MAP estimate based on assimilating $d_{obs,\ell}$ and the covariance matrix $C_{M,\ell}$ effectively provide the prior Gaussian model for the assimilation of the next set of data, $d_{obs,\ell+1}$. Using these ideas, it can be shown that RML can be used to assimilate data sequentially in time and that the conditional realizations obtained at the final time step t_n represents a correct sampling of $f(m|d_{\text{obs}})$ provided again that (i) the prior pdf for m used to assimilate data at the first time t_1 is $N(m_{prior}, C_M)$, (ii) $g(m) = g(m_0)+G(m-m_0)$ and (iii) data measurement errors are uncorrelated in time and satisfy Gaussian distributions with zero means. Under these conditions, application of Eq. 2.74 sequentially in time results in a set of y_j^u's after assimilation at t_n which reprsent a correct sampling of the pdf in Eq. 2.69 and the model parts of the y_j^u's represent a correct sampling of $f(m|d_{obs})$ where, d_{obs} is given by Eq. 2.83.

From this point on we assume that Eq. 2.74 is applied sequentially in time so that $d_{uc,j}$ represents data at some time t_ℓ where our vector of observed data is $d_{\text{obs}}(t_\ell)$. For $1 \leq j \leq N_e$, we let m_j^p denote the model part of the ensemble vector obtained by assimilating data at $t_{\ell-1}$. Letting $g(m)$ now denote predicted data corresponding to

$d_{obs} = d_{\text{obs}}(t_\ell)$, it makes sense to replace $y_{uc,j}$ in Eq. 2.74 by

$$y_j^p = \begin{bmatrix} m_j^p \\ g(m_j^p) \end{bmatrix}, \tag{2.84}$$

so that Eq. 2.74 becomes

$$y_j^u = y_j^p + C_Y H^T \Big(C_D + H C_Y H^T \Big)^{-1} \Big(d_{uc,j} - H y_j^p \Big), \tag{2.85}$$

for $1 \leq j \leq N_e$ we also replace $\bar{y}$ in the prior model by the average of the y_j^p's. This is the natural analogue of the procedure we established in the linear case. If we also estimate the covariance C_Y by

$$C_Y = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (y_j^p - \bar{y})(y_j^p - \bar{y})^T, \tag{2.86}$$

then Eq. 2.86 is mathematically equivalent to the basic EnKF algorithm for assimilating data. The approximation of Eq. 2.86 may be rank deficient and so even in the linear case, we can only show that EnKF is equivalent to RML and hence gives a correct sampling as $N_e \to \infty$. Similarly, we can approximate the covariance submatrices that appear in Eq. 2.62 by the following procedure, here we drop the superscript p on the right hand side of Eq. 2.85 and keep in mind all the matrices on the right hand side belongs to the prediction period. To simplify the equations, we define the $N_y \times N_e$ matrix of all ensembles Y by

$$Y = \begin{bmatrix} y_1 & y_2 & \dots & y_{N_e}, \end{bmatrix} \tag{2.87}$$

and the $N_y \times N_e$ matrix $\overline{Y}$, as the matrix with all columns equal to $\overline{y}$ given by

$$\overline{y} = \frac{1}{N_e} \sum_{j=1}^{N_e} y_j, \tag{2.88}$$

Then, we define the matrix ΔY as

$$\begin{aligned}\Delta Y &= Y - \overline{Y} \\ &= \begin{bmatrix} y_1 - \overline{y} & y_2 - \overline{y} & \dots & y_{N_e} - \overline{y} \end{bmatrix}, \\ &= \begin{bmatrix} m_1 - \overline{m} & \dots & m_{N_e} - \overline{m} \\ d_1 - \overline{d} & \dots & d_{N_e} - \overline{d} \end{bmatrix}\end{aligned} \tag{2.89}$$

where d_j is the predicted data for jth ensemble, $d_j = g(m_j)$, and $\overline{d}$ is the average of predicted data.

From Eq. 2.89, we write an approximation of C_Y as

$$\begin{aligned}C_Y = \begin{bmatrix} C_M & C_{M,G} \\ C_{G,M} & C_G \end{bmatrix} &\approx \frac{1}{N_e - 1}(Y - \overline{Y})(Y - \overline{Y})^T \\ &= \frac{1}{N_e - 1}\sum_{j=1}^{N_e}(y_j - \overline{y})(y_j - \overline{y})^T.\end{aligned} \tag{2.90}$$

To avoid calculating the $N_y \times N_e$ matrix C_Y, which may be extremely large, we define the matrix A by $A = \Delta Y^T H^T$ so

$$\begin{aligned}A^T &= H\Delta Y \\ &= \begin{bmatrix} O_{N_d \times N_m} & | & I_{N_d \times N_d} \end{bmatrix} \Delta Y \\ &= \begin{bmatrix} d_1 - \overline{d} & \dots & d_{N_e} - \overline{d} \end{bmatrix},\end{aligned} \tag{2.91}$$

represents the rows of ΔY corresponding to predicted data. Using the approximation of Eqs. 2.90 and 2.91, we may approximate Eq. 2.85 as

$$y_j^u = y_j^p + \frac{1}{N_e - 1}\Delta Y A\Big(C_D + \frac{A^T A}{N_e - 1}\Big)^{-1}\Big(d_{uc,j} - H y_j^p\Big), \tag{2.92}$$

where

$$\frac{1}{N_e-1}A^TA = \frac{1}{N_e-1}\begin{bmatrix} d_1-\overline{d} & \dots & d_{N_e}-\overline{d}\end{bmatrix}\begin{bmatrix}(d_1-\overline{d})^T\\ \vdots \\ (d_{N_e}-\overline{d})^T\end{bmatrix} \\ = \frac{1}{N_e-1}\sum_{j=1}^{N_e}(d_j-\overline{d})(d_j-\overline{d})^T, \tag{2.93}$$

and

$$\frac{1}{N_e-1}\Delta YA = \frac{1}{N_e-1}\begin{bmatrix} y_1-\overline{y} & \dots & y_{N_e}-\overline{y}\end{bmatrix}\begin{bmatrix}(d_1-\overline{d})^T\\ \vdots \\ (d_{N_e}-\overline{d})^T\end{bmatrix} \\ = \frac{1}{N_e-1}\begin{bmatrix} m_1-\overline{m} & \dots & m_{N_e}-\overline{m} \\ d_1-\overline{d} & \dots & d_{N_e}-\overline{d}\end{bmatrix}\begin{bmatrix}(d_1-\overline{d})^T\\ \vdots \\ (d_{N_e}-\overline{d})^T\end{bmatrix} \\ = \begin{bmatrix}\frac{1}{N_e-1}\sum_{j=1}^{N_e}(m_j-\overline{m})(d_j-\overline{d})^T \\ \frac{1}{N_e-1}\sum_{j=1}^{N_e}(d_j-\overline{d})(d_j-\overline{d})^T\end{bmatrix}. \tag{2.94}$$

Finally, by substituting Eqs. 2.93 and 2.94 into Eq. 2.92,

$$y_j^u = y_j^p + \begin{bmatrix}\frac{1}{N_e-1}\sum_{j=1}^{N_e}(m_j-\overline{m})(d_j-\overline{d})^T \\ \frac{1}{N_e-1}\sum_{j=1}^{N_e}(d_j-\overline{d})(d_j-\overline{d})^T\end{bmatrix} \\ \times\left(\left[\frac{1}{N_e-1}\sum_{j=1}^{N_e}(d_j-\overline{d})(d_j-\overline{d})^T\right]+C_D\right)^{-1}(d_{uc,j}-Hy_j^p). \tag{2.95}$$

2.3 EnKF from Multiple Linear Regression

2.3.1 Multiple Linear Regression

Multiple linear regression is based on a linear relationship between a main variable m and a set of auxiliary variables z_i. The multiple linear regression estimate, m^{est}, is

given by

$$m^{\text{est}} = \mu + \sum_{i=1}^{N_d} a_i(z_i - \zeta_i), \tag{2.96}$$

where N_d is the number of auxiliary variables, μ is the mean of the main variable m and ζ_i is the mean of the auxiliary variable z_i. In vector form,

$$m^{\text{est}} = \mu + a^T(z - \zeta). \tag{2.97}$$

Note if m is a random vector, a will be the matrix of coefficients, where the jth column of a represents the weights used to estimate jth entry of the vector m. Now we define the estimation error as

$$e = m - m^{\text{est}}. \tag{2.98}$$

The expectation of the estimation error is equal to zero (null vector for the case where m is a vector). Since the expectation of the estimation error is equal to zero, the mean square error and the covariance of estimation error are identical. Therefore, to minimize the mean square error we inscribe the covariance of the estimation error as

$$\begin{aligned}
\text{cov}(e,e) &= E\Big[\big(m - m^{\text{est}}\big)\big(m - m^{\text{est}}\big)^T\Big] \\
&= E\Big[\big(m - \mu - a^T(z-\zeta)\big)\big(m - \mu - a^T(z-\zeta)\big)^T\Big] \\
&= E\Big[(m-\mu)(m-\mu)^T - a^T(z-\zeta)(m-\mu)^T \\
&\qquad - (m-\mu)(z-\zeta)^T a + a^T(z-\zeta)(z-\zeta)^T a\Big] \\
&= C_M - a^T C_{ZM} - C_{MZ}a + a^T C_Z a,
\end{aligned} \tag{2.99}$$

where C_M and C_Z are the covariance matrices of m and z respectively and C_{MZ} is the cross-covariance matrix between m and z. Using the fact that the transpose of C_{ZM} is equal to C_{MZ} and taking partial derivatives with respect to a_i yields

$$\nabla_a\big(\text{cov}(e,e)\big) = -2C_{MZ}^T + 2C_Z a. \tag{2.100}$$

Therefore the minimum of the mean square error is obtained when

$$a = C_Z^{-1} C_{MZ}^T. \tag{2.101}$$

Using a given by Eq. 2.101 in Eq. 2.97 gives the best linear estimate of the model parameters. Note Eq. 2.101 is also the well known simple cokriging system of equations; see [31]. By substituting Eq. 2.101 into Eq. 2.97 we obtain

$$m^{\text{est}} = \mu + C_{MZ} C_Z^{-1} (z - \zeta). \tag{2.102}$$

Using Eq. 2.101 in Eq. 2.99 to compute the covariance matrix of the estimation error, we find that

$$\begin{aligned} \text{cov}(e, e) &= C_M - C_{MZ} C_Z^{-1} C_{MZ}^T - C_{MZ} C_Z^{-1} C_{MZ}^T + C_{MZ} C_Z^{-1} C_Z C_Z^{-1} C_{MZ}^T \\ &= C_M - C_{MZ} C_Z^{-1} C_{MZ}^T. \end{aligned} \tag{2.103}$$

Now we define the vector ϵ as the measurement error with mean equal to a null vector and covariance matrix C_D. Suppose we can always write vector of observations, z, as

$$z = d + \epsilon \tag{2.104}$$

where $d = g(m)$ is a function of the random vector m. Assuming the measurement errors are independent of d and m, we can rewrite the covariance matrices as

$$C_Z = \text{cov}(d + \epsilon, d + \epsilon) = \text{cov}(d, d) + C_D = C_{DD} + C_D \tag{2.105}$$

and

$$C_{MZ} = \text{cov}(m, d + \epsilon) = \text{cov}(m, d) = C_{MD}, \tag{2.106}$$

then Eq. 2.102 becomes

$$m^{\text{est}} = \mu + C_{MD} \big(C_{DD} + C_D \big)^{-1} (d + \epsilon - \delta). \tag{2.107}$$

where δ is the mean of the random vector d. Note, because the mean of the vector ϵ is a null vector, δ is formally equal to ζ. Now for a given set of observations, d_{obs}, as a realization of the random vector z (or $d+\epsilon$) and a set of model parameters with mean equal to m_{prior} and covariance matrix C_M, we can write the best linear estimate (Eq. 2.107) of the model parameters as

$$m^{\text{est}} = m_{\text{prior}} + C_{MD}\big(C_{DD} + C_D\big)^{-1}(d_{\text{obs}} - \delta). \tag{2.108}$$

2.3.2 An Ad hoc Approximate Sampling Method

The estimate calculated from the Eq. 2.108 is a smooth estimate of the model m. Here, we attempt to sample from the probability distribution of the estimate model parameters to construct rougher estimate of the model parameters. To attempt to sample the estimate, we use a method which is very similar to randomized maximum likelihood method. First, we generate a set of N_e realizations of the model parameters from

$$m_j = m_{\text{prior}} + L_M Z_{M_j}, \tag{2.109}$$

where L_M is a square root of the C_M and Z_M is a vector of independent normal deviates with zero mean and unit variance. In practice, if the number of model parameters is large we usually use a sequential Gaussian simulation program to generate the realizations. Then we generate a set of N_e realizations of the observed data as

$$d_{uc,j} = d_{\text{obs}} + L_D Z_{D_j}, \tag{2.110}$$

where L_D is a square root of the C_D and Z_D is a vector of independent normal deviates with zero mean and unit variance. Note this algorithm is valid for a Gaussian measurement error and a Gaussian prior model. To generate realizations of the data part, instead of sampling the probability distribution of data with mean δ and covariance matrix C_{DD}, we generate a set of N_e realizations of the data part by running the model m_j to sample $d_j = g(m_j)$. Therefore, for each realization from Eq. 2.108 we have an updated model

defined by

$$m_j^u = m_j + C_{MD}\big(C_{DD} + C_D\big)^{-1}(d_{uc,j} - d_j). \tag{2.111}$$

Although Eqs. 2.107 and 2.108 were theoretically derived, Eq. 2.111 is not; it is an ad hoc result based on the similar intuitive motivation that resulted in the randomized maximum likelihood (RML) method. A rough intuitive motivation follows. By replacing m_{prior} by $m_{uc,j}$, we replace the underlying prior pdf by one which now has expection equal to $m_{uc,j}$. By replacing δ by $d_j = g(m_{uc,j})$, we are effectively assuming that the mean of the set of predicted data generated from a set of models which represents a sampling of the prior pdf can be estimated by the data predicted using the mean (now assumed to be $m_{uc,j}$) of the prior distribution. Why noise should be added to d_{obs} seems less clear except that in RML case for simple cases we can show that it is necessary. Although the preceding is a less than compelling explanation, we can show that in the case, where the data is linear related to the model, the random variable m_j^u given in Eq. 2.111 has the correct expectation and covariance.

For a linear problem

$$d = Gm \tag{2.112}$$

using the fact that $E[Z_M] = 0$ and $E[Z_D] = 0$ the expectation of the updated models is equal to

$$\begin{aligned} E[m^u] &= E\big[m_{\text{prior}} + L_M Z_M + C_{MD}\big(C_{DD} + C_D\big)^{-1}\big(d_{\text{obs}} + L_D Z_D - G(m_{\text{prior}} + L_M Z_M)\big)\big] \\ &= m_{\text{prior}} + C_{MD}\big(C_{DD} + C_D\big)^{-1}(d_{\text{obs}} - Gm_{\text{prior}}). \end{aligned} \tag{2.113}$$

After substitution δ for Gm_{prior}, we obtain

$$E[m^u] = m^{\text{est}}. \tag{2.114}$$

Therefore, the covariance matrix of the updated models, using the fact that for the linear

problem, $C_M G^T = C_{MD}$ and $G C_M G^T = C_{DD}$, is equal to

$$\begin{aligned} E[(m^u - m^{\text{est}})(m^u - m^{\text{est}})^T] = \\ & E\Big[\Big(L_M Z_M + C_{MD}(C_{DD} + C_D)^{-1}(L_D Z_D - G L_M Z_M)\Big) \\ & \cdot \Big(L_M Z_M + C_{MD}(C_{DD} + C_D)^{-1}(L_D Z_D - G L_M Z_M)\Big)^T\Big] \\ & = C_M - C_{MD}(C_{DD} + C_D)^{-1} C_{MD}^T, \end{aligned} \tag{2.115}$$

where from Eqs. 2.105, 2.105 and 2.106, the covariance matrix of the updated models is equal to the covariance matrix of the estimation error. Therefore in our sampling procedure for a linear problem we can preserve the expectation and the covariance of the original estimate as the best linear estimate.

2.3.3 EnKF motivation

So far we have not discussed the way that we should calculate the matrices C_{MD} and C_{DD}. By definition

$$C_{MD} = E[(m - \mu)(d - \delta)^T], \tag{2.116}$$

and

$$C_{DD} = E[(d - \delta)(d - \delta)^T]. \tag{2.117}$$

By using the idea of ensemble Kalman filtering we can approximate these matrices by

$$C_{MD} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (m_j - \overline{m})(d_j - \overline{d})^T, \tag{2.118}$$

and

$$C_{DD} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (d_j - \overline{d})(d_j - \overline{d})^T, \tag{2.119}$$

where $d_j = g(m_j)$ and $\overline{d}$ is given by

$$\overline{d} = \frac{1}{N_e}\sum_{j=1}^{N_e} g(m_j). \tag{2.120}$$

Now we can write Eq. 2.111 as an approximation by

$$m_j^u = m_j + \Big(\frac{1}{N_e-1}\sum_{j=1}^{N_e}(m_j-\overline{m})(d_j-\overline{d})^T\Big)\Big(\frac{1}{N_e-1}\sum_{j=1}^{N_e}(d_j-\overline{d})(d_j-\overline{d})^T+C_D\Big)^{-1}(d_{uc,j}-d_j). \tag{2.121}$$

The preceding equation is the update equation of the ensemble Kalman filtering method for assimilation of observed data at each assimilation time step. Note that beside the Gaussian assumption for the measurement error we did not make any other important Gaussian assumption to derive Eq. 2.121. The assumption that the initial samples of the model parameters were drawn from a multi-Gaussian distribution may not be a crucial assumption for the case where we assimilate the data sequentially. It means, once we start to assimilate the data the updated model from the previous data assimilation step construct the set of initial samples for the next data assimilation step. This idea is well described in Bayesian framework by [10].

2.4 Updating of Time Dependent Parameters in EnKF

In the presence of time dependent parameters, the state vector for jth ensemble will be defined by

$$y_j = \begin{bmatrix} m_j \\ p_j \\ d_j \end{bmatrix} = \begin{bmatrix} m_j^T & p_j^T & d_j^T \end{bmatrix}^T, \tag{2.122}$$

where p is a N_p-dimensional column vector and N_p is the number of time dependent parameters for each ensemble. For a black oil reservoir simulator, p may include pressure, saturation and dissolved GOR. Here, we rewrite Eq. 2.85 in terms of matrix of all

ensembles,

$$Y = \begin{bmatrix} M \\ P \\ D \end{bmatrix}, \tag{2.123}$$

where M is an $N_m \times N_e$ matrix with the jth column equal to m_j, P is an $N_p \times N_e$ matrix with the jth column equal to p_j and D is an $N_d \times N_e$ matrix with the jth column equal to d_j, predicted data from the jth ensemble. In terms of matrix of ensembles, we can now rewrite Eq. 2.85 as

$$Y^u = Y + C_Y H^T (H C_Y H^T + C_D)^{-1} (D_{uc} - HY), \tag{2.124}$$

where D_{uc} is an $N_d \times N_e$ matrix with the jth column equal to $d_{uc,j}$. As in Eq. 2.92, we approximate Eq. 2.124 as

$$\begin{aligned} Y^u &= Y + \frac{(Y - \overline{Y})(Y - \overline{Y})^T}{N_e - 1} H^T \Big[H \frac{(Y - \overline{Y})(Y - \overline{Y})^T}{N_e - 1} H^T + C_D \Big]^{-1} (D_{uc} - HY) \\ &= Y + \frac{(Y - \overline{Y})(D - \overline{D})^T}{N_e - 1} \Big[\frac{(D - \overline{D})(D - \overline{D})^T}{N_e - 1} + C_D \Big]^{-1} (D_{uc} - D) \end{aligned} . \tag{2.125}$$

It is easy to show that $\overline{Y}(D - \overline{D})^T$ is equal to a null matrix because

$$\begin{aligned} \overline{Y}(D - \overline{D})^T &= \begin{bmatrix} \overline{y} & \overline{y} & \dots & \overline{y} \end{bmatrix} \begin{bmatrix} (d_1 - \overline{d})^T \\ \vdots \\ (d_{N_e} - \overline{d})^T \end{bmatrix} \\ &= \sum_{j=1}^{N_e} \overline{y} (d_j - \overline{d})^T \\ &= \overline{y} \sum_{j=1}^{N_e} (d_j - \overline{d})^T = 0. \end{aligned} \tag{2.126}$$

Therefore

$$\begin{aligned} Y^u &= Y\Big(I + \frac{(D-\overline{D})^T}{N_e-1}\Big[\frac{(D-\overline{D})(D-\overline{D})^T}{N_e-1} + C_D\Big]^{-1}(D_{uc}-D)\Big), \\ &= Y\big(I+\delta(D)\big) \end{aligned} \tag{2.127}$$

where $\delta(D)$ is defined by

$$\delta(D) = \frac{(D-\overline{D})^T}{N_e-1}\Big[\frac{(D-\overline{D})(D-\overline{D})^T}{N_e-1} + C_D\Big]^{-1}(D_{uc}-D) \tag{2.128}$$

Note that parts of Y, i.e., M, P and D, all get updated using the same coefficient matrix, $\big(I+\delta(D)\big)$, and Eq. 2.127 indicates the updating coefficient matrix is only a function of the matrix D.

2.4.1 EnKF Procedure and Implementation

Fig. 2.1 shows the general procedure that one should use for any type of model.

Eq. 2.127 shows the best way to implement the ensemble Kalman filter procedure:

- Generation of matrix D; for each ensemble the forward model will be run and the jth column of matrix D is equal to the predicted data from the jth ensemble.
- Calculation of vector $\overline{d}$, vector of average predicted data; this vector represents the whole matrix $\overline{D}$, because all the columns of matrix $\overline{D}$ are equal to $\overline{d}$.
- Generation of matrix D_{uc}, matrix of perturbed observations;
- Calculation of $\delta(D)$ term; for the inversion part we used singular value decomposition method from the LAPACK routine. (in our implementation, we use significant singular values up to the point that the summation of the significant singular values is less than 99.99% of the summation of all singular values.)
- Update the matrix Y; for this part if the number of model parameters is too large we can update the model parameters partially, e.g., we can update the porosity of all gridblocks ,store them on the disk and then update the permeability of all gridblocks and so on.

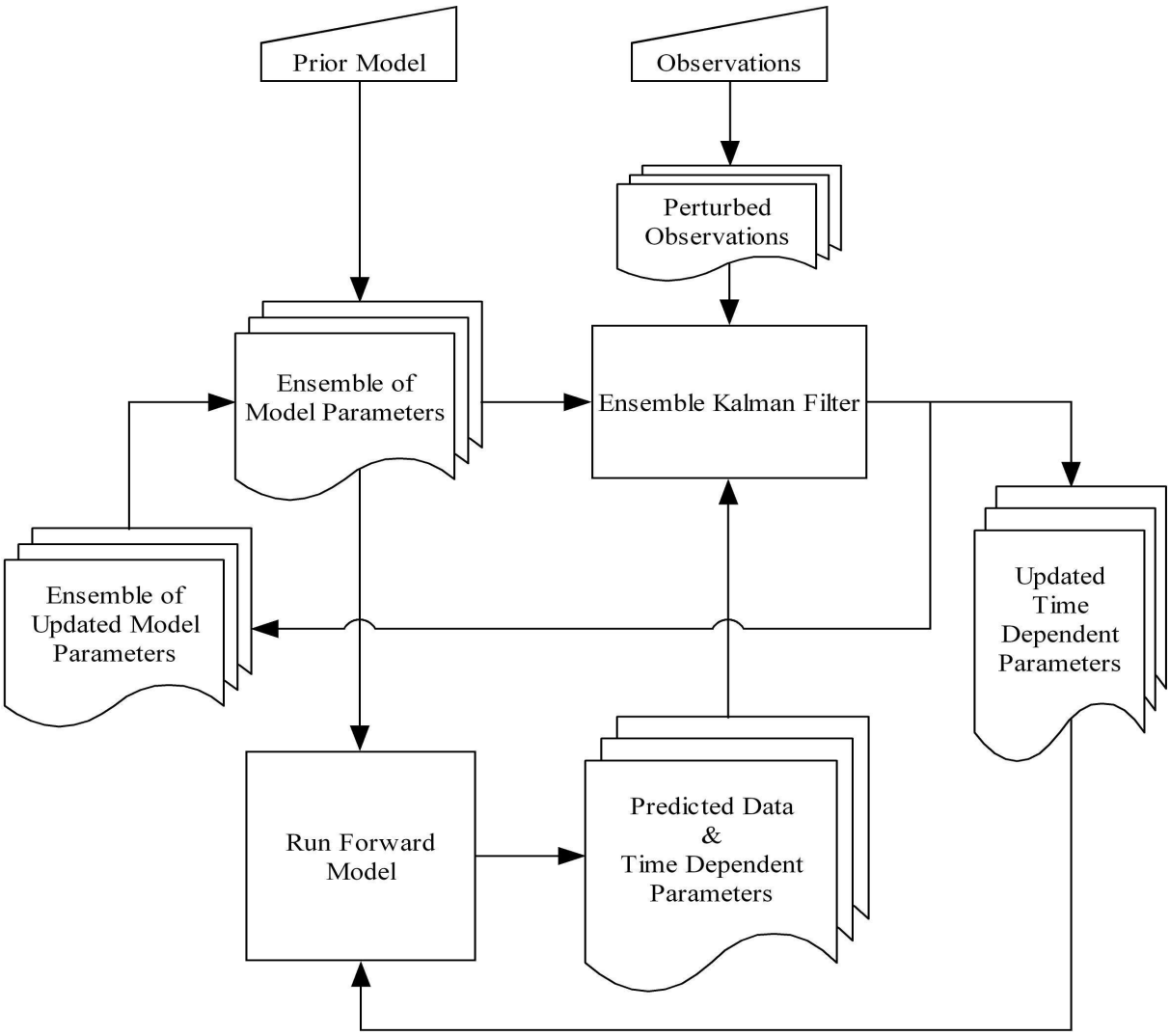


Figure 2.1: Ensemble Kalman filtering flowchart.

2.4.2 Linear Problem

As EnKF updates not only model parameters but primary variables at a data assimilation step, it is possible to obtain physically unreasonable values of pressures and saturations that must be modified to physically reasonable values. Intuitively, one might expect to eliminate this problem by rerunning the reservoir simulation time step. Here, we consider a simple linear problem to investigate the effect of updating of time dependent parameters (primary variables in reservoir simulation) by rerunning the forward model in EnKF. Suppose we have the following relationships between p, d and m,

$$p^{k+1} = F_k m^k + A_k p^k + \alpha_k, \tag{2.129a}$$

$$d^{k+1} = G_k m^k + B_k p^k + \beta_k, \tag{2.129b}$$

where F_k, A_k, G_k and B_k are matrices and α_k and β_k are known vectors. Here, k denotes the time step, p^{k+1} and d^{k+1} are at time t_{k+1} that we assimilate data. We note that if d^{k+1} is a function of m^k and p^{k+1}, we can still write Eq. 2.129b as a function of m^k and p^k by using a new set of coefficient matrices. We rewrite Eqs. 2.129a and 2.129b, in the form of a matrix of ensembles as

$$P^{k+1} = F_k M^k + A_k P^k + \alpha_k \tilde{I}, \tag{2.130a}$$

$$D^{k+1} = G_k M^k + B_k P^k + \beta_k \tilde{I}, \tag{2.130b}$$

where $\tilde{I}$ is a N_e-dimensional row vector where each entry is equal to 1.

Note that $\tilde{I}(D - \overline{D})^T$ is equal to a null matrix, i.e.,

$$\tilde{I}(D - \overline{D}) = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} (d_1 - \overline{d})^T \\ (d_1 - \overline{d})^T \\ \vdots \\ (d_{N_e} - \overline{d})^T \end{bmatrix} = \sum_{j=1}^{N_e} (d_j - \overline{d}) = O. \tag{2.131}$$

Then from Eq. 2.127, we see that $\tilde{I}\delta(D)$ is also a null matrix, i.e.

$$\tilde{I}\delta(D) = O. \tag{2.132}$$

Now suppose we have data at two time steps, the k+1st and k+2nd time steps, that we wish to assimilate. After assimilation of data at t_{k+1}, from Eqs. 2.127 and 2.130a, we see that for the standard EnKF method where we *update* the time dependent parameters, we have

$$\begin{aligned} P^{k+1,u} &= P^{k+1}\big(I + \delta(D^{k+1})\big) \\ &= F_k M^k \big(I + \delta(D^{k+1})\big) + A_k P^k \big(I + \delta(D^{k+1})\big) + \alpha_k \tilde{I}, \end{aligned} \tag{2.133}$$

whereas, for the case that we *rerun* the time step, we have

$$P^{k+1,r} = F_k M^k\big(I + \delta(D^{k+1})\big) + A_k P^k + \alpha_k \tilde{I}. \tag{2.134}$$

Now we continue assimilation of data for one more time step. After assimilation of second set of data at time t_{k+2}, the normal EnKF updating gives

$$\begin{aligned} P^{k+2,u} &= P^{k+2}\big(I + \delta(D^{k+2})\big) \\ &= F_{k+1} M^k\big(I + \delta(D^{k+1})\big)\big(I + \delta(D^{k+2})\big) + A_{k+1} P^{k+1,u}\big(I + \delta(D^{k+2})\big) + \alpha_{k+1}\tilde{I} \\ &= F_{k+1} M^k\big(I + \delta(D^{k+1})\big)\big(I + \delta(D^{k+2})\big) \\ &\quad + A_{k+1}\Big(F_k M^k\big(I + \delta(D^{k+1})\big) + A_k P^k\big(I + \delta(D^{k+1})\big) + \alpha_k \tilde{I}\Big)\big(I + \delta(D^{k+2})\big) \\ &\quad + \alpha_{k+1}\tilde{I} \end{aligned} \tag{2.135}$$

For the case that we rerun the time step,

$$\begin{aligned} P^{k+2,r} &= F_{k+1} M^k\big(I + \delta(D^{k+1})\big)\big(I + \delta(D^{k+2})\big) + A_{k+1} P^{k+1,r} + \alpha_{k+1}\tilde{I} \\ &= F_{k+1} M^k\big(I + \delta(D^{k+1})\big)\big(I + \delta(D^{k+2})\big) \\ &\quad + A_{k+1}\Big(F_k M^k\big(I + \delta(D^{k+1})\big) + A_k P^k + \alpha_k \tilde{I}\Big) \\ &\quad + \alpha_{k+1}\tilde{I} \end{aligned} \tag{2.136}$$

Although the $\delta(D^{k+2})$ in Eqs. 2.135 and 2.136 are not the same, by comparing the right hand sides of these equations, we can see that in the second term on RHS of Eq. 2.136 the model part has not been updated. If we use the final model parameters obtained from the case where we reran the data assimilation time step and apply the forward model (Eqs. 2.130a and 2.130b) forward from "time zero", the predicted P^{k+2} will not be consistent with $P^{k+2,r}$. Thus, using $P^{k+2,r}$ to predict forward to times greater than time t_{k+2} will give P^l results inconsistent with the estimate of the model. Therefore, in our implementation we simply truncate any such value back to a prescribed upper or lower bound for the variable. For example, water saturation physically can not be less than 0

or greater than 1. Therefore, in our study every time the updated water saturation of any gridblock is less than 0 we change the saturation to 0 for that particular gridblock.

CHAPTER 3

TOY PROBLEMS

3.1 Toy Problem 1

Here, we apply the EnKF to a simple non-linear problem. For this example, there is only one model parameter, i.e. m is a real random variable. Consider the following forward model,

$$d = g(m,t) = 1 - \frac{9}{2}\Big(m - \frac{2\pi}{3}\Big)^2 + (t-1)\sin(m), \tag{3.1}$$

where m is a scalar model parameter and t is time. To make this model similar to a problem where our forward model is a reservoir simulator, we define the following recursive equation for predicting data with the restart option of the simulator,

$$g(m, t + \Delta t) = g(m,t) + \sin(m)\Delta t. \tag{3.2}$$

We generate true synthetic data using the true model which is $m_{true} = 1.88358$, at five times $t = 1, 2, ..., 5$. At each time, there is a single datum. To generate synthetic observed data, we add normal random noise generated from $N(0, 0.01)$ as measurement error. We consider a normally distributed prior model, $m \sim N(2.4, 0.1)$. To assimilate the data at each time step, we update the time dependent parameters along with the model parameters during the assimilation of data. Note that in this problem, the data and time dependent parameter (p in Eq. 2.122) for each data assimilation time step are the same. Fig. 3.1 shows the prior pdf and the posterior distribution conditioned to 1, 2, 3, 4 and 5 data. The posteriori pdf's are non-Gaussian and the pdf's conditional to data at t_1, t_2 and t_3 have two distinct modes. In this case, we start the process with 5001 ensembles, 5000 generated as unconditional realizations from the prior plus one ensemble equal to the prior mean. Figs. 3.2 through 3.4 show the histogram of the model parameter

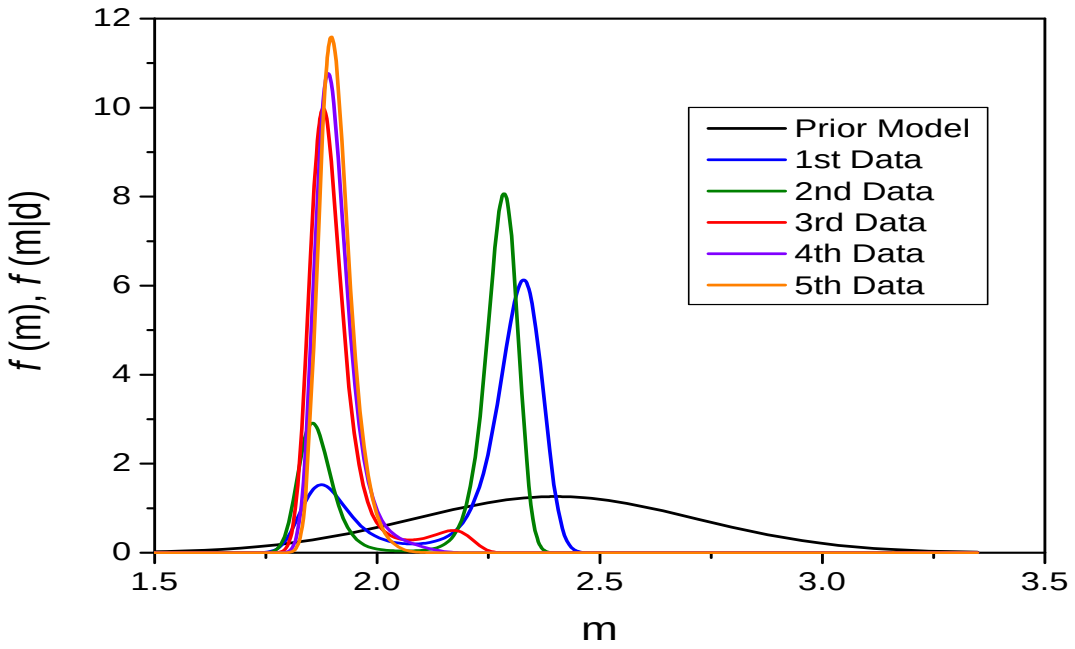


Figure 3.1: Prior pdf and posterior pdf after integrating 1, 2, 3, 4 and 5 data, toy problem 1.

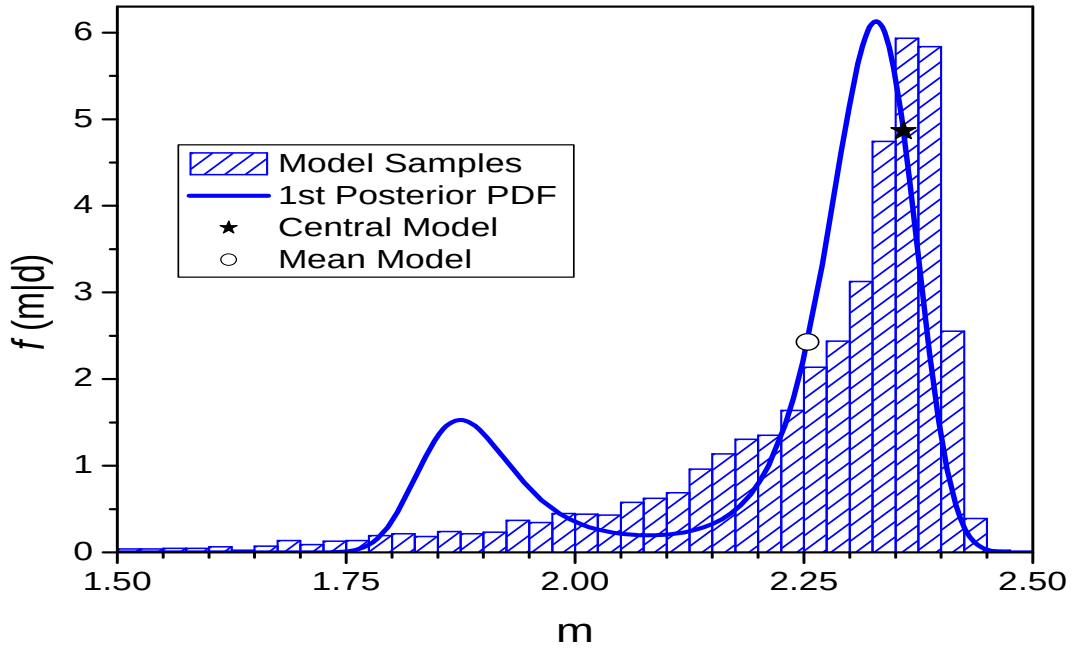


Figure 3.2: Posterior pdf and ensembles after assimilation of first data, toy problem 1.

and the posterior pdf after assimilation of the data at $t = 1, 3$ and 5. Here, we add the prior mean to the set of ensembles and during data assimilation, update this ensemble by assimilating d_{obs} (not a d_{uc}) to obtain an updated model which we refer to as the central model. For the linear case, it is easy to show that the central model is the same as the MAP estimate. In Fig. 3.2, the posterior pdf shows high probability around the model $m = 2.30521$ (second root of the quadratic part of the model) because for $t = 1$ the sine term in Eq. 3.1 does not have any effect on the predicted data. By assimilating more data the sine term in the model becomes dominant and $m = 2.30521$ is no longer a valid model. Figs. 3.3 and 3.4 show how the posterior pdf moves so that it finally has a single mode close to the truth. As we can see, the mean model, average of all ensembles, is always slightly closer to the true model for all data assimilation steps. After assimilation of the datum corresponding to $t = 5$, EnKF (see Fig. 3.4) gives a more reliable sampling of the correct pdf, but still gives a distribution which is too broad. Even as we continue to assimilate data beyond $t = 5$, the distribution obtained with EnKF remains broader than the true distribution. Overall, the results suggest that the standard EnKF method does not give a reliable estimate of the true pdf in a multi-modal case, i.e., does not give a proper characterization of the uncertainty in m and even in the single mode case it may over estimate the uncertainty in m.

Figs. 3.5 and 3.6 show the distribution of the model parameter after matching one and five data using the RML method. The unconditional realizations (starting ensembles) in RML method and EnKF are the same. For this non-linear problem RML approximate the posterior pdf much better than EnKF.

3.2 Toy Problem 2

In the previous problem, as we assimilate more data the posterior pdf moves toward a conditional pdf with a single mode. We now consider a problem when as more data are assimilated, the posterior pdf remains multi-modal. For this example the forward model is given by

$$d = g(m, t) = 1 - \frac{9t}{2}\Big(m - \frac{2\pi}{3}\Big)^2. \tag{3.3}$$

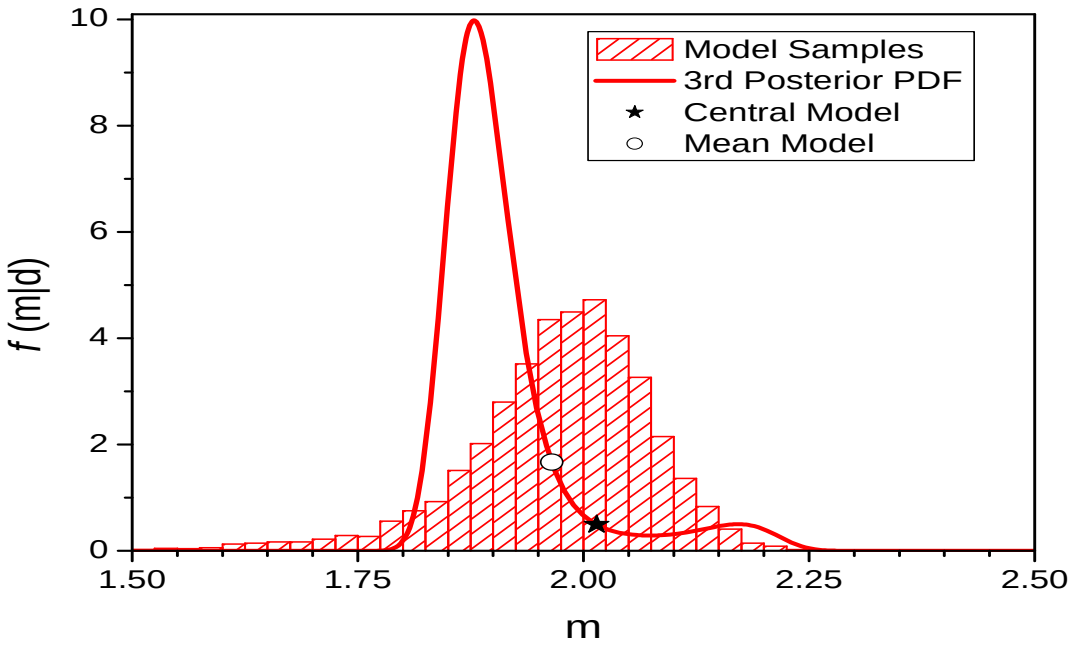


Figure 3.3: Posterior pdf and ensembles after assimilation of third data, toy problem 1.

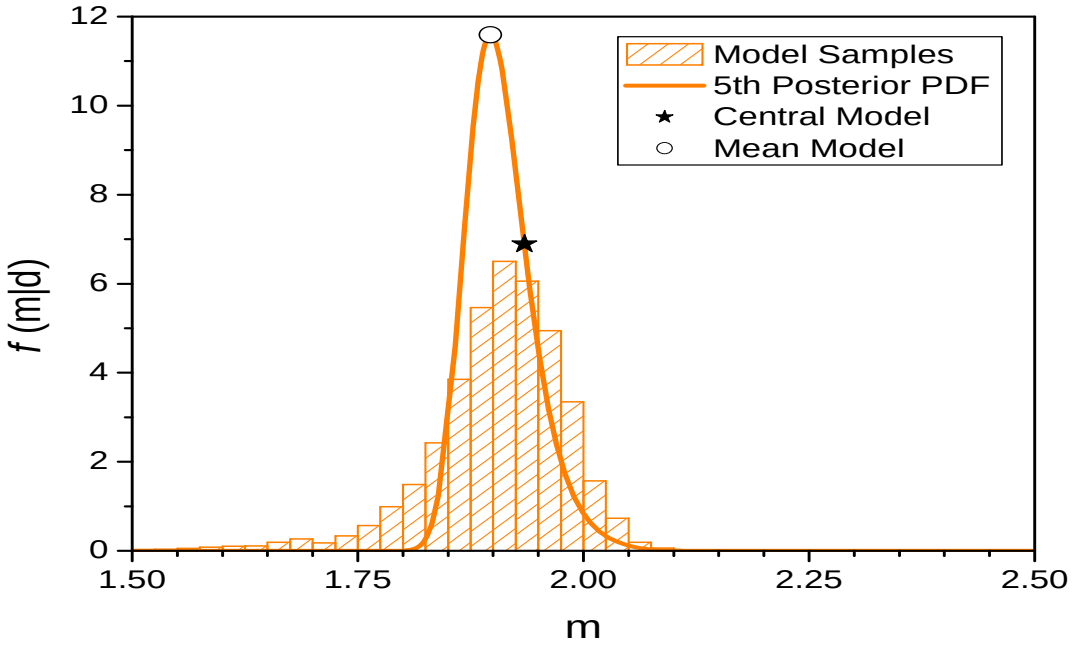


Figure 3.4: Posterior pdf and ensembles after assimilation of fifth data, toy problem 1.

To make this data prediction similar to a problem where our forward model is a reservoir

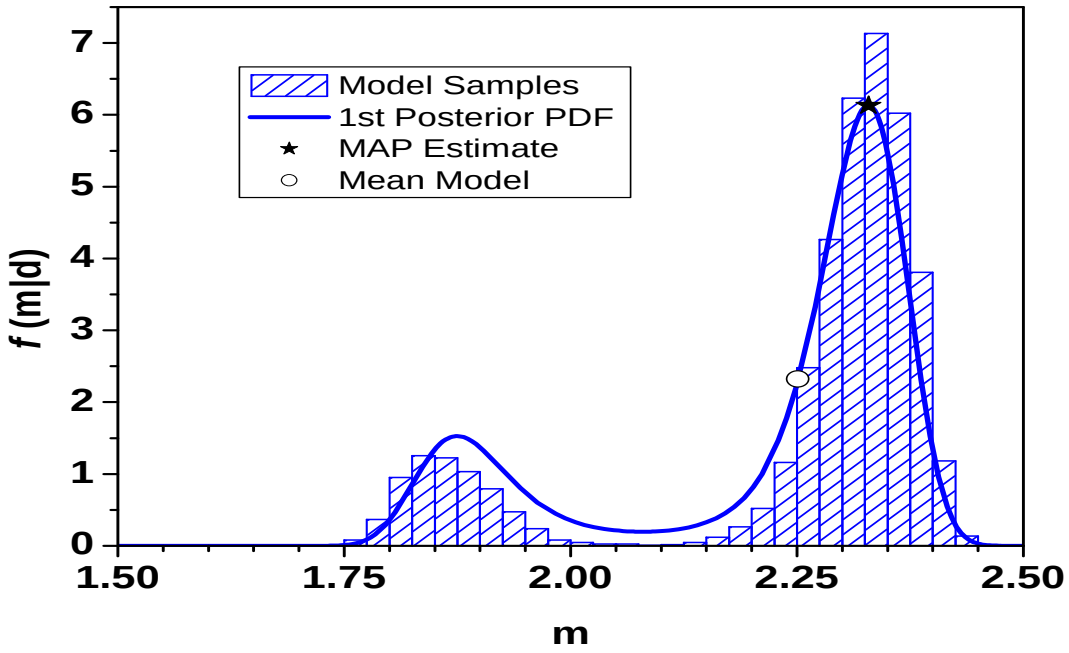


Figure 3.5: Posterior pdf and distribution of conditional realizations (ensembles) after matching one data by the RML method, toy problem 1.

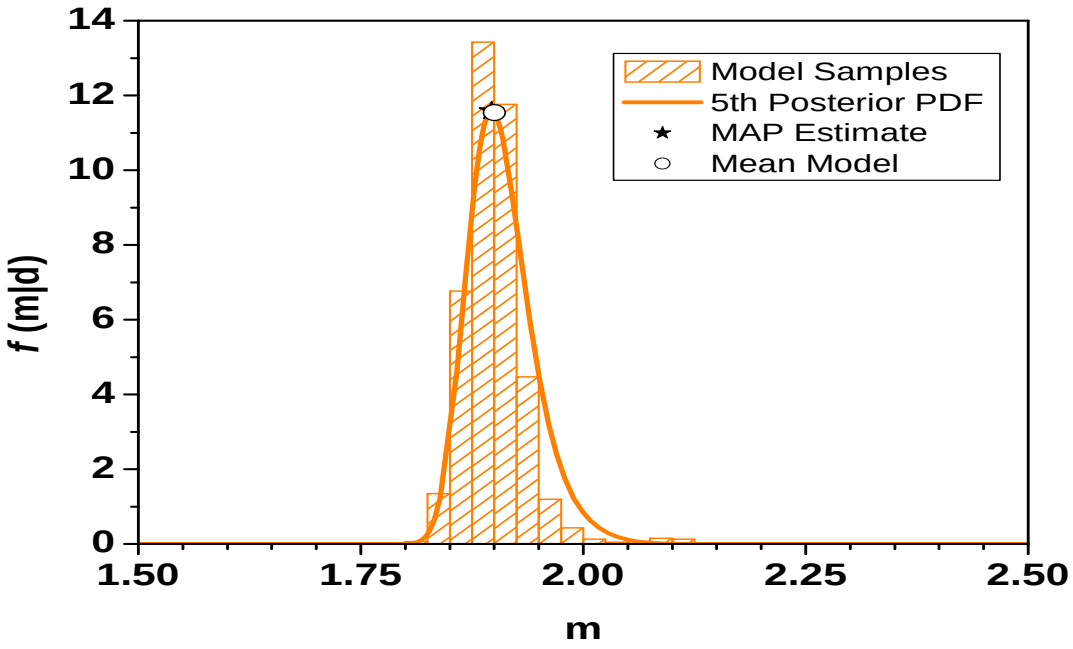


Figure 3.6: Posterior pdf and distribution of conditional realizations (ensembles) after matching five data by RML method, toy problem 1.

simulator, we rewrite Eq. 3.3 as,

$$g(m, t + \Delta t) = g(m, t) - \frac{9\Delta t}{2}\Big(m - \frac{2\pi}{3}\Big)^2. \tag{3.4}$$

We generate true synthetic data using the true model given by $m_{\text{true}} = 1.88358$, at five times $t = 1, 2, ..., 5$ (Note each datum which can be produced by the true model, can also be generated by $m = 2.30521$). At each time, there is a single datum. To generate synthetic observed data, we add normal random noise generated from $N(0, 0.01)$ as measurement error. We consider a normally distributed prior model, $m \sim N(2.1, 0.2)$. To assimilate the data at each time step, we update the time dependent parameters along with the model parameters during the assimilation of data. Fig. 3.7 shows the posterior pdf after matching 5 data at $t = 5$ and Fig. 3.8 shows the histogram of the model parameter after assimilation of the data at $t = 5$. We can see the real posterior pdf is much narrower than the histogram generated by the ensembles. This example confirms again that EnKF may provide a relatively poor characterization of uncertainty in the model in a multi-modal case.

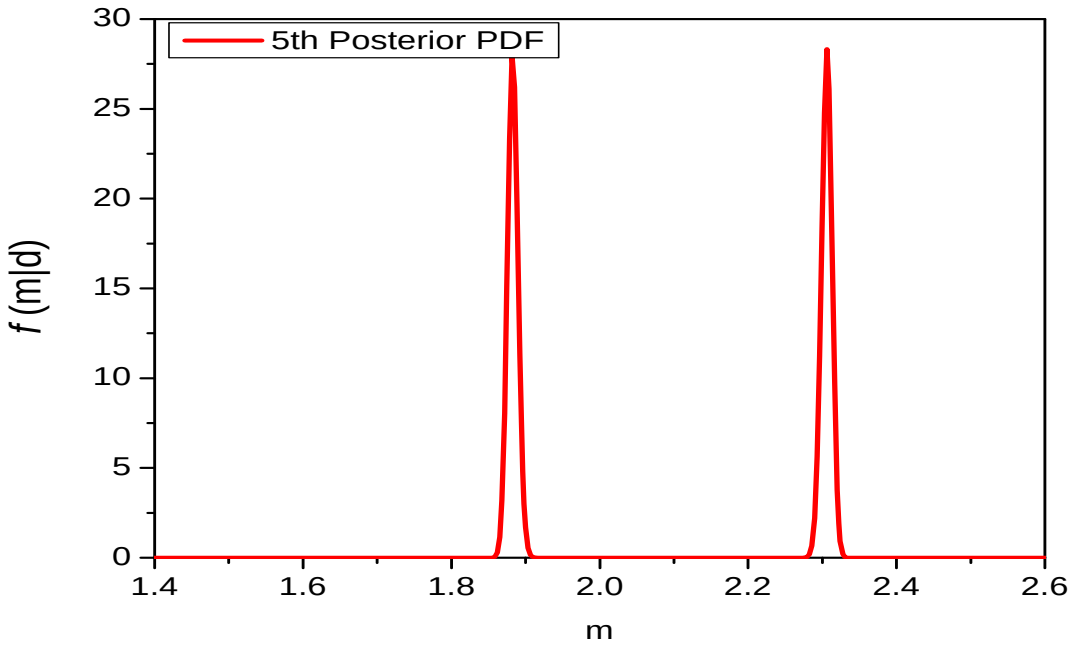


Figure 3.7: Posterior pdf after assimilation of fifth data, toy problem 2.

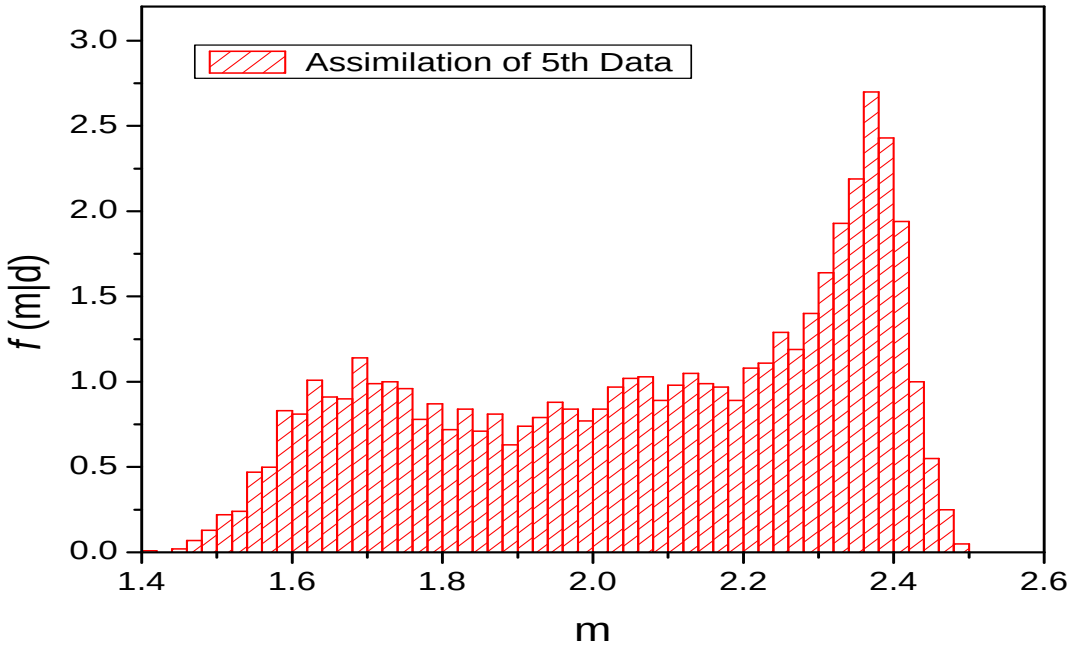


Figure 3.8: Ensembles after assimilation of fifth data, toy problem 2.

Fig. 3.9 and shows the distribution of the model parameter after matching five data using the RML method. The unconditional realizations (starting ensembles) in RML method and EnKF are the same. For this non-linear problem RML provides an almost perfect approximate to the posterior pdf.

3.3 Metric

Fig. 3.10 shows the difference between the average model and the truth after assimilation of each datum. For the first toy problem the difference decreases after each data assimilation but in the second toy problem, the difference between the average model, which we usually consider as our estimate from the EnKF method, is not becoming closer to the truth as more data are assimilated. Fig. 3.11 illustrates the average of the difference between each ensemble and true model, i.e., the average of $|m - m_{\text{true}}|$. We can see for the second problem the average difference between the ensembles and the true model decreases as data at $t = 4$ and $t = 5$ are assimilated. For toy problem 1, EnKF clearly provides a better characterization of the conditional pdf for m as more data are

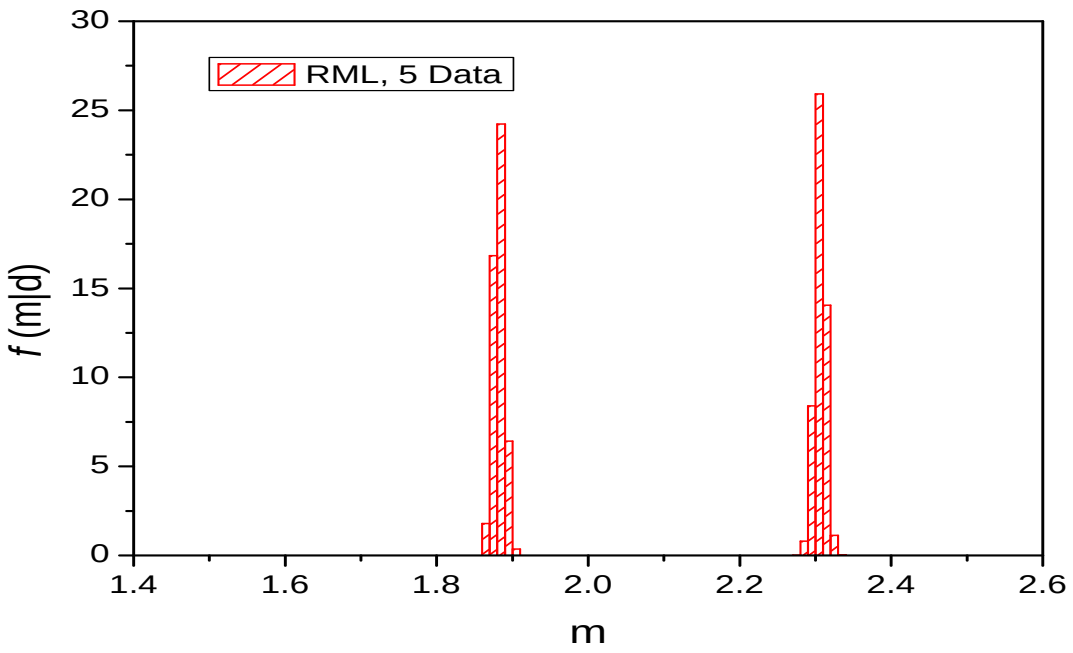


Figure 3.9: Posterior pdf and distribution of conditional realizations (ensembles) after matching five data by the RML method, toy problem 2.

assimilated, see Figs. 3.2 through 3.4. This is also true in toy problem 2 although the improvement is slight and is not apparent from the results we have provided. Now the question is what metric should be used to quantify the reliability of EnKF in terms of providing an estimate of the truth. The mean of the ensembles is conventionally used as an estimate even though we have shown that the mean is not guaranteed to provide a good estimate. This is consistent with the fact that EnKF gives a slightly better characterization of the relevant pdf as more data is assimilated and suggests that the average of the difference between each ensemble and the truth is a better indicator of how EnKF is performing even if the average model is used as the estimate. The results of Fig. 3.10 are based on using the mean model as an estimator of the true model, but similar results are obtained if the central model is used as the estimator of the truth. Note the central model becomes more important in the presence of time dependent parameters, i.e. after assimilation of production data, to predict the performance of the reservoir, the central model can be run by its updated time dependent parameters.

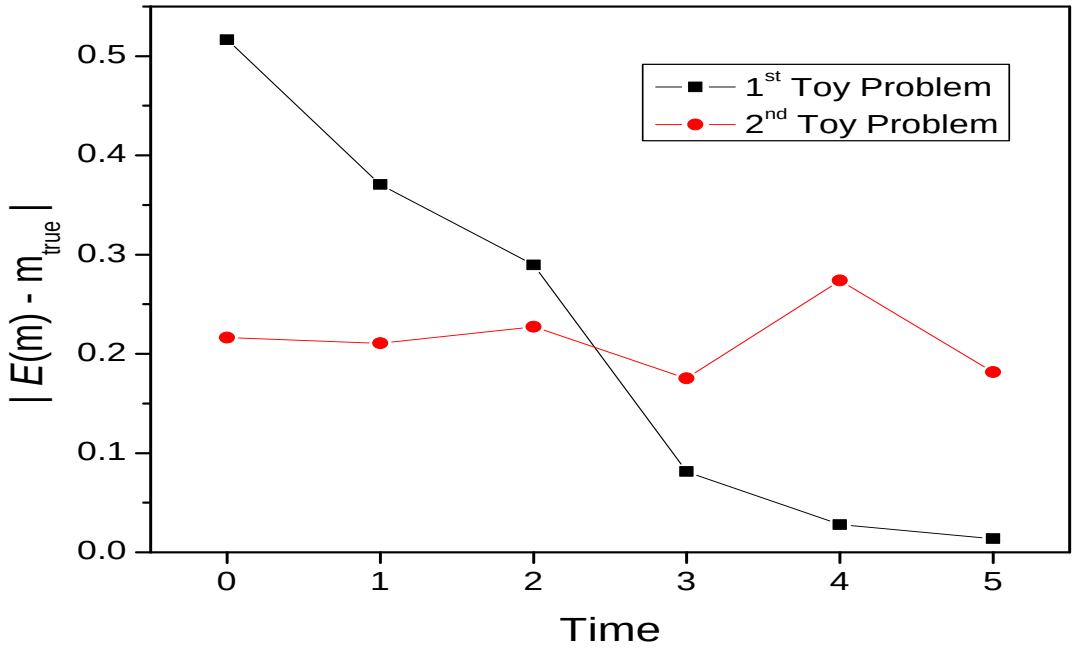


Figure 3.10: Difference of the average of ensembles and true model after each data assimilation.

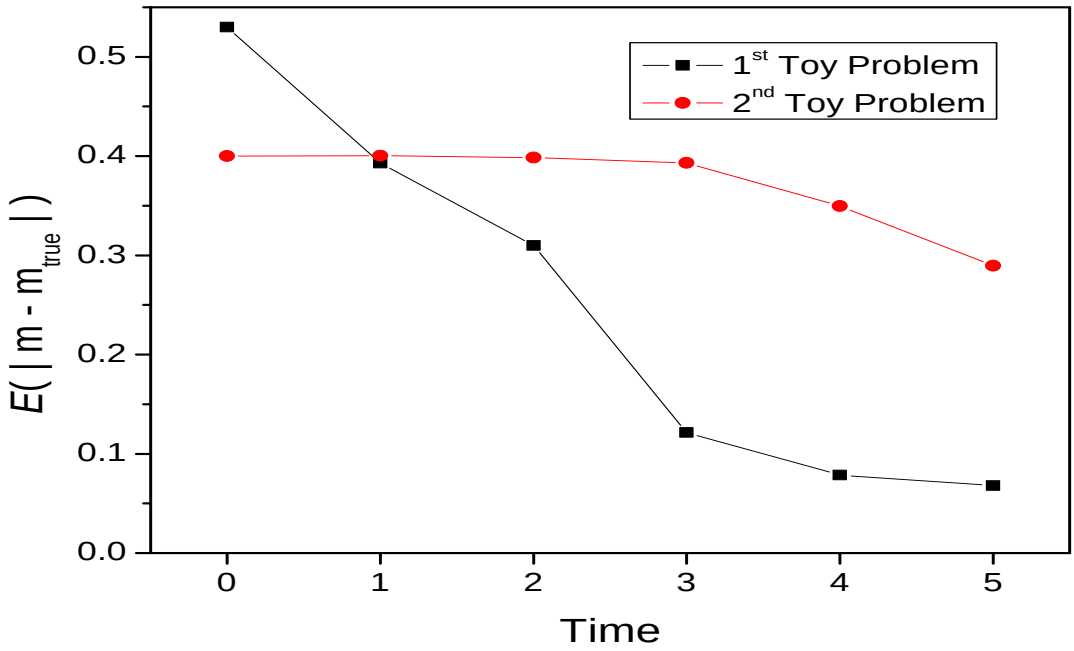


Figure 3.11: Average of the differences between each ensemble and true model after each data assimilation.

CHAPTER 4

2D SYNTHETIC PROBLEM

4.1 Production History

The horizontal 2D problem consists of a 20×30 grid with $\Delta x = \Delta y = 300$ feet and 600 active gridblocks. There are 6 producing wells. The production consists of 90 days of production from well PRO-1 with all other wells shut in. After this initial 90 day period, well PRO-2 begins production while well PRO-1 continues to produce. This sequence is continued until we reach a total time of $t = 540$ days at which well PRO-6 has been producing for exactly 90 days. At each well, the production well is specified to be $q_o = 100$ STB/day from the time the well is put on production up to 540 days. After this 540 day period, we change well PRO-1 from a producer to an injector, INJ-1, and continue to produce from the other wells until 5850 days. At day 5850, we close well PRO-5 and continue production from the rest of the wells, until 7290 days. From $t = 540$ to 5850 days, the water injection rate is 1000 STB/day and during this time period, the wellbore constraint at the five producing wells is $q_o = 200$ STB/day. At $t = 5850$ days, the injection rate is increased to 2000 STB/day, well PRO-5 is shut in and the other wells produce based on a target rate of $q_o = 300$ STB/day.

The true rock property fields were generated using sequential Gaussian co-simulation. The key geostatistical parameters used to generate the truth are listed in Table. 4.1. The production data generated consist of (i) flowing bottom hole pressure, producing gas-oil ratio (GOR) and water cut at producing wells; (ii) bottom hole pressure of the injector. We generated these true data using the ECLIPSE simulator. To generate the observed data, d_{obs}, we added Gaussian noise to the true production data. The relative error added to the true data is 5% for bottom hole pressures, 5% for producing gas-oil ratio and 0.1% for water cut.

4.1.1 Generation of Ensembles

A suite of unconditional realizations (ensembles) were generated using joint sequential multi-Gaussian code developed by Gómez-Hernández et al. [13]. For the ensemble Kalman filter, 90 unconditional realizations were generated as the initial set of ensembles based on the true geostatistical parameters given in Table. 4.1. r_1 and r_2 are the primary and secondary correlation lengths of the structure in the direction of anisotropy axes, α.

ϕ_{mean}	0.2
$[\ln k]_{\text{mean}}$	4.0
σ_ϕ	0.05
$\sigma_{\ln k}$	2.0
$\rho_{\phi,\ln k}$	0.8
α	40°
r_1 (ft)	8400
r_2 (ft)	1500

Table 4.1: Geostatistical parameters of 2D synthetic problem.

4.2 Metric

Fig. 4.1 shows the difference between the average model and the true model for the porosity and log-permeability fields. Here we define the function Ψ by

$$\Psi(m) = \frac{1}{N_m}\sum_{j=1}^{N_m}\Big([\frac{1}{N_e}\sum_{i=1}^{N_e} m_j^i] - m_j^{\text{true}}\Big)^2, \tag{4.1}$$

where N_m is the number of a particular type of model parameters and $\frac{1}{N_e}\sum_{i=1}^{N_e} m_j^i$ equals the average of all ensembles for the jth grid block. Similar to the second toy problem, the metric of Eq. 4.1 shows that we are not obtaining more reliable estimate of the model as more data are assimilated. But when we compare the 2D map of the porosity and log-permeability field with respect to the truth, Figs. 4.3 through 4.8, we can see that there is an improvement in estimating the structure of the true model. Therefore we define a new function Γ by

$$\Gamma(m) = \frac{1}{N_m}\sum_{j=1}^{N_m}\frac{1}{N_e}\sum_{i=1}^{N_e}\big(m_j^i - m_j^{\text{true}}\big)^2, \tag{4.2}$$

Fig. 4.2 shows the behavior of function Γ for both porosity and log-permeability after each data assimilation. We can see that in general, this metric suggest that the ensembles are getting closer to the truth as more data are assimilated. Note that $\Gamma(m)$ shows a noticeable increase at the two times ($t = 540$ and $t = 5850$ days) where the well producing conditions were changed sharply.

Figs. 4.9 and 4.10 show the water saturation profile for the true model and average of all ensembles after assimilation of data at 7290 days. We can see that there is good agreement between the truth and the average of saturation field. Calculations show that there is 5% error in the water content of the reservoir after 7290 days. The two black regions on the average predicted saturation profile belongs to the places where the updated water saturation is less than irreducible water saturation.

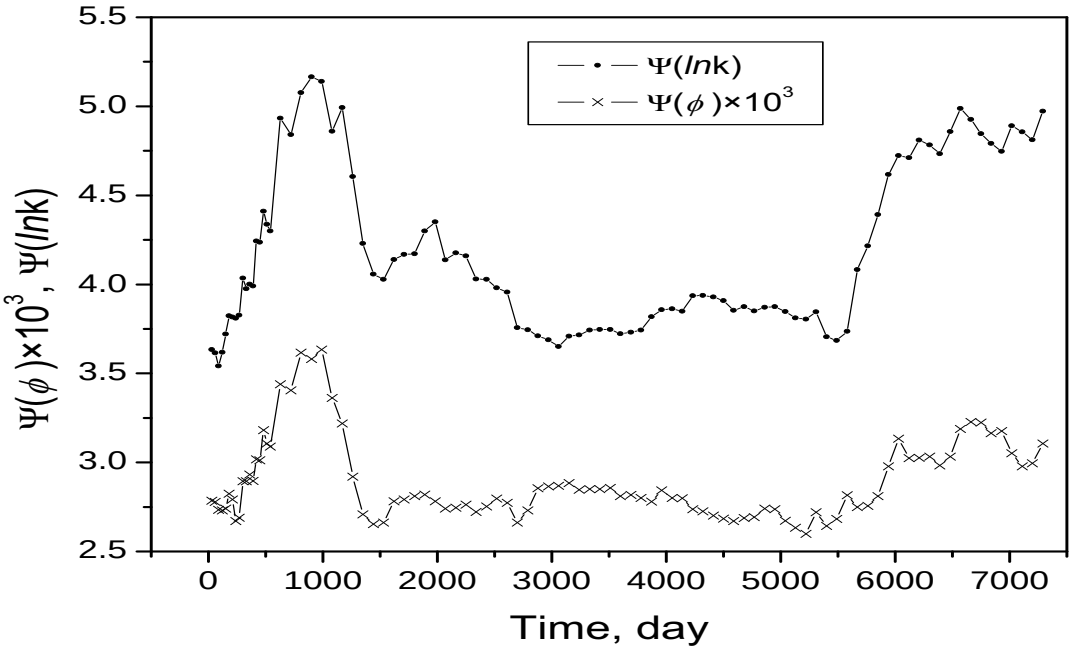


Figure 4.1: Difference of the average of ensembles and true model after each data assimilation, 2D synthetic problem.

4.3 Data Match and Performance Prediction.

To investigate the EnKF for predicting the future performance of the reservoir, first we predict the cumulative oil production for the initial ensembles (it is possible that

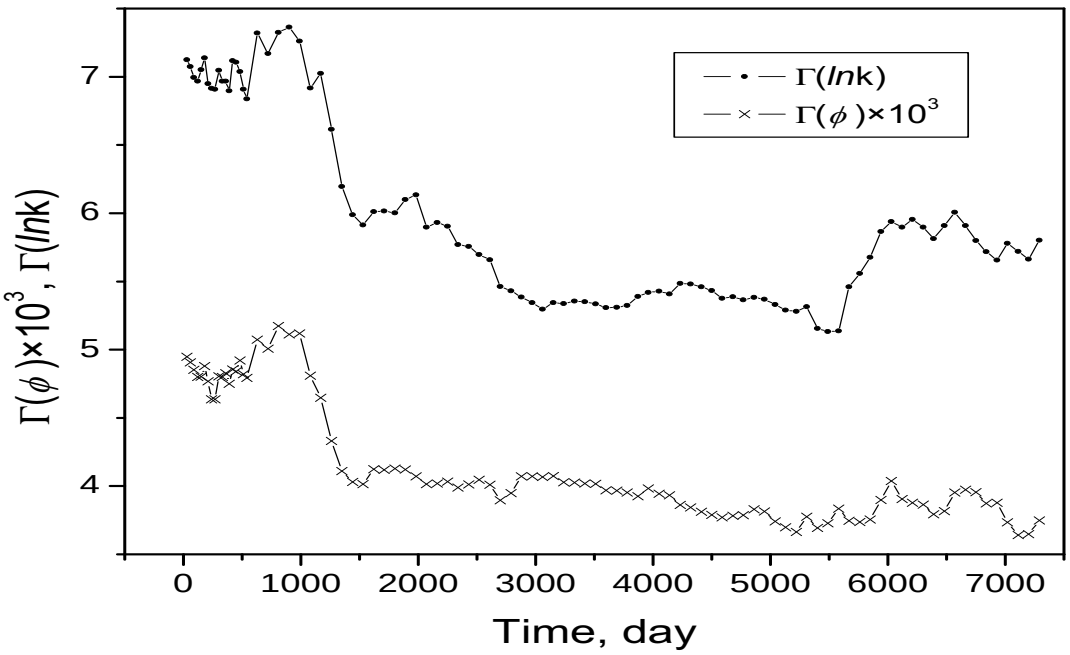


Figure 4.2: Average of the differences between each ensemble and true model after each data assimilation, 2D synthetic problem.

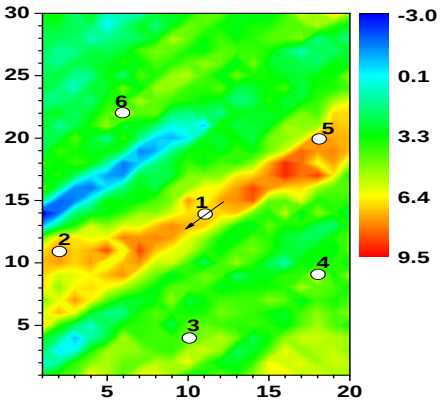


Figure 4.3: Log-permeability field, truth, 2D synthetic problem.

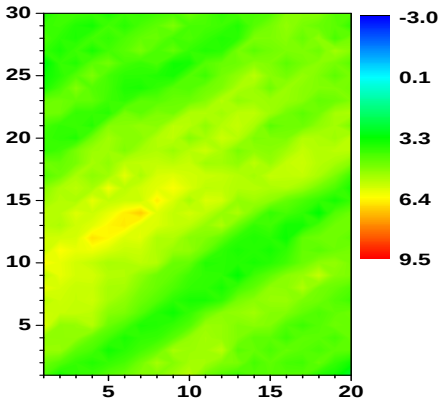


Figure 4.4: Average log-permeability after 540 days data assimilation, 2D synthetic problem.

some of the ensembles are not physically able to produce under the desired conditions for the 10000 day period considered). Then we assimilate the observed data for 5400 days, and predict until day 10000. In the next case, we continue the assimilation of data until 7290 days, then predict until day 10000. The reason that we use two different steps to

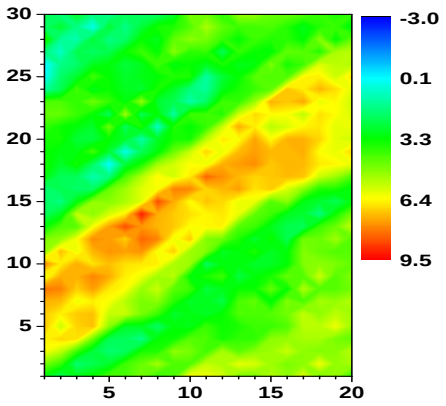


Figure 4.5: Average log-permeability after 5580 days data assimilation, 2D synthetic problem.

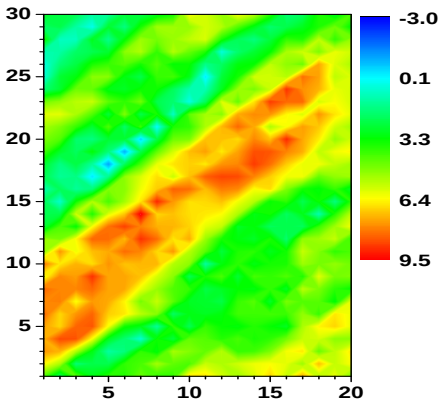


Figure 4.6: Average log-permeability after 7290 days data assimilation, 2D synthetic problem.

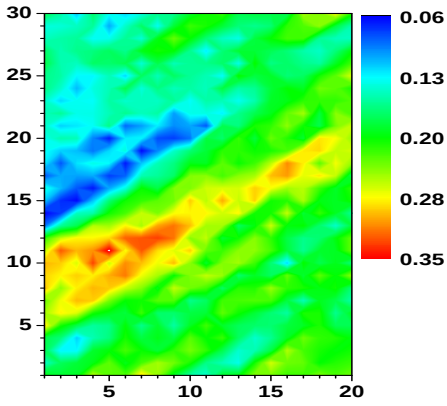


Figure 4.7: Porosity field, truth, 2D synthetic problem.

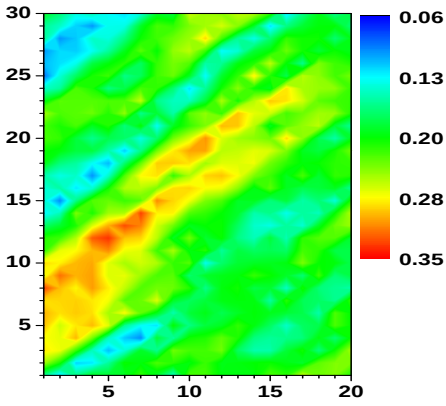


Figure 4.8: Average porosity after 7290 days data assimilation, 2D synthetic problem.

assimilate the data is to see the effect of data assimilation in the late production history. After 7290 days all the production wells (well PRO-5 will remain shut in) will produce at constant bottom hole pressure constraint. The bottom hole pressure constraint for wells PRO-2, 3, 4, and 6 are 2500, 1000, 1500 and 300 psi respectively. The water injection rate will remain 2000 STB/day as before.

Figs. 4.11 through 4.13 show the predicted cumulative oil production after 7290 days for three cases (i) set of initial ensembles (ii) after 5400 days of data assimilation and (iii) after 7290 days of data assimilation (before day 7290 all the wells are producing

under the constant oil rate constraint therefore the cumulative oil production is a fixed number for all the ensembles except for some of the unconditional realizations that can not produce under the constant oil rate constraint). Note that in second and third cases after assimilation of production data the band of predictions decreases significantly as compared to the predictions from initial ensembles, which means that the assimilation of data has greatly reduced the uncertainty in the production forecast. The results from the third case show that most of the ensembles over predict the cumulative oil production during the forecasting period. This occurs when the water cut in the wells producing under the constant bottom hole pressure constraint tends to be lower than the true water cut.

Fig. 4.14 shows the match of bottom hole pressure for well INJ-1(PRO-1 first 540 days) and prediction of bottom hole pressure until day 10000. Fig. 4.15 illustrate the same data except we continue assimilation of data for 7290 days then predict until day 10000. We can see that by assimilating the data between 5400 and 7290 days, the prediction is improved. In Fig. 4.14, after all the production wells switch to the constant bottom hole pressure constraint at $t = 7290$ days the injection pressure for some of the ensembles starts to increase. This is because at $t = 7290$ days, for some of the ensembles the predicted bottom hole pressure in production wells are below the constant bottom hole pressure constraint. This can be shown in Fig. 4.16 where at the time 7290 days there are some ensembles where the predicted bottom hole pressure for the well PRO-3 is below the constraint and ECLIPSE, the simulator that we use for forward model, considers them as shut in wells. Fig. 4.17 shows the bottom hole pressure of the well PRO-3 after 7290 day data assimilation. Here all of the ensembles are able to produce under the bottom hole pressure constraint.

Figs. 4.18 and 4.19 show the match of producing gas oil ratio for the well PRO-3. Again we can see after assimilation of data between 5400 to 7290 days the ensembles results in lower uncertainty in performance prediction. However, the predicted data from the central model does not change much.

In this problem, during the early producing period, water break through does not

occur in any of the wells. At time 5400 days water break through has occurred only in well PRO-5. Here, there is an interesting problem, at the time of water break through in well PRO-5, only one of the ensembles predict a significant water cut, therefore ΔD is almost a null matrix (all the ensembles predict almost the same value) and based on Eq. 2.127 there will be no correction to the model parameters. Fig. 4.20 shows the water cut in well PRO-5, we can see there is almost no improvement in the water cut prediction before the predicted water cuts become significant in some of the ensembles. Figs. 4.21 and 4.22 illustrate the water cut for well PRO-2 (the second production well in the high permeable region). Assimilation of more observations between 5400 days and 7290 days in general improves the predicted water cut from the ensembles but during that period the predicted water cut from the central model seems to get worse.

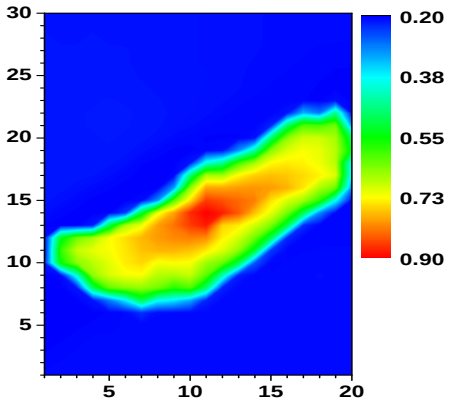


Figure 4.9: Water saturation profile after 7290 days, truth, 2D synthetic problem.

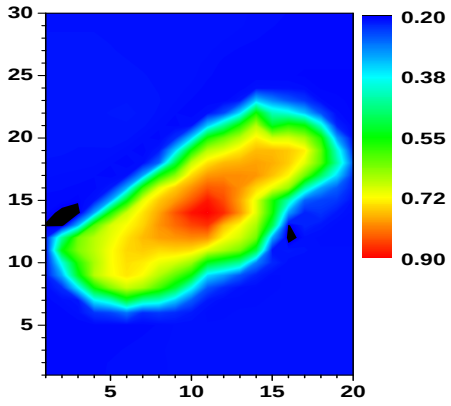


Figure 4.10: Average water saturation profile after 7290 days data assimilation, 2D synthetic problem.

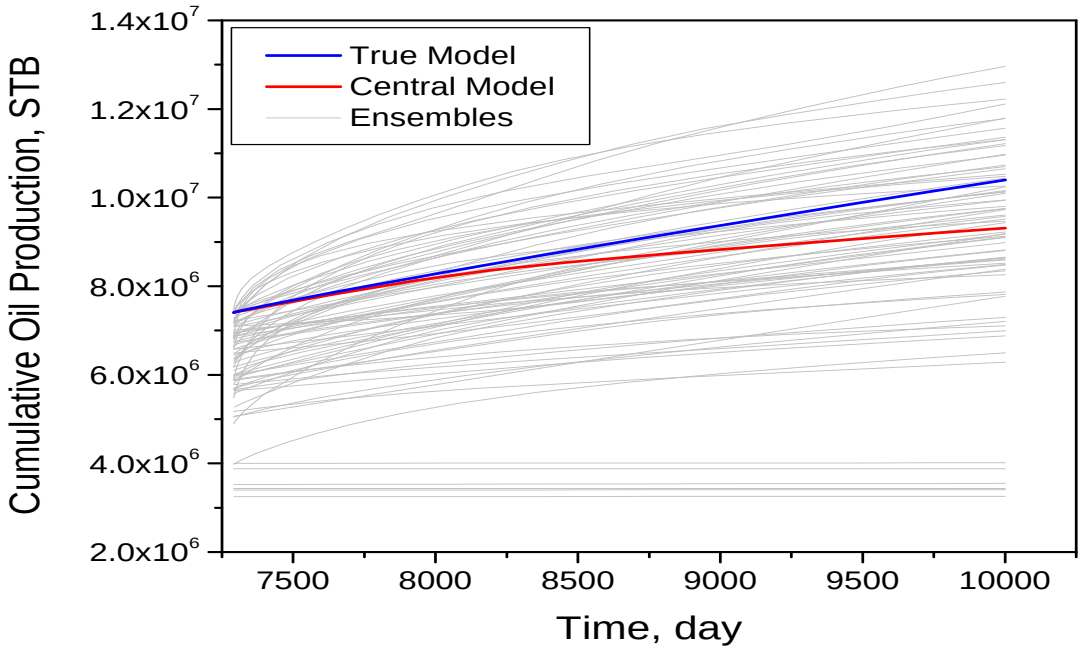


Figure 4.11: Cumulative oil production prediction for initial ensembles, 2D synthetic problem.

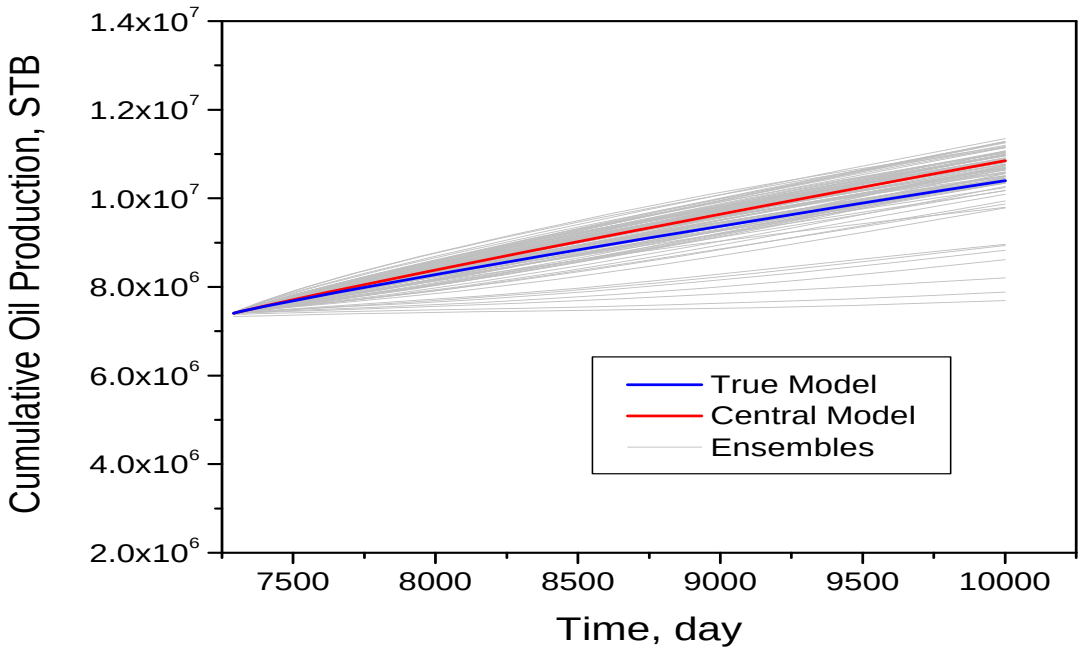


Figure 4.12: Cumulative oil production prediction after 5400 days data assimilation, 2D synthetic problem.

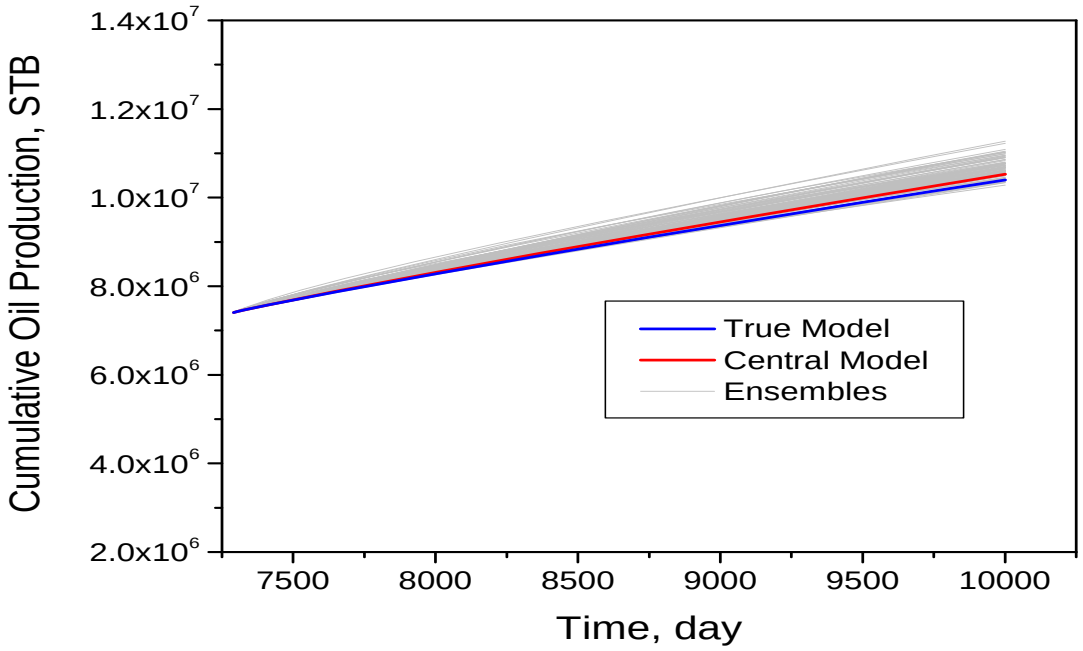


Figure 4.13: Cumulative oil production prediction after 7290 days data assimilation, 2D synthetic problem.

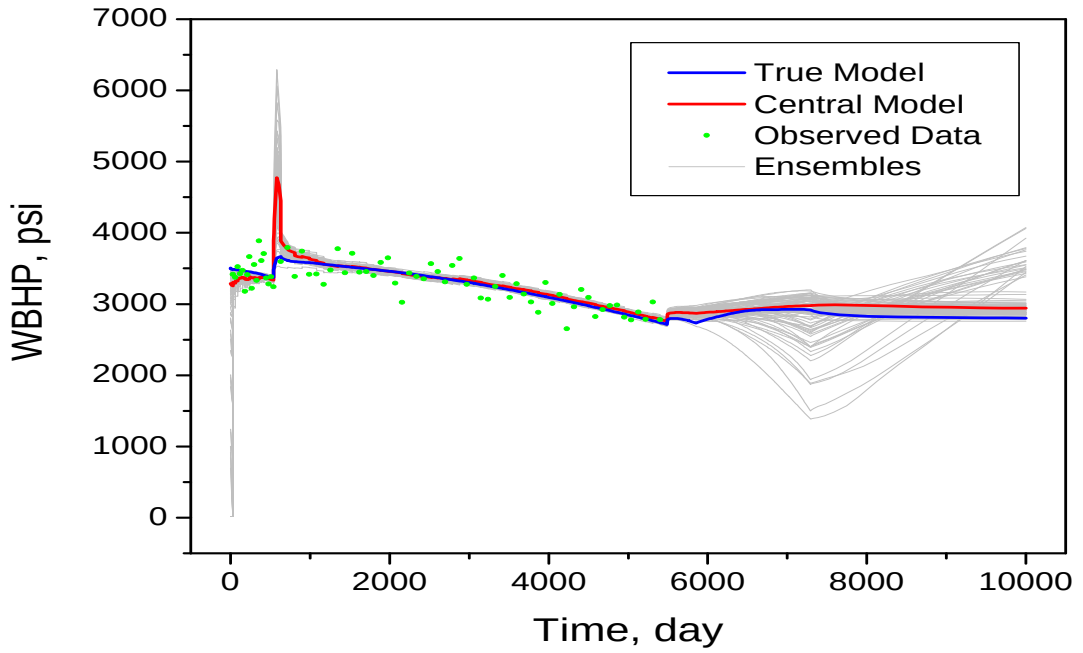


Figure 4.14: Well bottom hole pressure for INJ-1(PRO-1), data assimilation for 5400 days, prediction to 10000 days, 2D synthetic problem.

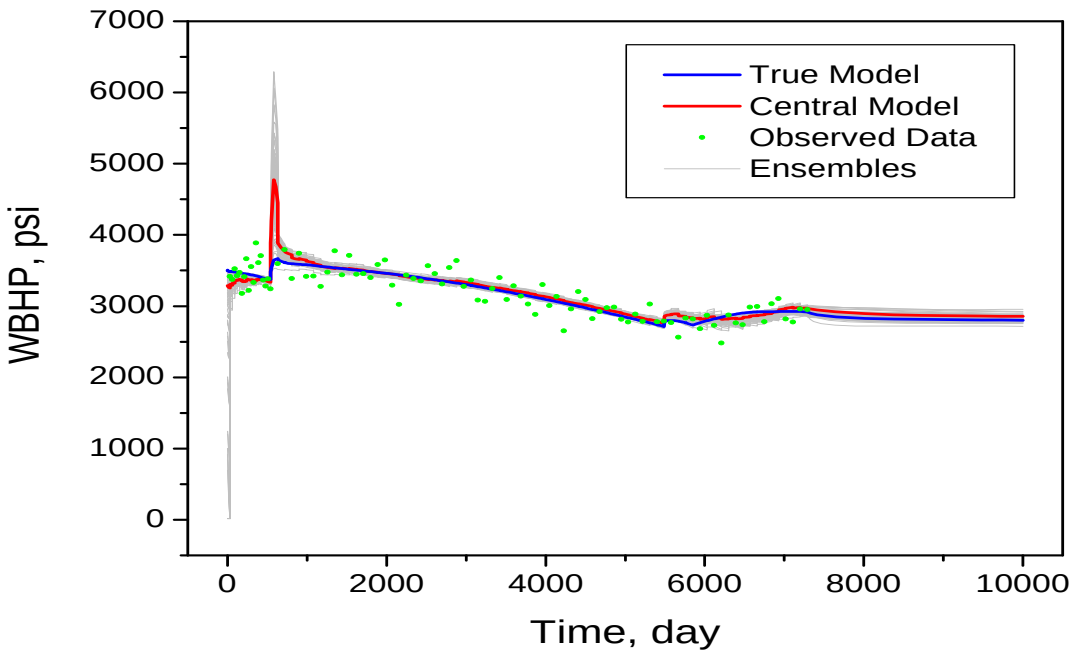


Figure 4.15: Well bottom hole pressure for INJ-1(PRO-1), data assimilation for 7290 days, prediction to 10000 days, 2D synthetic problem.

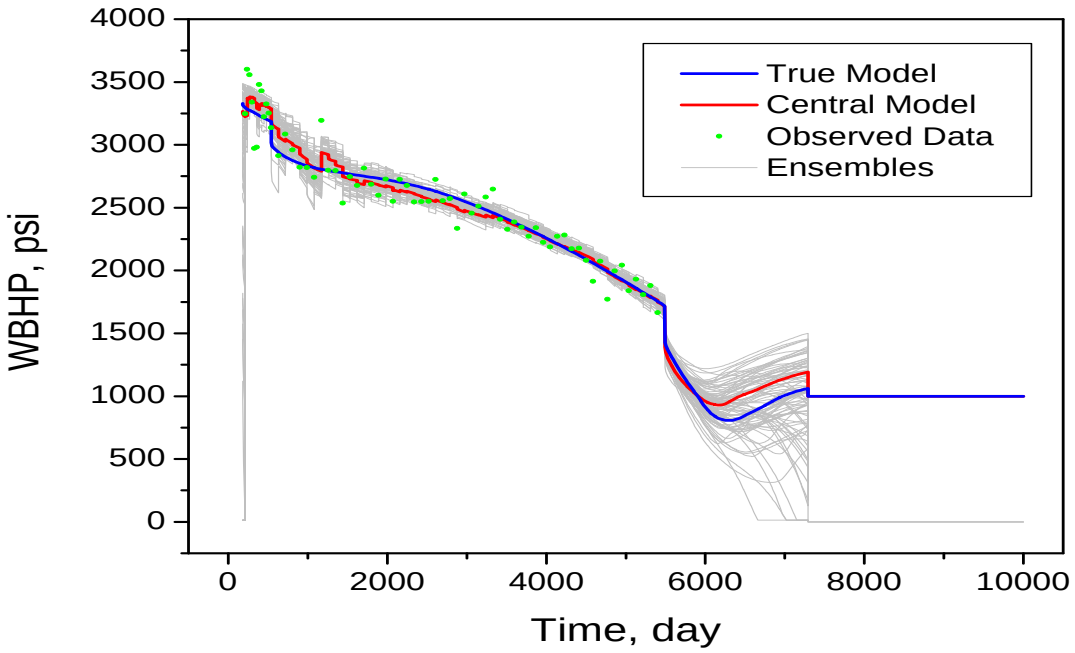


Figure 4.16: Well bottom hole pressure for PRO-3, data assimilation for 5400 days, prediction to 7290 days, 2D synthetic problem.

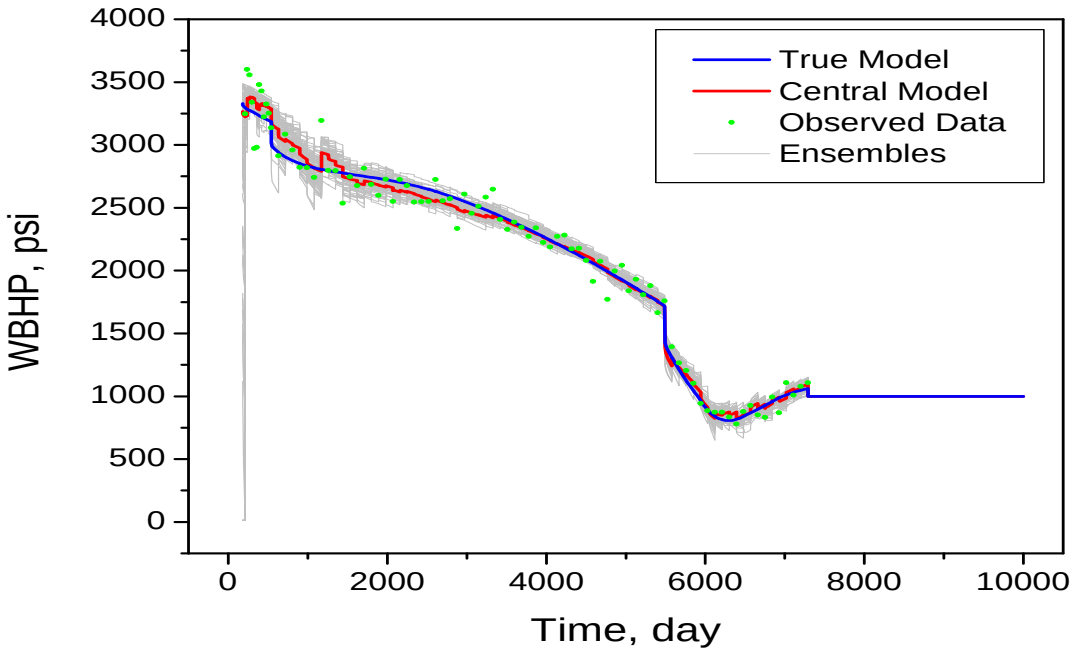


Figure 4.17: Well bottom hole pressure for PRO-3, data assimilation for 7290 days, 2D synthetic problem.

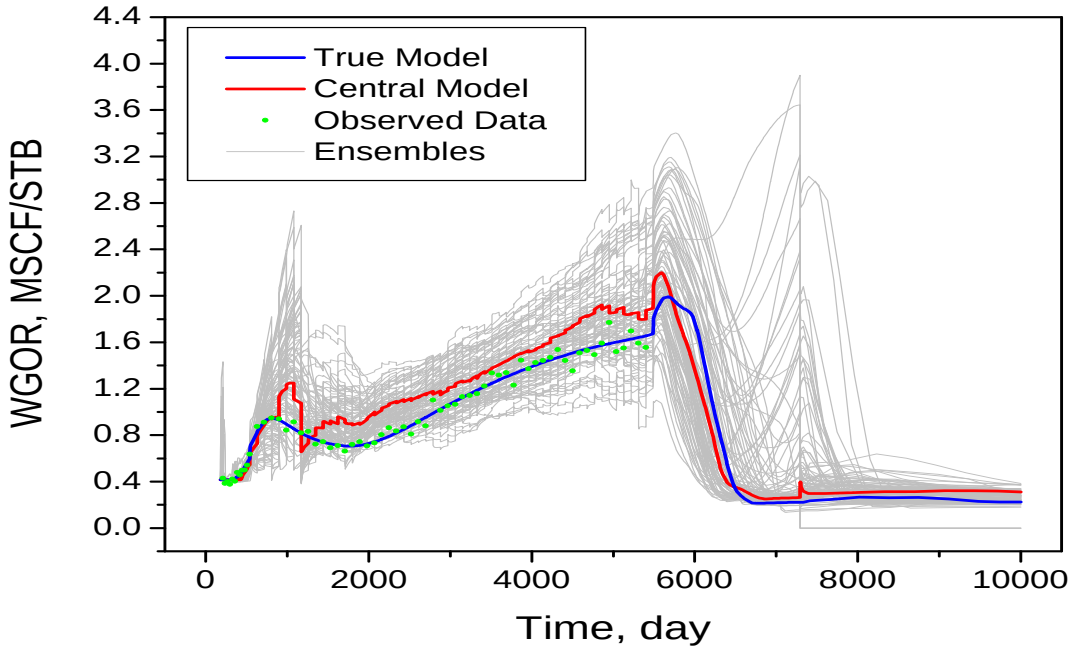


Figure 4.18: Production GOR for PRO-3, data assimilation for 5400 days, prediction to 10000 days, 2D synthetic problem.

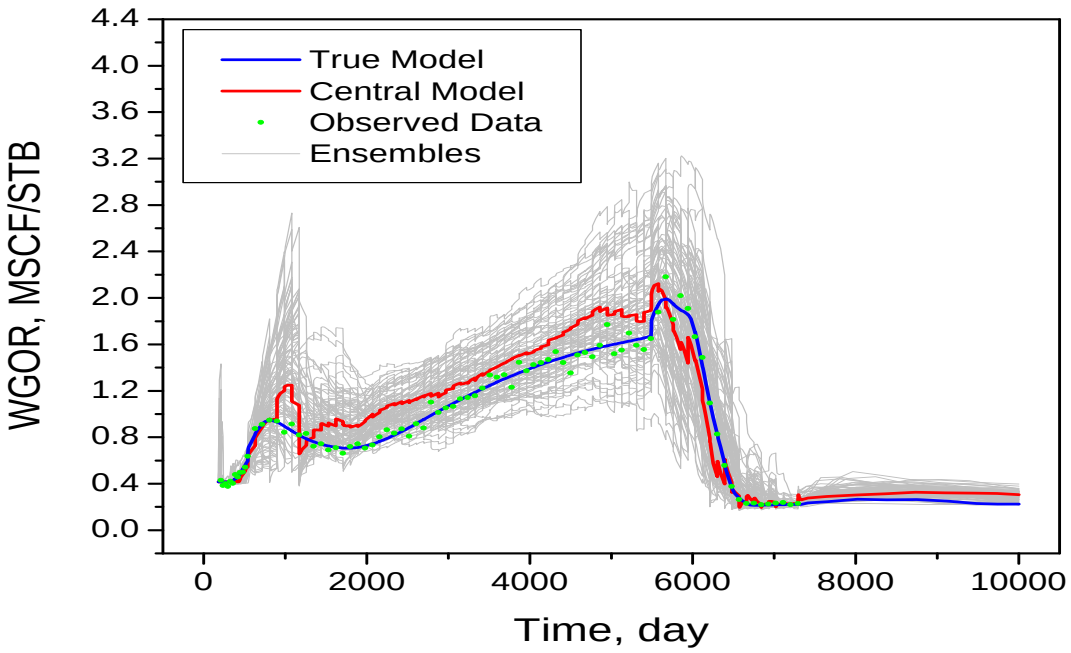


Figure 4.19: Production GOR for PRO-3, data assimilation for 7290 days, prediction to 10000 days, 2D synthetic problem.

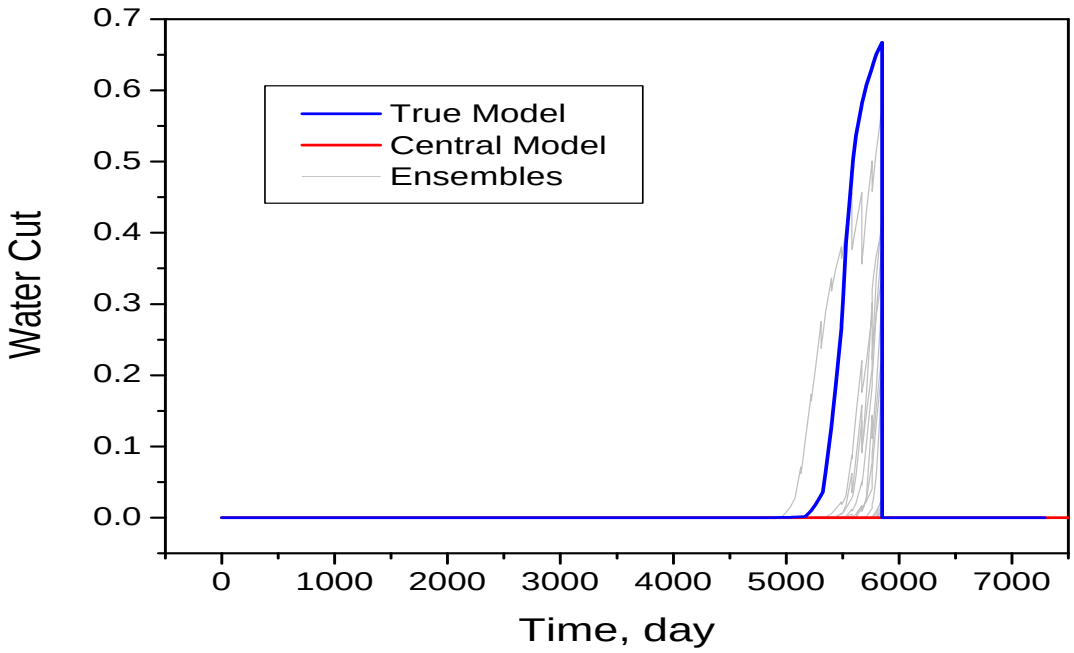


Figure 4.20: Water cut for PRO-5, data assimilation for 7290 days, 2D synthetic problem.

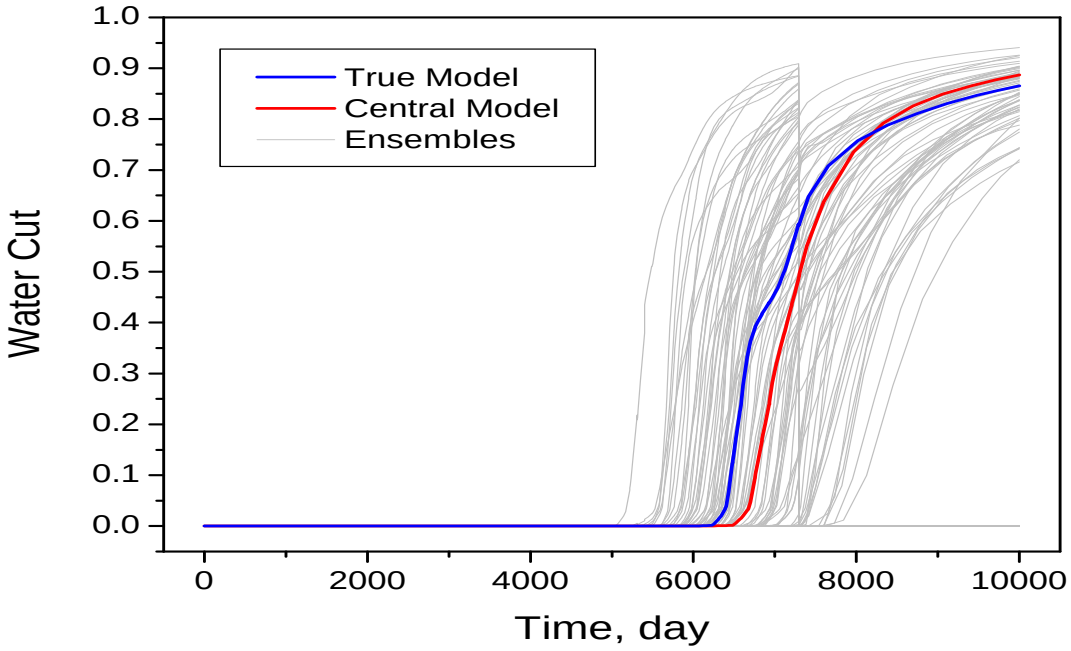


Figure 4.21: Water cut for PRO-2, data assimilation for 5400 days, prediction to 10000 days, 2D synthetic problem.

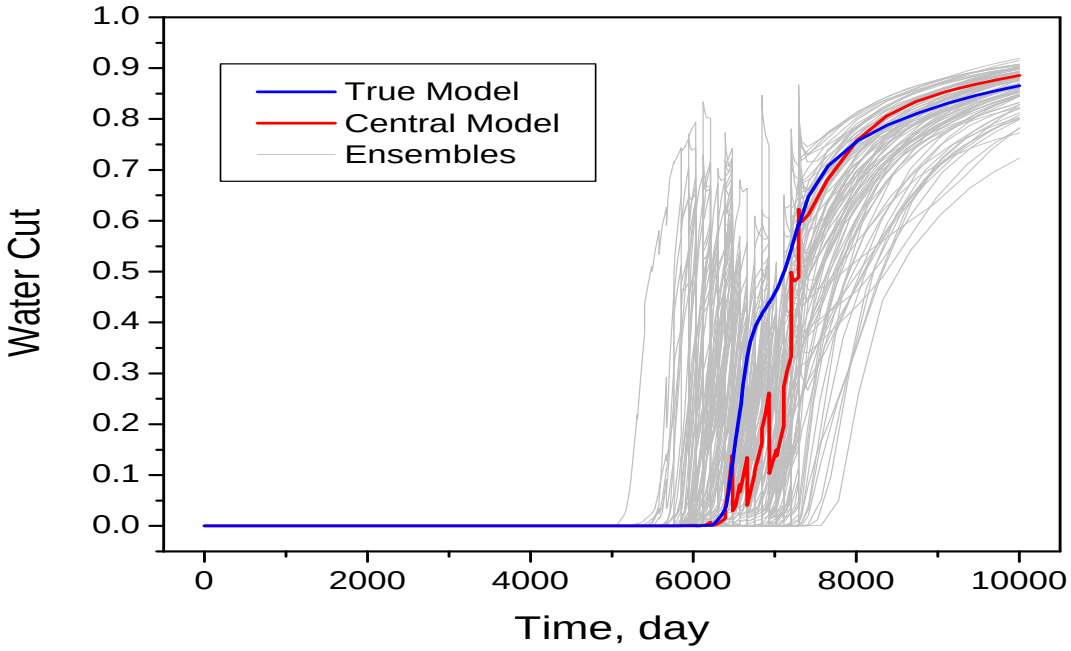


Figure 4.22: Water cut for PRO-2, data assimilation for 7290 days, prediction to 10000 days, 2D synthetic problem.

CHAPTER 5

PUNQ-S3 PROBLEM

5.1 Observed Data for PUNQ-S3

The PUNQ-S3 reservoir represents a synthetic model based on an actual North Sea reservoir, [11]. The problem was set up as a test case to allow various research groups to test their ability to characterize the uncertainty in reservoir performance predictions given some geologic information on the reservoir, hard data at well grid blocks and some scattered production data from the first eight years of production. Participants were asked to predict cumulative oil production for 16.5 years of total production and characterize the uncertainty in this prediction. The original simulation grid for the PUNQ-S3 problem consists of a $19 \times 28 \times 5$ grid with $\Delta x = \Delta y = 180$ meters and 1761 active gridblocks. The top structure map of the field, as shown in Fig. 5.1, shows that the field is bounded to the east and south by a fault, and links to the north and west to a fairly strong aquifer. A small gas cap is located in the center of the dome shaped structure. The field initially contains 6 production wells located around the gas-oil contact. Positions for 5 extra in-fill wells (X1-X5) were also defined. We do not consider these wells in this work.

5.1.1 Production Data

The production consists of a first year of extended well testing, followed by a three year shut-in period, and then 12.5 years of production with 14 days of shutin annually to collect buildup pressure data. The production rates for each well during the first year for each of four three-month periods are 629 (STB/day), 1258 (STB/day), 629 (STB/day) and 314.5 (STB/day) respectively. Subsequently, the production rate at each of the six wells was set to 943.5 (STB/day) except during shut in periods. However, this rate is a target rate, and if the bottom hole flowing pressure for a well falls below a limiting bottom

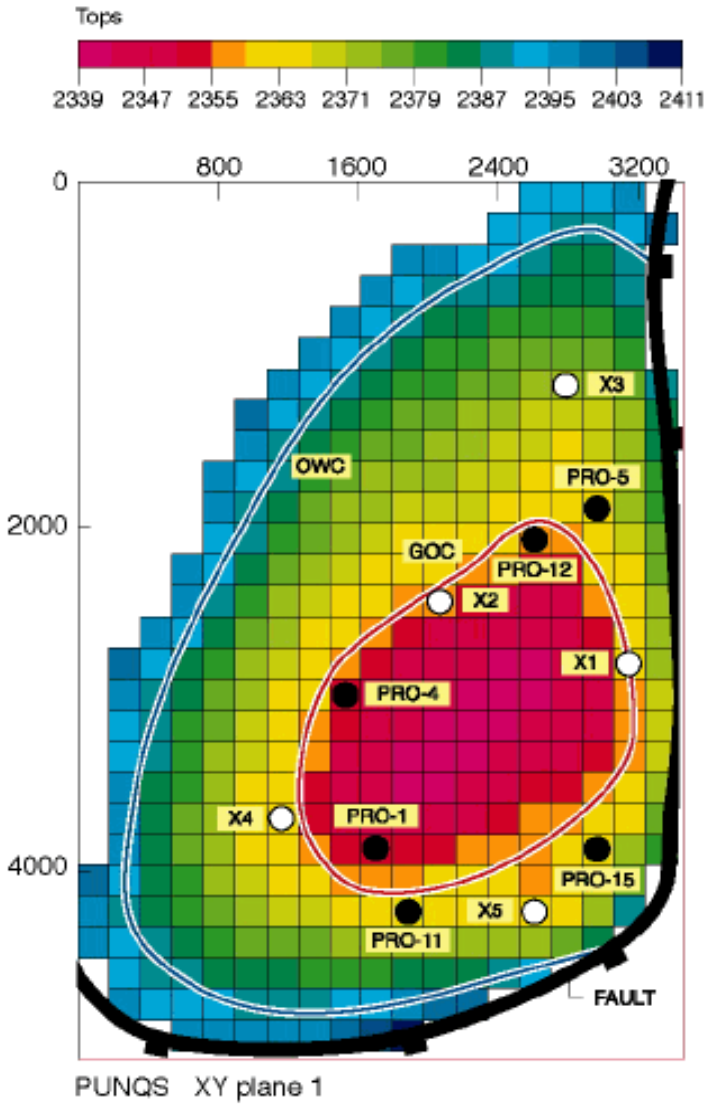


Figure 5.1: PUNQS3 structure map.

hole pressure of 1764 psi, the well is constrained to produce at a bottomhole pressure of 1764 psi. A complete description of how the truth was generated and how the true porosity, horizontal permeability and vertical permeability fields were used to generate the production data are available on the TNO web site (http://www.nitg.tno.nl/punq/). The true rock property fields were generated using sequential Gaussian co-simulation. Some of the key geostatistical parameters used to generate the truth are listed in Tables A-1 and A-2 of [2]. The true fields, or how they were generated, were unknown to the participants in the original PUNQ-S3 study. The production data generated consist of (i) 6 shut-in bottom hole pressure data in each well after 1, 4, 5, 6, 7 and 8 years production;

(ii) 8 flowing bottom hole pressure data in each well after 1 day, 3, 6, and 9 months of production, and immediately before the shut-in period after 5, 6, 7, and 8 years of production; (iii) 5 gas-oil ratio (GOR) data obtained after 4.5, 5, 5.5, 6.5 and 7.5 years of production in well PRO1, 4 GOR data after 5, 5.5, 6.5 and 7.5 years of production in well PRO2, and 4 GOR data after 5, 6, 7, and 8 years production in all other wells; (iv) 3 water-oil ratio (WOR) data after 7, 7.5, and 8 years of production in well PRO11 that experienced water breakthrough, and 1 WOR data after 8 years of production in all other wells. There are total of 84 bottom hole pressure data, 25 GOR data and 8 WOR data. These are the same number of data and collected at the same times under the same operating conditions as in the original PUNQ-S3 study [11, 2]. The standard deviations of the noise added to the true data are 1 bar (14.5 psi) for shut-in bottom hole pressures, 3 bars (43.5 psi) for flowing bottom hole pressures, 10% of the true values for GOR data less than 90 sm^3/sm^3 (or 505 scf/STB), 25% of the true values for GOR data greater than 90 sm^3/sm^3 (or 505 scf/STB), and the greater of 25% of the true value or 0.02 for WOR data.

5.1.2 Hard Data

The observed hard data at well gridblocks, $d_{\text{obs},h}$, including porosities, log horizontal permeabilities ($\ln(k)$) and log vertical permeabilities ($\ln(k_z)$), are shown in Table 5.1. They were generated by adding Gaussian noises with standard deviation equal to 15% of their true values. These hard data are used in both the RML method and the ensemble Kalman filter method. We should note that these data are not the same as used in the PUNQ-S3 study as in their results, the normalized hard data for porosity, horizontal log-permeability and vertical log-permeability were identical which is unreasonable unless only porosity is measured and the other hard data are based on a deterministic relation with porosity.

5.1.3 Data Assimilation and Production Prediction

For the EnKF method, 90 unconditional realizations of the prior were generated with sequential Gaussian co-simulation. The prior model is the same generated by [12].

Well	1	2	3	4	5	6
Layer 1-ϕ	0.0825	0.2298	0.2412	0.0807	0.0832	0.2535
Layer 1-ln(k)	3.6743	6.5140	6.0850	4.0519	3.5502	5.5903
Layer 1-ln(k_z)	2.2059	0.2298	6.1176	3.7693	3.0413	6.1981
Layer 2-ϕ	0.0631	0.0684	0.0716	0.0867	0.0954	0.1044
Layer 2-ln(k)	3.2129	3.0137	2.5084	2.7883	4.2405	3.8122
Layer 2-ln(k_z)	1.2633	0.9875	1.0094	1.8578	3.4015	2.4284
Layer 3-ϕ	0.1219	0.0995	0.2382	0.2887	0.0799	0.1521
Layer 3-ln(k)	4.6476	3.8683	5.7408	6.2875	4.2277	4.7780
Layer 3-ln(k_z)	2.5500	3.3104	5.4433	5.5541	3.2929	4.6214
Layer 4-ϕ	0.1618	0.1504	0.1660	0.1599	0.1484	0.1994
Layer 4-ln(k)	5.5268	6.5330	4.8816	4.9159	5.6424	6.1363
Layer 4-ln(k_z)	4.3646	3.6502	4.1680	3.1715	3.5704	3.8065
Layer 5-ϕ	0.2383	0.1625	0.0987	0.1271	0.2909	0.2418
Layer 5-ln(k)	5.4004	6.1618	3.0654	4.8846	6.0195	7.3068
Layer 5-ln(k_z)	5.6233	5.9546	2.5284	2.9698	5.5591	6.0218

Table 5.1: Observed hard data for PUNQ-S3.

Each of the realizations were truncated using the bounds specified by the minimum and maximum values and the realization hard data obtained by adding noise to the observed hard data of Table. 5.1 was also truncated based on these bounds. The truncated hard data was assimilated using the ensemble Kalman filter and the resulting updated ensembles were then truncated again. The resulting 90 ensembles provide the suite of 90 starting models for the assimilation of production data, which is done on a step by step basis using only one forward run of the reservoir simulator for each member of the ensemble. In the assimilation of production data, no truncation is done and no constraints are imposed. In the PUNQ-S3 example, at each well, a target oil rate is specified and a minimum bottomhole pressure is specified as a constraint. If the well can produce at the specified oil rate without falling below the minimum bottomhole pressure, the well produces at this rate. Otherwise, the well is produced at a constant bottomhole pressure equal to the minimum specified. This makes the problem more interesting. From the history matching viewpoint, we are trying to match observed bottomhole pressure, but when the simulator operates at a fixed bottomhole pressure, the bottomhole pressure ceases to be useful data because the predicted bottomhole pressure is insensitive to all model parameters. In the LBFGS algorithm, when this occurs at some iteration, we use the target oil rate as

pseudo-data as this enhances the rate of convergence and quickly yields a model which can produce at the specified rate while remaining above the allowable minimum bottomhole pressure. The difficulty caused by the change in the well's operating condition causes a more interesting problem when assimilating data by the ensemble Kalman filter as data are included in the state vector. If one uses wellbore pressure data in the state vector, but in the prediction step, the well can not produce at the specified oil rate and stay above the minimum bottomhole pressure, then the "predicted" bottomhole pressure is set equal to the specified minimum bottomhole pressure and thus no longer represents a prediction based on the model generated at the last time at which data was assimilated. Experiments we have done indicates that this tends to significantly diminish the reliability of the ensemble Kalman filter method. Perhaps because of the problem caused by a minimum bottomhole pressure constraint, [15] did not use only pressure, water cut and producing GOR data in their application of the Ensemble Kalman filter to the PUNQ problem, they also added rate "data" with a small amount of noise. We use a quite different procedure. We do not use rate data, but instead use only water cut, wellbore pressure and GOR data, but during the assimilation of production data, we set atmospheric pressure as the minimum bottomhole pressure. Effectively, this removes the minimum bottomhole pressure constraint. Thus, with rare exception, a well can meet the target rate and the wellbore pressure generated from the simulator does reflect a prediction based on the state vector at the previous assimilation step. In the final future prediction phase, after assimilation of data during the first 8 years, we of course use the constraints set by the originators of the PUNQ-S3 model to predict performance for the remaining 8.5 years so we have set of reservoir predictions for a 16.5 year producing period. In this problem, included in the state vector are gridblock pressures, water saturations, gas saturations and dissolved gas-oil ratios, gridblock porosities, horizontal and vertical log-permeabilities, and data (bottomhole pressure, GOR and water cut).

Figs. 5.2 and 5.3 show the bottom hole pressure and producing GOR in well PRO1 generated by ECLIPSE for the true model and all 90 ensembles after assimilation of both hard data and production data. The blue curves represent data from the true model.

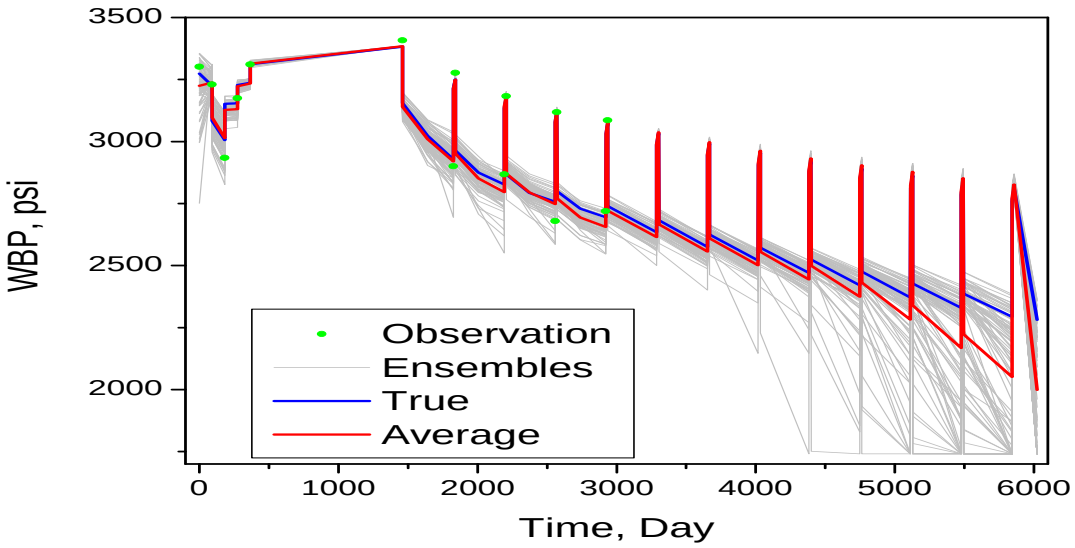


Figure 5.2: Pressure match, well PRO1, EnKF.

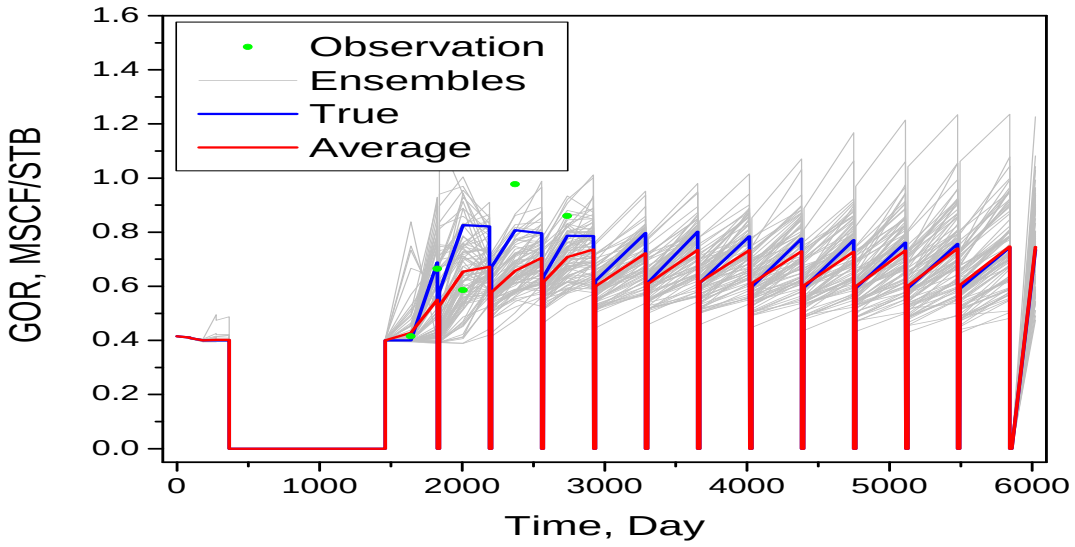


Figure 5.3: GOR match, well PRO1, EnKF.

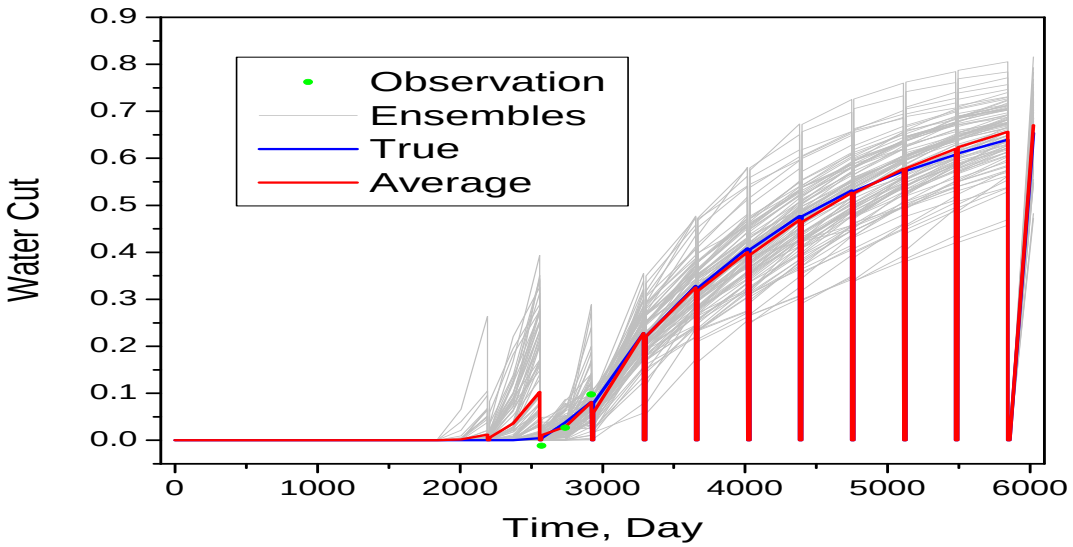
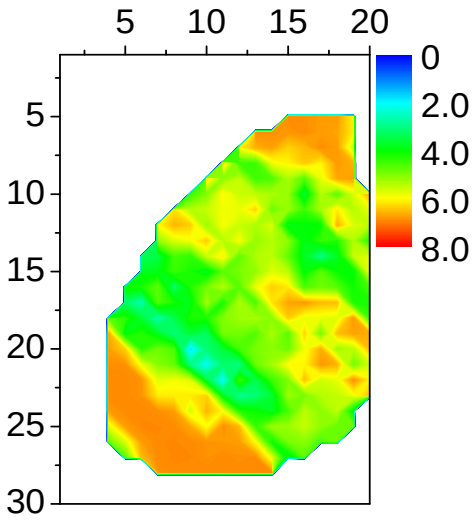


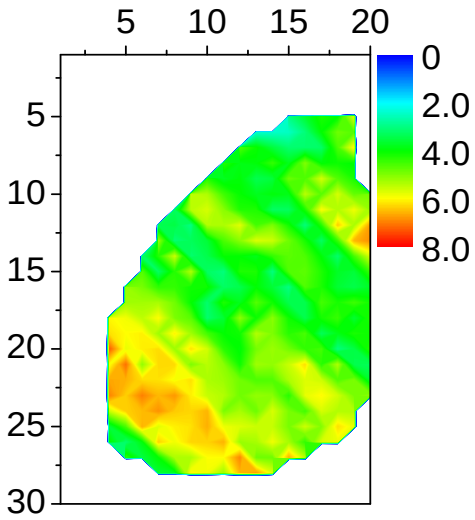
Figure 5.4: WOR match, well PRO11 EnKF.

The gray curves show the results from the ensembles during both the data assimilation period (first 8 years) and the subsequent 8.5 year prediction period. Red curves represent the average production data from all ensembles. The average bottom hole pressure is in good agreement with the true data until the late part of the prediction period, when the average of all ensemble bottomhole pressures falls below the truth. However, the true bottom hole pressure always is fairly near the center of predictions from the ensembles. In Fig. 5.3, the average of GOR data from the 90 ensembles is somewhat different from the truth in the time period following gas breakthrough, but as more data are assimilated, the average GOR of the ensembles becomes quite close to the truth. Fig. 5.4 shows the water cut in producing well PRO11. Before water breakthrough occurs for the true model at this well, some of ensembles predict water breakthrough, but later when the first water cut data (a low value) is encountered, almost all the ensembles give a close match of that data, and hence the early peak in the water cut data is diminished towards zero. During the prediction period the average water cut for the ensembles is very close to the truth.

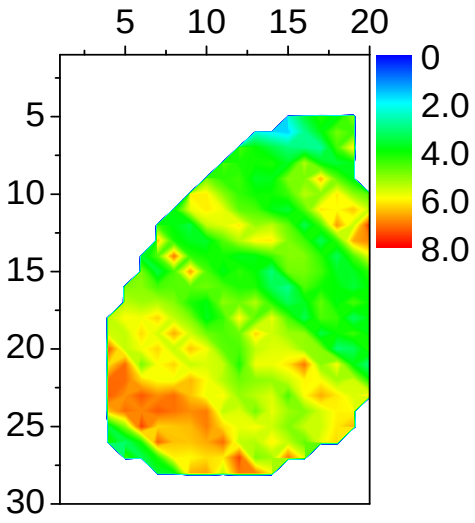
Fig. 5.5(a) shows the true logarithm of horizontal permeability and Fig. 5.5(b) shows the average model after assimilation of production data only, where Fig. 5.5(c) illustrate the central model after assimilation of production data. We can see the central



(a) True Model



(b) Mean Model



(c) Central Model

Figure 5.5: Comparison between true, mean and central model after assimilation of production data.

model and the mean model give reasonable agreement with the true model, although the central model represents the geological features better than mean model.

To complete this discussion instead of comparing the data predicted from the true

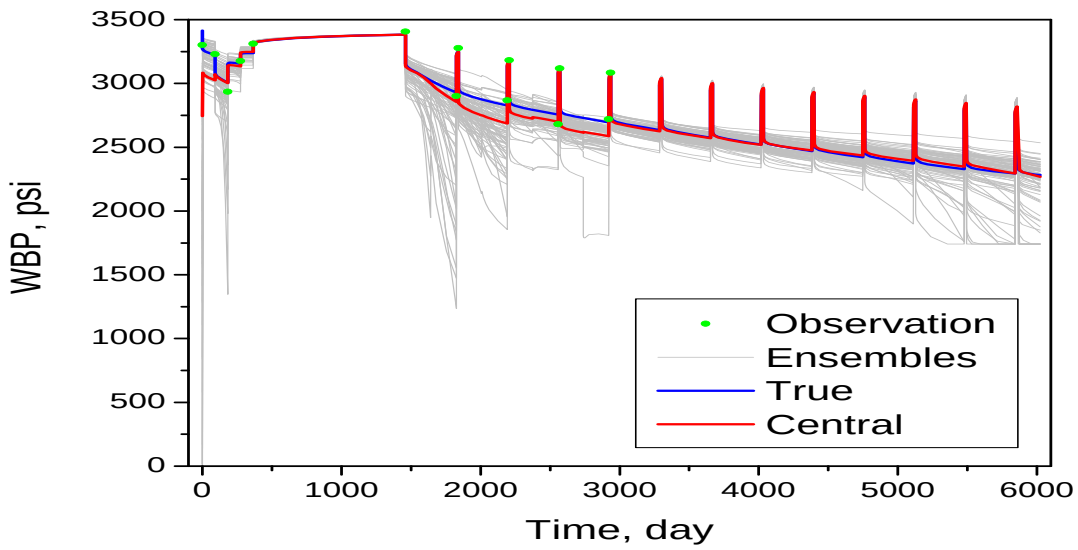


Figure 5.6: Pressure match,assimilation of production data, well PRO1, EnKF.

model with the average of predicted data from each ensemble, in Figs. 5.6 through 5.8, we compare the data predicted from the true model with the data predicted from the central model. Fig. 5.6 shows the bottom hole pressure in well PRO1 for the true model and all 90 ensembles after assimilation of only production data. Fig. 5.7 illustrate the GOR data in well PRO4. The blue curves represent data predicted from the true model. The gray curves show the results from the ensembles during both the data assimilation period (first 8 years) and the subsequent 8.5 year prediction period. Red curves represent the predicted production data from the central model. The jumps in the red and gray curves occur at assimilation times at which we update the solution variables in the state vector.

5.1.4 Reservoir Performance Predictions

Fig. 5.9(a) shows the cumulative oil production performances predicted with 10 unconditional realizations. The blue curve in Fig. 5.9(a) represents the prediction generated with the true model. The results of Fig. 5.9(a) indicate that the cumulative oil production predicted with the unconditional realizations are biased; Fig. 5.9(b) show the cumulative oil productions predicted with 10 conditional realizations obtained by conditioning to

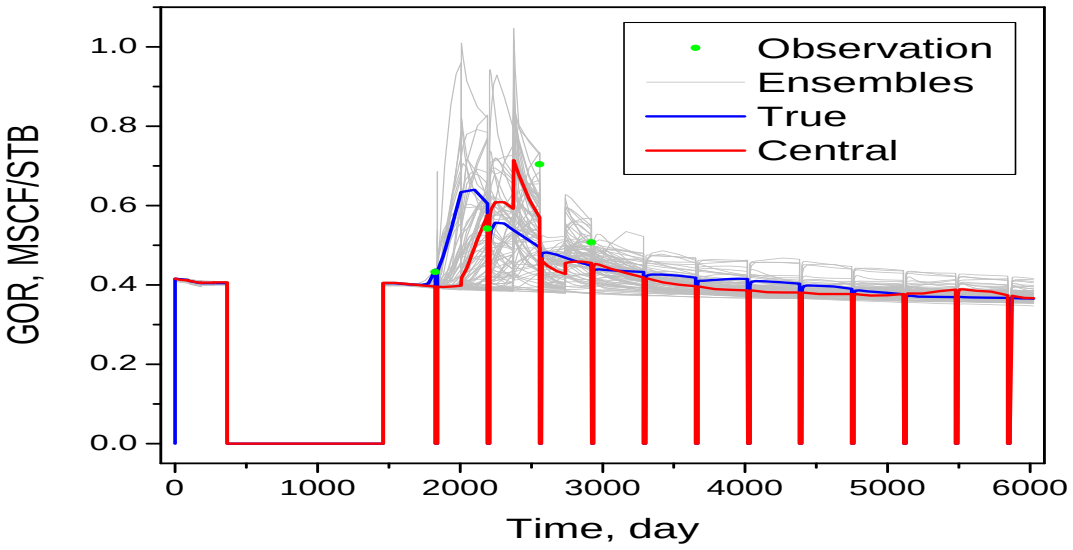


Figure 5.7: GOR match, assimilation of production data, well PRO4, EnKF.

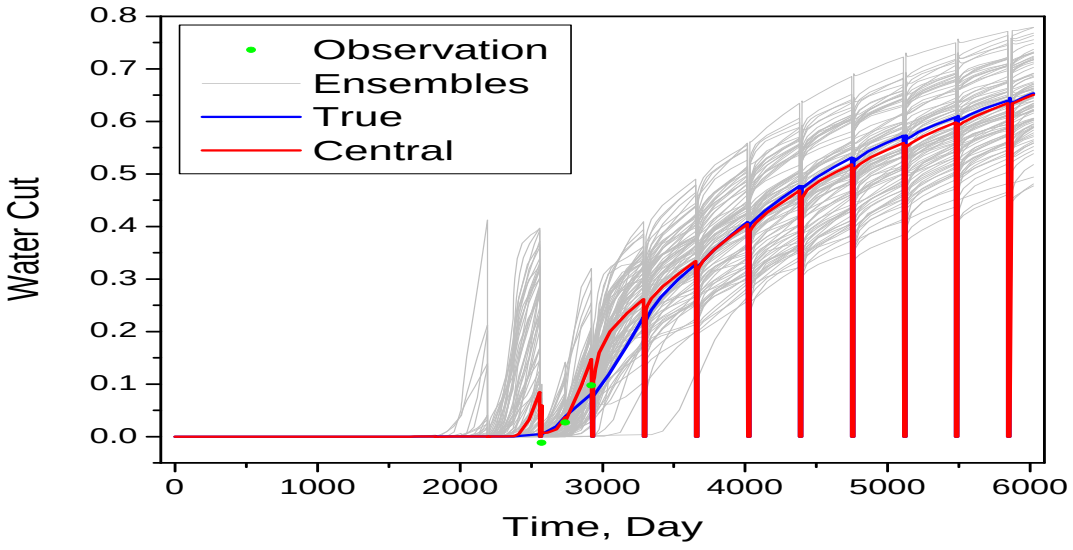
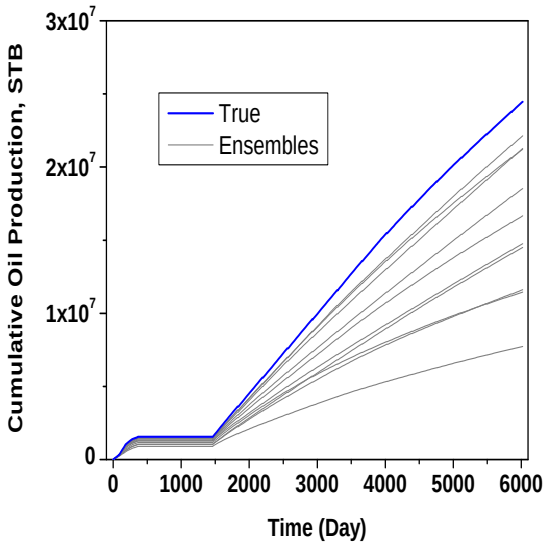
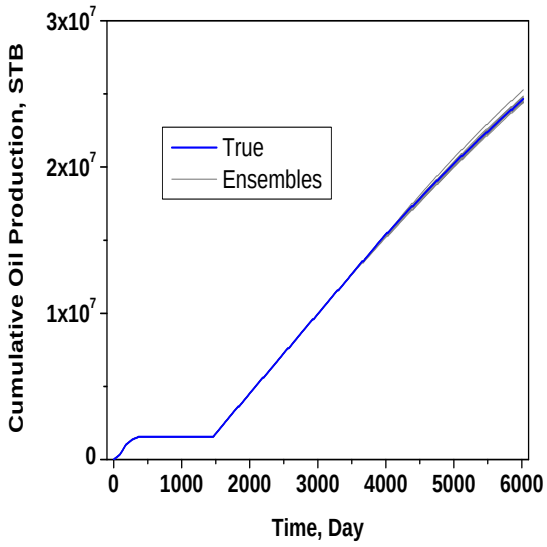


Figure 5.8: WOR match, assimilation of production data, well PRO11, EnKF.

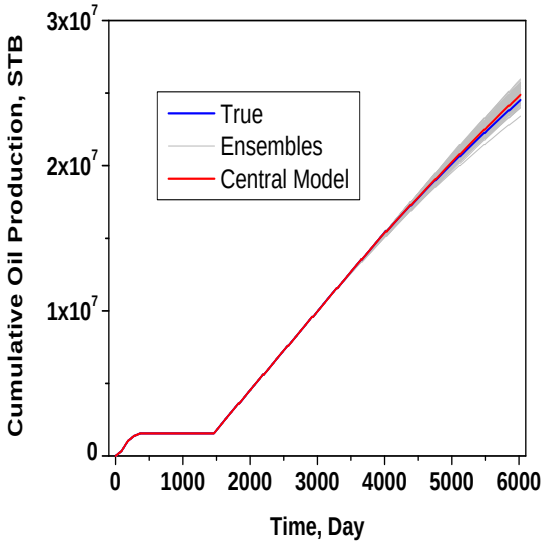
both hard data and production data obtained with the RML method. Fig. 5.9(c) shows corresponding results generated with the 90 ensembles used in the EnKF method. The truth (blue curves) lies within the band of predictions generated with both the RML method and the EnKF method; i.e., both methods give a reasonable characterization of uncertainty. Note the uncertainty in predicted reservoir performance is quite small, especially, compared to the uncertainty in predictions generated with unconditional realizations. Similar results are observed for the cumulative gas and water production, see Figs. 5.10(a) through 5.11(c).



(a) Unconditional

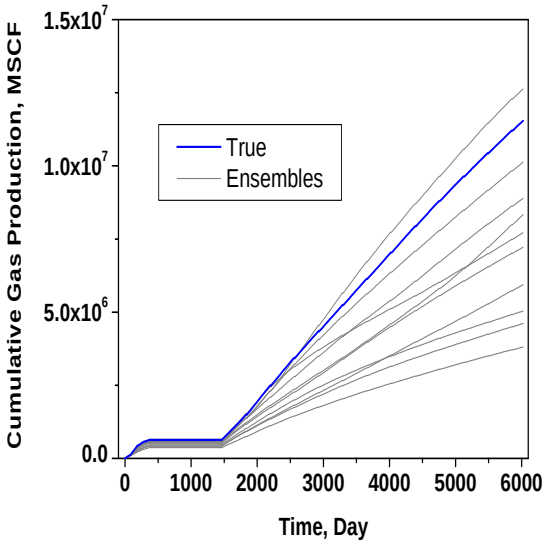


(b) RML

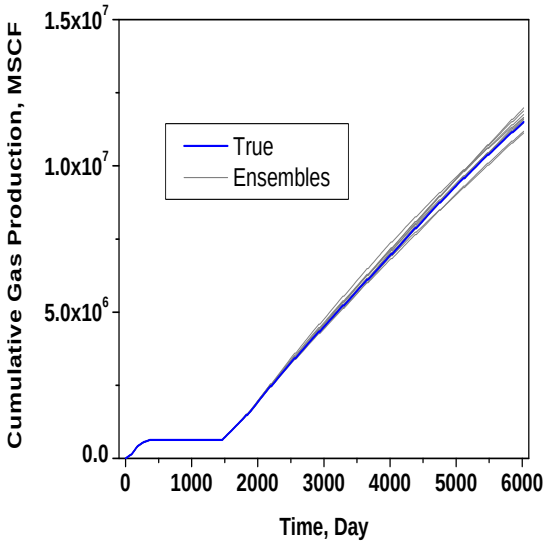


(c) EnKF

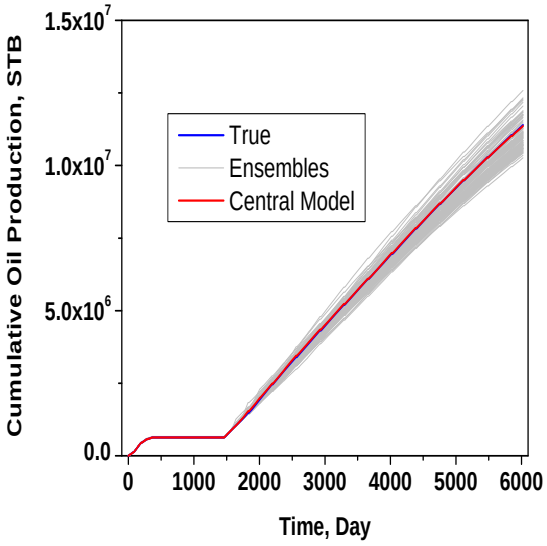
Figure 5.9: Comparison between RML and EnKF in reservoir performance prediction.



(a) Unconditional

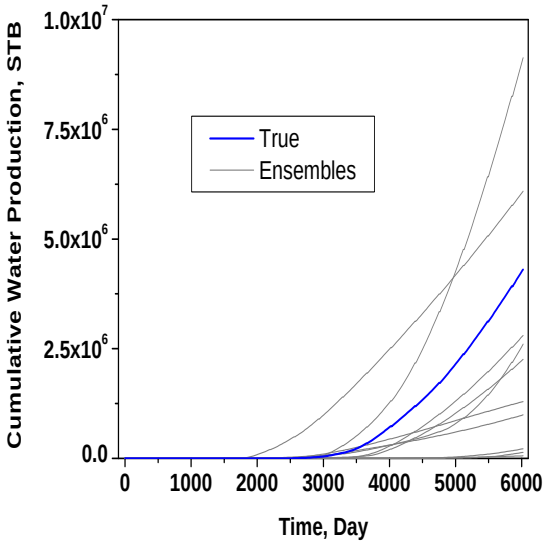


(b) RML

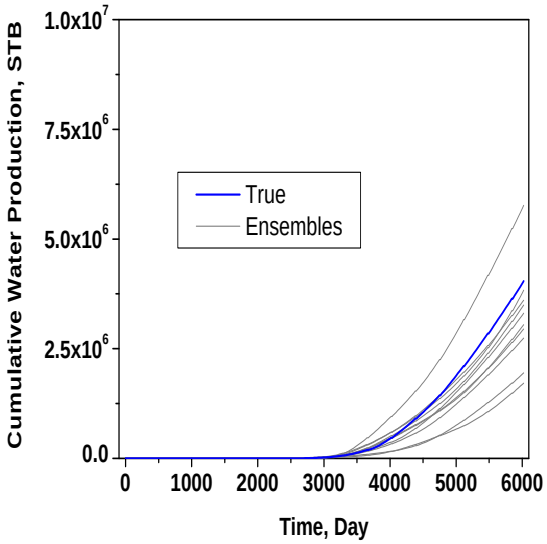


(c) EnKF

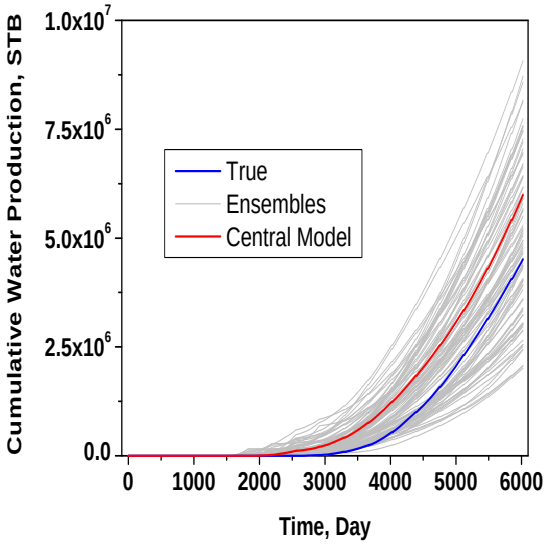
Figure 5.10: Comparison between RML and EnKF in cumulative gas production prediction.



(a) Unconditional



(b) RML



(c) EnKF

Figure 5.11: Comparison between RML and EnKF in cumulative water production prediction.

CHAPTER 6

DISCUSSIONS AND CONCLUSIONS

Any method used to integrate dynamic data into reservoir description should exhibit several desirable characteristics; the method should be (i) computationally robust and efficient, (ii) able to give a match of the dynamic data that is consistent with the noise level in the data, (iii) able to preserve the underlying geology, (iv) able to correctly characterize the uncertainty in performance predictions, (v) able to characterize the uncertainty in reservoir description with regard to the distribution of rock properties and important geologic features such as faults and boundaries between facies, (vi) able to characterize the uncertainty in the distribution of fluids which is critical for infill drilling programs and (vii) be available in or easily incorporated into software. Many methods satisfy well one or more or these characteristics, but do a poor job of satisfying others. Based on our study, the ensemble Kalman filter (EnKF) satisfies (i) through (vii), but as this is a relatively new method, it is not known whether this result is general.

Although the ensemble mean is frequently used as estimation of the truth, we have shown that the deviation of this mean from the truth does not give a reliable metric for quantifying how well the posterior pdf is characterized by the distribution of the ensembles.

As noted above, the EnKF algorithm can be quickly coupled with any reservoir simulator and the ensemble Kalman filter method has the advantage of simplicity in terms of its implementation, and is more flexible to use and easily adapted to diverse applications. However, if the number of data to be assimilated at a particular time step becomes large as in seismic data, the overhead cost of the numerical linear algebra required in the EnKF method will be significant.

The Gaussian assumption of EnKF is very critical. Even for large number of

ensembles, in some of the multi-modal problems, EnKF can not sample the posterior pdf correctly. For the linear case, we showed that EnKF samples correctly from the posterior pdf as the number of ensembles goes to infinity and is equivalent to randomized maximum likelihood method. For two toy problems discussed, we found that the central model is not a good estimate of the true model, but for the 2D synthetic problem the data predicted from the central model is in a good agreement with the truth and the central model provide a reasonable approximation of future performance prediction. This id also true for the well known PUNQ-S3 problem. Although during the late time of data assimilation the model parameters start to diverge from the truth and many of the ensembles over predict the ultimate oil recovery, the final prediction performance is still reasonable.

For the linear case, we have shown that rerunning the time step to recompute time dependent parameters is inappropriate.

Although, we have focused on problems where EnKF encounters some difficulties, it is likely that modifications will be found to further improve its reliability. Moreover, the advantage of the method in terms of computational efficiency and simplicity in implementation, suggest that it will become widely used in practice even if it provides a less than perfect characterization of variability in reservoir variables, fluid distributions and performance predictions.

Relative to the original motivation for this study, we conclude that, at least for problems similar to real reservoir problems, the EnKF method gives a reliable assessment of the uncertainty in reservoir performance predictions.

BIBLIOGRAPHY

[1] Jeffrey L. Anderson and Stephen L. Anderson. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, pages 2741–2758, 1999.

[2] John W. Barker, Maarten Cuypers, and Lars Holden. Quantifying uncertainty in production forecasts: Another look at the PUNQ-S3 problem. *SPE Journal*, 6(4):433–441, 2001.

[3] Luciane Bonet-Cunha, D. S. Oliver, R. A. Rednar, and A. C. Reynolds. A hybrid Markov chain Monte Carlo method for generating permeability fields conditioned to multiwell pressure data and prior information. *SPE Journal*, 3(3):261–271, 1998.

[4] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126:1719–1724, 1998.

[5] Guy M. Chavent, M. Dupuy, and P. Lemonnier. History matching by use of optimal control theory. *Soc. Petrol. Eng. J.*, 15(1):74–86, 1975.

[6] W. H. Chen, G. R. Gavalas, John H. Seinfeld, and Mel L. Wasserman. A new algorithm for automatic history matching. *Soc. Petrol. Eng. J.*, pages 593–608, 1974.

[7] Lifu Chu, Albert C. Reynolds, and Dean S. Oliver. Computation of sensitivity coefficients for conditioning the permeability field to well-test data. *In Situ*, 19(2):179–223, 1995.

[8] Yannong Dong and Dean S. Oliver. Quantitative use of 4D seismic data for reservoir characterization (SPE-84571). In *Proceedings of 2003 SPE Annual Technical Conference and Exhibition*, 2003.

[9] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99:10143–10162, 1994.

[10] Geir Evensen. The combined parameter and state estimation problem. *Computational Geosciences*, page submitted, 2005.

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

[12] Guohua Gao, Mohammad Zafari, and A. C. Reynolds. Quantifying uncertainty for the PUNQ-S3 problem in a Bayesian setting with RML and EnKF (SPE-93324). In *SPE Reservoir Simulation Symposium*, page 19, 2005.

[13] J. Jaime Gómez-Hernández and André G. Journel. Joint sequential simulation of multigaussian fields. In A. Soares, editor, *Geostatistic Troia 92*, pages 133–144. 1992.

[14] Yaqing Gu and Dean S. Oliver. The ensemble Kalman filter for continuous updating of reservoir simulation models. *TUPREP Research Report 21*, pages 150–164, 2004.

[15] Yaqing Gu and Dean S. Oliver. History matching of the punq-s3 reservoir model using the ensemble Kalman filter. *SPE-89942*, 2004.

[16] Thomas M. Hamill and Jeffrey S. Whitaker. Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Monthly Weather Review*, pages 2776–2790, 2001.

[17] N. He, A. C. Reynolds, and D. S. Oliver. Three-dimensional reservoir description from multiwell pressure data and prior information. *Soc. Pet. Eng. J.*, pages 312–327, 1997.

[18] P. L. Houtekamer and Herschel L. Mitchell. Data assimilation using an ensemble kalman filter technique. *Monthly Weather Review*, pages 796–811, 1998.

[19] Peter K. Kitanidis. Quasi-linear geostatistical theory for inversing. *Water Resour. Res.*, 31(10):2411–2419, 1995.

[20] Peter Jan Van Leeuwen. Comment on data assimilation using an ensemble kalman filter technique. *Monthly Weather Review*, pages 1374–1377, 1999.

[21] Ruijian Li, A. C. Reynolds, and D. S. Oliver. History matching of three-phase flow production data. *SPE J.*, 8(4):328–340, 2003.

[22] Ning Liu and Dean S. Oliver. Evaluation of Monte Carlo methods for assessing uncertainty. *SPE Journal*, 8(2):188–195, 2003.

[23] G. Naevdal, L. M. Johnsen, S. I. Aanonsen, and E. H. Vefring. Reservoir monitoring and continuous model updating using ensemble Kalman filter (SPE-84372). In *2003 SPE Annual Technical Conference and Exhibition*, 2003.

[24] G. Naevdal, T. Mannseth, and E. H. Vefring. Near-well reservoir monitoring through ensemble Kalman filter (SPE-75235). In *Proceeding of SPE/DOE Improved Oil Recovery Symposium*, 2002.

[25] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, 1999.

[26] Dean S. Oliver. On conditional simulation to inaccurate data. *Math. Geology*, 28(6):811–817, 1996.

[27] Dean S. Oliver, Nanqun He, and Albert C. Reynolds. Conditioning permeability fields to pressure data. In *European Conference for the Mathematics of Oil Recovery, V*, pages 1–11, 1996.

[28] Albert C. Reynolds, Nanqun He, and Dean S. Oliver. Reducing uncertainty in geostatistical description with well testing pressure data. In Richard A. Schatzinger and John F. Jordan, editors, *Reservoir Characterization—Recent Advances*, pages 149–162. American Association of Petroleum Geologists, 1999.

[29] Albert C. Reynolds, Nanqun He, and Dean S. Oliver. Reducing uncertainty in geostatistical description with well testing pressure data. In *Fourth International Reservoir Characterization Technical Conference*, 2–4 March 1997.

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

[31] Hans Wackernagel. *Multivariate Geostatistics*. Springer, Germany, 1995.

[32] Jeffrey S. Whitaker and Thomas M. Hamill. Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, pages 1913–1924, 2002.

[33] Mohammad Zafari and Albert C. Reynolds. Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter. *SPE-95750*, 2005.

[34] Mohammad Zafari and Albert C. Reynolds. Enkf versus rml, theoretical notes and computational experiments. *TUPREP Research Report 22*, pages 195–228, 2005.

[35] F. Zhang and A. C. Reynolds. Optimization algorithms for automatic history matching of production data. *Proceedings of 8th European Conference on the Mathematics of Oil Recovery*, pages 1–10, 2002.