

UNIVERSITY OF OKLAHOMA  
GRADUATE COLLEGE

HISTORY MATCHING PRODUCTION DATA USING  
THE ENSEMBLE KALMAN FILTER

A DISSERTATION  
SUBMITTED TO THE GRADUATE FACULTY  
in partial fulfillment of the requirements for the  
degree of  
Doctor of Philosophy

By  
YAQING GU  
Norman, Oklahoma  
2006

HISTORY MATCHING PRODUCTION DATA USING  
THE ENSEMBLE KALMAN FILTER

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

BY

---

Dean S. Oliver, Committee Chair

---

Dongxiao Zhang

---

Faruk Civan

---

Chandra S. Rai

---

S. LakshmiVarahan

© Copyright by YAQING GU 2006  
All Rights Reserved.

## ACKNOWLEDGEMENTS

First of all, I want to express my sincerest appreciation to Dean S. Oliver, Eberly Chair Professor, who has been supervising my research progress with invaluable guidance and providing me with generous financial support during the past four years in OU; and to his wife, Mary Oliver, who makes me find Norman more like my second hometown.

I appreciate very much that Dongxiao Zhang, Miller Chair Professor, Faruk Civan, Alumni Professor, Chandra Rai, Mewbourne Chair Professor, and S. Lakshmivarahan, George Lynn Cross Research Professor, serve on my committee. Their insightful suggestions and comments definitely increase the quality of my dissertation.

Special thanks to Henry Neeman, Director of OU Supercomputing Center for Education and Research (OSCER), and Brandon George, Manager of Operations of OSCER for their always-available and expert-level helps whenever I come across difficulties.

I would like to extend my thanks to all the faculty members in the Mewbourne School of Petroleum and Geological Engineering (MPGE), who provided me opportunities to broaden my view beyond my dissertation; to all staff members in MPGE, whose hard work makes the school run as smoothly as possible; to all the fellow students in my research group for their encouragements and help; to all my classmates and friends in MPGE.

Also, I want to thank Chevron for providing me two internship opportunities. The helpful discussions with Wen H. Chen, Xian-Huan Wen, and Yuguang Chen during these internships are gratefully acknowledged. Special thanks to Alan Bernath for his availability and assistance when I had questions.

Licenses for the **ECLIPSE** Black Oil Reservoir Simulator were provided by Schlumberger, for **CHEARS** were provided by Chevron.

I am truly thankful for the dear friends who are especially encouraging to me: Ken and Karen Winfrey, Beau and Mary Ann Jennings, Jerry and Ann Lout, Ryan and Lori Lawrence, Barbara Fehervari, Xiaolu Wang, Jim and Jerry White, Doug and Karen Windol, Bill and Pam Kennedy, Jerry and Susan McCaskill, and Nella White.

I am deeply indebted to my husband, Yannong Dong, for his unfailing love, tremendous support, and strong belief in me during hard times.

# TABLE OF CONTENTS

|                                                                                                                          |      |
|--------------------------------------------------------------------------------------------------------------------------|------|
| ACKNOWLEDGEMENTS                                                                                                         | iv   |
| LIST OF TABLES                                                                                                           | ix   |
| LIST OF FIGURES                                                                                                          | xi   |
| ABSTRACT                                                                                                                 | xvii |
| I INTRODUCTION                                                                                                           | 1    |
| II THE ENSEMBLE KALMAN FILTER                                                                                            | 8    |
| 2.1 Background and Terminology . . . . .                                                                                 | 8    |
| 2.1.1 Reservoir flow and transport equations . . . . .                                                                   | 8    |
| 2.1.2 History matching terminology . . . . .                                                                             | 10   |
| 2.2 Outline of the Kalman Filter Algorithm . . . . .                                                                     | 13   |
| 2.3 The Ensemble Kalman Filter . . . . .                                                                                 | 16   |
| 2.3.1 State vectors for reservoir models . . . . .                                                                       | 18   |
| 2.3.2 Observations for ensemble models . . . . .                                                                         | 19   |
| 2.3.3 Forecast step for reservoir models . . . . .                                                                       | 19   |
| 2.3.4 Update step . . . . .                                                                                              | 21   |
| 2.4 Implementation of the EnKF . . . . .                                                                                 | 22   |
| 2.5 Parallel Implementation of the EnKF . . . . .                                                                        | 26   |
| 2.6 Heuristics Measures of the EnKF Performance . . . . .                                                                | 28   |
| III INVESTIGATION OF THE ENSEMBLE KALMAN FILTER TO<br>APPLICATIONS OF CONTINUOUS RESERVOIR SIMULATION<br>MODELS UPDATING | 31   |
| 3.1 Introduction . . . . .                                                                                               | 31   |
| 3.2 Example 1: One-dimensional Waterflood Problem . . . . .                                                              | 32   |
| 3.2.1 Generation of initial reservoir models . . . . .                                                                   | 33   |
| 3.2.2 Solving problems with updated saturations . . . . .                                                                | 35   |
| 3.2.3 Goodness of the model parameter estimates . . . . .                                                                | 43   |

|           |                                                                                                      |           |
|-----------|------------------------------------------------------------------------------------------------------|-----------|
| 3.3       | Example 2: Two-dimensional Waterflood Problem . . . . .                                              | 43        |
| 3.3.1     | History matching results . . . . .                                                                   | 44        |
| 3.3.2     | Model parameters estimates . . . . .                                                                 | 45        |
| 3.3.3     | Consistency check . . . . .                                                                          | 49        |
| 3.4       | Discussion . . . . .                                                                                 | 51        |
| <b>IV</b> | <b>HISTORY MATCHING OF THE PUNQ-S3 RESERVOIR SIMULATION MODEL USING THE ENSEMBLE KALMAN FILTER</b>   | <b>52</b> |
| 4.1       | Problem Description . . . . .                                                                        | 52        |
| 4.2       | Generation of Initial Realizations . . . . .                                                         | 54        |
| 4.3       | Production Data . . . . .                                                                            | 55        |
| 4.4       | Production Data Matched . . . . .                                                                    | 60        |
| 4.5       | Prediction of Total Oil Production . . . . .                                                         | 62        |
| 4.6       | Porosity and Permeability Estimates . . . . .                                                        | 62        |
| 4.7       | Discussion . . . . .                                                                                 | 64        |
| <b>V</b>  | <b>INVESTIGATION OF ITERATIVE SCHEMES HANDLING NON-LINEARITY BASED ON THE ENSEMBLE KALMAN FILTER</b> | <b>78</b> |
| 5.1       | Possible Problems of the EnKF . . . . .                                                              | 78        |
| 5.1.1     | Wen and Chen's remedy . . . . .                                                                      | 79        |
| 5.2       | Ensemble Randomized Maximum Likelihood Filter . . . . .                                              | 80        |
| 5.2.1     | Linear dynamic system . . . . .                                                                      | 81        |
| 5.2.2     | Non-linear dynamic system . . . . .                                                                  | 82        |
| 5.2.3     | Implementation and computational cost of the EnRMLF . .                                              | 84        |
| 5.2.4     | Computation of sensitivity coefficient matrix . . . . .                                              | 86        |
| 5.2.5     | EnRMLF with iteration when needed . . . . .                                                          | 88        |
| 5.3       | A Linear Example . . . . .                                                                           | 89        |
| 5.3.1     | Problem description . . . . .                                                                        | 89        |
| 5.3.2     | Observations . . . . .                                                                               | 91        |
| 5.3.3     | Criteria used to stop iterations . . . . .                                                           | 93        |
| 5.3.4     | Comparison of the two iterative methods . . . . .                                                    | 94        |

|                   |                                                                                           |            |
|-------------------|-------------------------------------------------------------------------------------------|------------|
| 5.3.5             | Computed sensitivity coefficients . . . . .                                               | 97         |
| 5.4               | A Non-linear Example . . . . .                                                            | 104        |
| 5.4.1             | Problem description . . . . .                                                             | 104        |
| 5.4.2             | Observations . . . . .                                                                    | 105        |
| 5.4.3             | Case I: first data assimilated before water breaks through the observation well . . . . . | 106        |
| 5.4.4             | Summary for Case I . . . . .                                                              | 113        |
| 5.4.5             | Case II: first data assimilated after water breaks through the observation well . . . . . | 114        |
| 5.4.6             | Summary for Case II . . . . .                                                             | 116        |
| 5.5               | Discussion . . . . .                                                                      | 116        |
| <b>VI</b>         | <b>A FIELD CASE STUDY WITH THE ENSEMBLE KALMAN FILTER</b>                                 | <b>137</b> |
| 6.1               | Field Description . . . . .                                                               | 137        |
| 6.2               | Production Data . . . . .                                                                 | 139        |
| 6.2.1             | Data pre-processing . . . . .                                                             | 139        |
| 6.2.2             | Well constraint and computed data sensitivity from ensemble                               | 140        |
| 6.2.3             | Data selection criterion . . . . .                                                        | 141        |
| 6.3               | Model Parameters . . . . .                                                                | 142        |
| 6.4               | Result Analysis . . . . .                                                                 | 143        |
| 6.4.1             | History matching results . . . . .                                                        | 143        |
| 6.4.2             | Prediction results . . . . .                                                              | 146        |
| 6.4.3             | Evolution of estimated permeability field . . . . .                                       | 147        |
| 6.5               | Discussion . . . . .                                                                      | 148        |
| <b>VII</b>        | <b>CONCLUSIONS</b>                                                                        | <b>159</b> |
| <b>REFERENCES</b> |                                                                                           | <b>161</b> |

## LIST OF TABLES

|     |                                                                                                                                                                                                               |     |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 2.1 | Summary of the two-step procedure of the Kalman filter at measurement time $t_k$ . . . . .                                                                                                                    | 15  |
| 3.1 | Parameters used for generating initial reservoir models for the one-dimensional waterflood problem. . . . .                                                                                                   | 33  |
| 3.2 | Prior water saturation versus distance. . . . .                                                                                                                                                               | 39  |
| 3.3 | Prior water saturation versus updated distance. . . . .                                                                                                                                                       | 39  |
| 3.4 | Interpolated water saturation versus distance. The interpolation is based on Table 3.3. . . . .                                                                                                               | 39  |
| 4.1 | Well locations of the six producers and perforated layers of each well.                                                                                                                                       | 54  |
| 4.2 | Parameters of spherical variogram models used to generate initial realizations for the five layers of the reservoir. . . . .                                                                                  | 55  |
| 4.3 | Mean and standard deviation values of the initial realizations for different layers (all cells have same format: mean/std) . . . . .                                                                          | 55  |
| 4.4 | The kinds and numbers of data available at different measurement times.                                                                                                                                       | 58  |
| 4.5 | Standard deviations of the noises added to the synthetic data computed from the true reservoir simulation model. . . . .                                                                                      | 59  |
| 5.1 | Constant parameters related to the calculation of the tracer arrival time for the linear example. . . . .                                                                                                     | 90  |
| 5.2 | Parameters used to generate the initial realizations of the porosity distribution for the linear example. . . . .                                                                                             | 91  |
| 5.3 | The uncontaminated synthetic observations, computed using the true porosity values, and the standard deviations of the unbiased Gaussian measurement errors, at the four measurement times/locations. . . . . | 92  |
| 5.4 | Final data mismatch, number of iterations applied, and criteria satisfied when iterations are terminated, for different methods at all measurement times with 30 ensemble members for the linear example. . . | 95  |
| 5.5 | Parameters used to generate initial realizations for the non-linear example. . . . .                                                                                                                          | 105 |
| 5.6 | Synthetic data computed from the selected true reservoir model before adding measurement noise at 8 measurement times. . . . .                                                                                | 107 |
| 5.7 | Computational cost for the three iterative methods at different measurement times for Case I. . . . .                                                                                                         | 108 |

|                                                                                                            |     |
|------------------------------------------------------------------------------------------------------------|-----|
| 5.8 Computational cost for the three iterative methods at different measurement times for Case II. . . . . | 115 |
| 6.1 Data processing schemes used and their advantages and disadvantages. . . . .                           | 141 |

## LIST OF FIGURES

|      |                                                                                                                                                                                    |    |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.1  | Schematic setup for the one-dimensional waterflood problem. Note: BHP means bottom-hole pressure, WIR means water injection rate, OPR/WPR means oil/water production rate. . . . . | 33 |
| 3.2  | Water saturation profiles and histogram before the first application of the Kalman correction at 110 days. . . . .                                                                 | 36 |
| 3.3  | Water saturation profiles and histogram after the first application of the Kalman correction at 110 days. . . . .                                                                  | 36 |
| 3.4  | Normal score transformed profiles before and after the first application of the Kalman correction at 110 days. . . . .                                                             | 37 |
| 3.5  | Water saturation profiles and histogram after transforming back the updated normal scores in Fig. 3.4(b) at 110 days. . . . .                                                      | 38 |
| 3.6  | Water saturation profiles and histogram after the first application of the Kalman correction at 110 days, with saturation front location as a state variable. . . . .              | 40 |
| 3.7  | Saturation profiles after one extra iteration of Kalman correction at 110 days. . . . .                                                                                            | 41 |
| 3.8  | The observed and computed water and oil production rates from the ensemble after Kalman correction, without iteration. . . . .                                                     | 42 |
| 3.9  | The observed and computed water and oil production rates from the ensemble after Kalman correction, with one extra iteration. . . . .                                              | 42 |
| 3.10 | RMSEs and spreads of both the porosity (left) and natural logarithm permeability (right) estimates with and without iteration. . . . .                                             | 43 |
| 3.11 | The locations of the injector and four producers for the two-dimensional waterflood problem. . . . .                                                                               | 44 |
| 3.12 | Oil production rates at the four producers after Kalman correction. .                                                                                                              | 45 |
| 3.13 | Water saturation profiles at the middle column before and after the first application of the Kalman correction at 10 days. . . . .                                                 | 46 |
| 3.14 | The true porosity (left) and permeability (right) fields. . . . .                                                                                                                  | 47 |
| 3.15 | The permeability field for model 1 before (left) and after (right) the first application of Kalman correction at day 10. . . . .                                                   | 47 |
| 3.16 | The final porosity (left) and permeability (right) fields for model 1 after 200 days of Kalman correction. . . . .                                                                 | 48 |

|      |                                                                                                                                                                                                                                                                                                                                                                                            |    |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.17 | The mean of the ensemble permeability estimates at day 10 (left) and day 200 (right) after Kalman correction. . . . .                                                                                                                                                                                                                                                                      | 48 |
| 3.18 | RMSEs and spreads of both the porosity (left) and natural logarithm permeability (right) estimates. . . . .                                                                                                                                                                                                                                                                                | 48 |
| 3.19 | Material balance checking after 200 days of data assimilation. . . . .                                                                                                                                                                                                                                                                                                                     | 50 |
| 3.20 | Comparison of the observed water injection rates (red line) to its computed counterparts (black lines) from the final updated ensemble models at day 200 by rerunning the reservoir simulator from time 0. . . . .                                                                                                                                                                         | 50 |
| 4.1  | Top structural map of the PUNQ-S3 reservoir, from the PUNQ web page: <a href="http://www.nitg.tno.nl/punq/cases/punqs3/index.htm">http://www.nitg.tno.nl/punq/cases/punqs3/index.htm</a> . . . . .                                                                                                                                                                                         | 53 |
| 4.2  | Oil production rate of the true reservoir simulation model versus time for one of the production wells. . . . .                                                                                                                                                                                                                                                                            | 56 |
| 4.3  | Summary plot of how often data are assimilated and what and how many data are available at the well testing, shut-in, and production periods of the first 8 years. . . . .                                                                                                                                                                                                                 | 57 |
| 4.4  | Measurement time intervals of the 20 measurement times. . . . .                                                                                                                                                                                                                                                                                                                            | 59 |
| 4.5  | The production data at PRO-1 from the initial models and during the HM process & prediction. The vertical black line divides the time axis into history matching phase and prediction phase. The red line denotes the simulated data from the true simulation model. The black lines denote the results from different ensemble members. The blue crosses represent the data used. . . . . | 65 |
| 4.6  | The production data at PRO-4 from the initial models and during the the HM process & prediction. The legends used are the same with Fig. 4.5. . . . .                                                                                                                                                                                                                                      | 66 |
| 4.7  | The production data at PRO-5 from the initial models and during the the HM process & prediction. The legends used are the same with Fig. 4.5. . . . .                                                                                                                                                                                                                                      | 67 |
| 4.8  | The production data at PRO-11 from the initial models and during the the HM process & prediction. The legends used are the same with Fig. 4.5. . . . .                                                                                                                                                                                                                                     | 68 |
| 4.9  | The production data at PRO-12 from the initial models and during the the HM process & prediction. The legends used are the same with Fig. 4.5. . . . .                                                                                                                                                                                                                                     | 69 |
| 4.10 | The production data at PRO-15 from the initial models and during the the HM process & prediction. The legends used are the same with Fig. 4.5. . . . .                                                                                                                                                                                                                                     | 70 |

|                                                                                                                                                                                                                                                                                                                                  |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.11 Quantiles of 10, 50, and 90, of the cumulative oil production after 16.5 years from the EnKF, and results summarized by Floris et al. (2001). The horizontal line crossing the entire plot denotes the computed total oil production value from the true reservoir simulation model. . . . .                                | 71  |
| 4.12 Cumulative distribution function (CDF) of the cumulative oil production after 16.5 years from the EnKF. The vertical black line stands for the true total oil production value. The smooth red curve is the referenced Gaussian CDF with mean equal to the true production value and STD of $71,845 \text{ sm}^3$ . . . . . | 71  |
| 4.13 True porosity, evolution of the mean, gridblock RMSE, and STD of porosity estimates in layer 1. . . . .                                                                                                                                                                                                                     | 72  |
| 4.14 True porosity, evolution of the mean, gridblock RMSE, and STD of porosity estimates in layer 2. . . . .                                                                                                                                                                                                                     | 73  |
| 4.15 True porosity, evolution of the mean, gridblock RMSE, and STD of porosity estimates in layer 3. . . . .                                                                                                                                                                                                                     | 74  |
| 4.16 True porosity, evolution of the mean, gridblock RMSE, and STD of porosity estimates in layer 4. . . . .                                                                                                                                                                                                                     | 75  |
| 4.17 True porosity, evolution of the mean, gridblock RMSE, and STD of porosity estimates in layer 5. . . . .                                                                                                                                                                                                                     | 76  |
| 4.18 RMSEs and STDs of for the porosity, $\ln K_h$ , and $\ln K_v$ estimates. . . .                                                                                                                                                                                                                                              | 77  |
| 5.1 Locations where the tracer arrival times are measured for the linear example. . . . .                                                                                                                                                                                                                                        | 90  |
| 5.2 Histogram of the computed tracer travel times with 30 ensemble members at the first measurement times prior to data assimilation. . . . .                                                                                                                                                                                    | 97  |
| 5.3 Histograms of the computed tracer travel times with 30 ensemble members from the Conforming EnKF and EnRMLF both with half step length at the end of their iterations for all measurement times. . . . .                                                                                                                     | 98  |
| 5.4 Ensemble mean porosity with 30 ensemble members from the EnKF, Conforming EnKF and EnRMLF both with half step length at the end of their iterations for all measurement times. . . . .                                                                                                                                       | 99  |
| 5.5 Gridblock STD of the porosity estimates with 30 ensemble members from the EnKF, Conforming EnKF and EnRMLF both with half step length at the end of their iterations for all measurement times. . . . .                                                                                                                      | 100 |
| 5.6 RMSE and STD of the ensemble porosity estimates with 30 ensemble members from the Conforming EnKF and EnRMLF both with half step length during iterations at all measurement times. . . . .                                                                                                                                  | 100 |

|      |                                                                                                                                                                                 |     |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.7  | Sensitivity coefficients computed by singular value decomposition (SVD) with different ensemble size at the third measurement time (location=12) . . . . .                      | 101 |
| 5.8  | Ensemble mean porosity and gridblock STD of the ensemble porosity estimates at the fourth measurement time (location=20) from the EnRMLF with different ensemble size . . . . . | 102 |
| 5.9  | RMSE and STD of the ensemble porosity estimates with different ensemble size from the EnRMLF. . . . .                                                                           | 103 |
| 5.10 | Schematic setup for the non-linear example problem. . . . .                                                                                                                     | 105 |
| 5.11 | Synthetic data before adding measurement noise at different times computed from the selected true reservoir model. . . . .                                                      | 106 |
| 5.12 | Conforming EnKF: $S_w$ profiles at day 140 for Case I. . . . .                                                                                                                  | 117 |
| 5.13 | Conforming EnKF: $S_w$ profiles at day 190 for Case I. . . . .                                                                                                                  | 118 |
| 5.14 | Conforming EnKF: gridblock mean and STD of the ensemble porosity estimates at day 190 from different iterations for Case I. . . . .                                             | 119 |
| 5.15 | Conforming EnKF: RMSE and spread of porosity and $S_w$ estimates at day 190 from different iterations for Case I. . . . .                                                       | 120 |
| 5.16 | Conforming EnKF: the final $S_w$ profiles at days 210, 230, 250, and 270 for Case I. . . . .                                                                                    | 121 |
| 5.17 | RMSE and spread of porosity and $S_w$ estimates from the three iterative methods for Case I. . . . .                                                                            | 122 |
| 5.18 | EnRMLF with iteration at every measurement time: $S_w$ profiles at day 140 before and after G-N iteration. . . . .                                                              | 122 |
| 5.19 | EnRMLF with iteration at every measurement time: $S_w$ profile at day 190. . . . .                                                                                              | 123 |
| 5.20 | EnRMLF with iteration at every measurement time: mean and STD of porosity at day 190 from different iterations for Case I. . . . .                                              | 124 |
| 5.21 | EnRMLF with iteration at every measurement time: RMSE and spread of porosity and $S_w$ estimates at day 190 from different iterations for Case I. . . . .                       | 125 |
| 5.22 | EnRMLF with iteration at every measurement time: the final $S_w$ profiles at days 210, 230, 250, and 270 for Case I. . . . .                                                    | 126 |
| 5.23 | EnRMLF iterate when needed: $S_w$ profiles at day 190 for Case I. . .                                                                                                           | 127 |
| 5.24 | EnRMLF iterate when needed: the final $S_w$ profiles at days 210, 230, 250, and 270 for Case I. . . . .                                                                         | 128 |

|                                                                                                                                                                                                                                                                                                                                        |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.25 Traditional EnKF: $S_w$ profiles at days 190, 210 and 270 for Case I . . . . .                                                                                                                                                                                                                                                    | 128 |
| 5.26 Traditional EnKF with truncation of $S_w$ after Kalman correction: $S_w$ profiles at day 190, 210 and 270 for Case I . . . . .                                                                                                                                                                                                    | 129 |
| 5.27 RMSE and spread for estimates of porosity and $S_w$ for the traditional EnKF with and without truncation of $S_w$ after Kalman corrections for Case I . . . . .                                                                                                                                                                   | 129 |
| 5.28 Data match at the 8 different measurement times from the three iterative methods for Case I. The solid red dots on each of the plot are the synthetic data computed directly from the selected true reservoir model. The multiple hollow black dots at each measurement time are the final data from the ensemble models. . . . . | 130 |
| 5.29 RMSE and spread of both estimates of porosity and $S_w$ from various methods for Case I . . . . .                                                                                                                                                                                                                                 | 131 |
| 5.30 Conforming EnKF: $S_w$ profiles at days 190 for Case II. . . . .                                                                                                                                                                                                                                                                  | 132 |
| 5.31 Conforming EnKF: RMSE and spread of porosity and $S_w$ estimates at day 190 from different iterations for Case II. . . . .                                                                                                                                                                                                        | 133 |
| 5.32 The comparison of the porosity estimates at day 190 of Case I and II. .                                                                                                                                                                                                                                                           | 133 |
| 5.33 Conforming EnKF: the final $S_w$ profiles at days 210, 230, 250, and 270 for Case II. . . . .                                                                                                                                                                                                                                     | 134 |
| 5.34 Conforming EnKF: data match at different measurement times for Case II. The solid red dots on each of the plot are the synthetic data computed directly from the selected true reservoir model. The multiple hollow black dots at each measurement time are the final data from the ensemble models. . . . .                      | 135 |
| 5.35 RMSE and spread of porosity and $S_w$ estimates from various methods for Case II. . . . .                                                                                                                                                                                                                                         | 136 |
| 6.1 Porosity distribution for the BBCK sector model. . . . .                                                                                                                                                                                                                                                                           | 137 |
| 6.2 Initial water saturation at the bottom layer (layer 39) of the BBCK sector model (the blue-colored areas represent the aquifer). . . . .                                                                                                                                                                                           | 138 |
| 6.3 Water cut for well 7D_27 computed with different data processing schemes. . . . .                                                                                                                                                                                                                                                  | 140 |
| 6.4 perturbed measurements and computed oil production rates, from 100 models, for well 6D_89 at Year 1965.5. . . . .                                                                                                                                                                                                                  | 142 |
| 6.5 Wells achieving different history matching results projected on the bottom layer with initial water saturation shown. The blue color represents the aquifer. . . . .                                                                                                                                                               | 149 |

|                                                                                                                                                                                                                                                                       |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 6.6 Water cut at well 7D_27. Similar behavior wells include: 6D_89, 7C_58, 7E_41, 8D_24, and 8D_33. They are all located close to reservoir boundaries and directly above the bottom aquifer which provides resource for water production (see Fig. 6.5(a)) . . . . . | 150 |
| 6.7 Water cut at well 8D_21. Similar behavior wells include: 7D_47, 7D_55, 7D_58, 7D_83, 7D_88, 8D_12, and 8D_15. They are all located directly above the bottom aquifer and water resource is guaranteed (see Fig. 6.5(a)). . . . .                                  | 151 |
| 6.8 Water cut at well 7D_11. Similar behavior well include: 7D_12. They are all located close to one of the reservoir boundaries and not much direct aquifer support for this fault block at early times (see Fig. 6.5(b)).                                           | 151 |
| 6.9 Water cut at well 7D_44. Similar behavior well include: 7D_24. They are all located close to right side the longer fault and not much direct aquifer support for this fault block at early times (see Fig. 6.5(b)). . .                                           | 152 |
| 6.10 Water cut at well 7D_74. Similar behavior wells include: 7D_71, 7D_77, and 8C_18 (see Fig. 6.5(c)). . . . .                                                                                                                                                      | 152 |
| 6.11 Initial and predicted water cut at well 7E_41. The blue-colored line on the second plot is the history-matching result obtained by generalized travel-time. . . . .                                                                                              | 152 |
| 6.12 Initial and predicted water cut at well 7D_44. Same legends are used as Fig. 6.11. . . . .                                                                                                                                                                       | 153 |
| 6.13 Initial and predicted water cut at well 7D_27. Same legends are used as Fig. 6.11. . . . .                                                                                                                                                                       | 153 |
| 6.14 Initial and predicted water cut at well 7D_55. Same legends are used as Fig. 6.11. . . . .                                                                                                                                                                       | 153 |
| 6.15 Initial and predicted field wide water cut. . . . .                                                                                                                                                                                                              | 154 |
| 6.16 Evolution of the mean $\log k_x$ estimate at layer 34 based on the 6-month interval data. . . . .                                                                                                                                                                | 155 |
| 6.17 Evolution of the mean $\log k_x$ estimate at layer 8 based on the 6-month interval data. . . . .                                                                                                                                                                 | 156 |
| 6.18 Evolution of the mean $\log k_x$ estimate at layer 3 based on the 6-month interval data. . . . .                                                                                                                                                                 | 157 |
| 6.19 Gridlock STD for the $\log k_x$ estimates at layer 34. . . . .                                                                                                                                                                                                   | 158 |

# ABSTRACT

In this dissertation, the ensemble Kalman filter (EnKF) is refined for applications to the problem of history matching. Like the Kalman filter and its variants, this method assimilates data sequentially according to the chronological order at which the data are acquired. This conceptual framework is suitable for real-time reservoir monitoring with data from permanent down-hole gauges. The EnKF is a Monte Carlo approach, in which an ensemble of models is used. The correlations between model variables and theoretical data are computed directly from the ensemble. By so doing, this method avoids the derivation and implementation of the complex adjoint equations required by efficient gradient-based history matching approaches. Thus it is a self-contained method and highly versatile and can be built upon any reservoir simulator. Multiple history-matched models will be obtained after the initial ensemble is conditioned to all production data. The final models can be used to assess the uncertainty in future reservoir performance. The computational cost for generating the multiple history-matched models is approximately the simulation runs for all the ensemble models plus the overhead time spent at the update steps.

The plausibility of the EnKF as an alternative method for history matching to reservoir applications is shown by two waterflood problems. We found that when the adjustments to reservoir variables are large at a measurement time, the EnKF may produce non-physical values for the state variables, e.g.  $S_w > 1.0$  or  $S_w < 0.0$ ; but when the adjustments made at each measurement time are small, the EnKF is able to correct both the model parameters and state variables with acceptable reliability.

The effectiveness of the EnKF is more thoroughly demonstrated with the PUNQ-S3 reservoir model. The PUNQ-S3 model has been used in several comprehensive studies to benchmark uncertainty quantification methods. With the EnKF, the production history is honored well and the history-matched models are able to provide good prediction in total oil production comparable to other methods.

The EnKF takes both model parameters and state variables as well as calculated data into its state vectors. When new observations are assimilated, all variables in the state vectors are adjusted simultaneously. For linear dynamic systems, the system governing equations are honored by the updated model parameters and state variables. For non-linear dynamic systems, however, when both variables are updated, in essence a linearized approximation to the non-linear relationship is used to make a prediction of the changes in state variables that would result from the changes in the model parameters. When the non-linearity is severe, e.g. dramatic changes of pressure and saturation in the reservoir simulation models, it may be impossible to update the state variables to be consistent with the updated model parameters without re-solving the non-linear forward problem.

Wen and Chen suggested adding a “conforming step” at each measurement time to compute the state variables with the updated model parameters by re-initializing the dynamic equations at the previous measurement time. Though applications have shown that their method achieves plausible results, we demonstrate in this dissertation through both linear and non-linear examples that the results of the conforming method are incorrect. We propose the ensemble Randomized Maximum Likelihood filter (EnRMLF) as a modification to the traditional EnKF to handle the non-linearity. For linear dynamic systems, the EnRMLF produces equivalent solutions as the EnKF where the validity of the solution is guaranteed when the ensemble is large enough. When dealing with strong non-linear systems, the EnRMLF out-performs the EnKF.

As a conclusion of this study, we were able to apply the EnKF and the Conforming EnKF of Wen and Chen to a real case study with availability of a courtesy data set from Chevron. It is a mature oil field in east Asia with long production history. The EnKF provided reasonable data match with quality comparable to previous studies from other methods. In this case, the Conforming EnKF did not provide better results. We identified that the uncertainties in the initial conditions have strong impacts on the history-matching outputs.

# CHAPTER I

## INTRODUCTION

History matching is the process of adjusting the variables in a reservoir simulation model so that the computed values of observables such as rates, or bottom-hole pressures, or gas-oil ratios, or water cuts, at individual wells are in reasonable agreement with actual measurements of those quantities. Due to the large scale of real petroleum reservoirs and the complexity of fluid flow and transport in the reservoirs, it is never an easy task to find a set of reservoir parameters to achieve the match. Fortunately, great progress has been made towards the automation of the adjustment procedure, and it is now possible to perform a history match of multi-phase flow data at the cost of approximately 100 reservoir simulation runs.

Within the context of Bayesian statistics, automatic history matching can be reduced to a non-linear minimization problem. The objective function contains the square of the mismatch of all measurements and the computed values, and the square of the mismatch of the current model parameters and the prior model parameters. There are two categories of optimization algorithms used in reservoir applications. One category is gradient based algorithms, e.g. steepest descent, Gauss-Newton, Levenberg-Marquardt, conjugate gradient, and etc. The other one is non-gradient based algorithms, such as simulated annealing, genetic algorithm, and MCMC. Large computational effort is required for the algorithms in both categories, either in objective function evaluation (non-gradient based minimization methods, thousands of simulation runs are needed), or in gradient computation (gradient based minimization methods). For gradient based algorithms, the adjoint method has been shown as the most efficient method to compute the gradient of a defined objective function (Zhang

and Reynolds, 2002). However, there are some drawbacks about the adjoint system hindering it from wide engineering practices. First, the adjoint equations are very time-consuming and complicated for deployment. Second, they are highly dependent on the reservoir simulator, and therefore it is not flexible and hard to transfer from one simulator to another.

In the traditional history matching, the data throughout the history are collected and used once. However, the increase in deployment of permanent sensors for monitoring pressure, temperature, resistivity, or flow rate has added impetus to the related problem of continuous model updating. Because the data output frequency in this case can be very high, to simultaneously use all recorded data to generate a reservoir flow model is impractical. Instead, it has become important to incorporate the data as soon as they are obtained so that the reservoir model is always up to date. Both the heavy computational burden and the high data sampling frequency require a new kind of history-matching method.

The Kalman filter has historically been the most widely applied method for assimilating new measurements to continuously update the estimate of state variables. Kalman filters have occasionally been applied to the problem of estimating values of petroleum model variables (Eisenmann et al., 1994; Corser et al., 2000), but they are most appropriate when the problems are characterized by relatively small numbers of variables and when the variables to be estimated are linearly related to the observations. Most data assimilation problems in petroleum reservoir engineering are highly non-linear and are characterized by many variables, often two or more variables per simulator gridblock.

Application to non-linear problems was at least partially solved by the development of the extended Kalman filter. However, it did not solve the critical problem with non-linear unstable dynamics, where it leads to a linear instability in the error covariance evolution (Evensen, 1994). The problem of weather forecasting is in many

respects similar to the problem of predicting future petroleum reservoir performance. The economic impact of inaccurate predictions is substantial in both cases, as is the difficulty of assimilating very large data sets and updating very large numerical models. One method that has been recently developed for assimilating data in weather forecasting is ensemble Kalman filtering (Evensen, 1994; Evensen and van Leeuwen, 1996; Evensen, 1997, 2003; Houtekamer and Mitchell, 1998, 2001; Anderson and Anderson, 1999; Hamill et al., 2000). The method is now beginning to be applied for data assimilation in groundwater hydrology (Reichle et al., 2002; Chen and Zhang, 2006) and in petroleum engineering (Nævdal et al., 2002, 2005; Gu and Oliver, 2006, 2005; Wen and Chen, 2005a,b; Liu and Oliver, 2005; Zafari and Reynolds, 2005; Gao et al., 2005; Lorentzen et al., 2005; Skjervheim et al., 2005; Dong et al., 2006).

The ensemble Kalman filter (hereafter, EnKF) is a Monte Carlo approach, in which an ensemble of models, instead of only one model as in traditional history matching methods and other Kalman filter related methods, is used. The correlation between reservoir response (e.g. rates, bottom-hole pressures, gas-oil ratios, and water cuts) and reservoir variables (e.g. porosities and permeabilities) is estimated directly from the ensemble, which is different from the explicit evolution of the covariance matrix in the Kalman filter and its non-linear extensions. Researchers have showed that the EnKF outperforms the extended Kalman filter for severe non-linear problems (Zang and Malanotte-Rizzoli, 2003; Bertino et al., 2003) because the latter ignores the higher order statistics, while the information about them are kept in the EnKF when the system governing equations propagate forward in time for the multiple reservoir models. By estimating the state error covariance function directly from the ensemble, the EnKF avoids computing the adjoint equations and thus the implementation becomes significantly simplified and transferable since it does not depend on the system dynamics used.

Like other Kalman filter related sequential estimation algorithms, the EnKF divides the history matching process into two major steps at each measurement time: (1) a forecast step which propagates the state of the system/model, usually, from previous measurement time to current measurement time, and (2) an update step which corrects the system/model variables by taking the current measurements into account. Generally, for petroleum reservoir applications, the system/model evolution is dictated by equations for multi-phase fluid flow and transport in a reservoir; the state variables are phase pressures and saturations at each simulation gridblock; the model parameters are properties intrinsic to the reservoir rock and not time varying themselves, such as porosity and absolute permeability at each simulation grid, although the estimates of them change with time when new measurements arrives; the model variables include both the model parameters and state variables; the forecast step is done by running a reservoir simulator, in this dissertation, Schlumberger's commercial simulator, **ECLIPSE** 100, and Chevron's in-house simulator, **CHEARS** (Chevron Extended Application Reservoir Simulator), are both used; the measurements could be production rates of fluid phases, bottom-hole pressures, gas-oil ratios, or water cuts for individual wells. The number of measured production data at one measurement time is usually moderate, of course, it can be large for other kinds of data, e.g. time-lapse seismic data. The interval between two consecutive measurement times might be as short as a few seconds, such as data from permanent gauges, or as long as years, such as time-lapse seismic data from two surveys. Cares need to be taken when the interval is extremely short or long.

The two-step procedure is repeated at each measurement time till the last measurements are assimilated. When the number of measurements available at one measurement time is moderate, as we considered throughout the dissertation, the reservoir simulation time dominates the computational cost of the EnKF. It is obvious that the

computing time for generating multiple history-matched reservoir models is approximately the same as the time to make simulation runs for the same number of models plus some overhead time spent at the updated steps. The multiple history-matched reservoir models can be used to assess the uncertainty in future reservoir performance.

When new measurements become available, all the variables, including both the model parameters, such as porosities and permeabilities, and state variables, such as pressures and saturations, are updated simultaneously. If the relationship between the state variables and model parameters is linear, the model parameters and the state variables can be adjusted simultaneously with consistency, i.e. the system equations are honored by the updated model parameters and state variables. The result after data assimilation contains an improved estimate of the (non-varying) model parameters and also an improved estimate of the current value of the state variables. However, for a non-linear problem, when both kinds of variables are updated, in essence a linearized approximation to the simulator is used to make a prediction of the pressure changes and saturation changes that would result from the porosity and permeability changes. When the changes to the state variables are small, the linearized approximation to the reservoir simulator is acceptable. However, when the changes are big, it may be impossible to update the state variables to be consistent with the updated model parameters without re-solving the non-linear forward equations. This is one potential problem with updating both the model parameters and state variables simultaneously in the EnKF. The update equation in the EnKF is based on Gaussian error statistics. The update of state variables whose density functions are bi-modal with the EnKF has been shown as problematic (Gu and Oliver, 2006). In reservoir applications, such state variables are water saturations at some circumstances. For a simple water flooding problem, the water saturations take large values behind the water flood front, and small values ahead of the front. Its distribution is bi-modal in this case and is not well modeled by the mean and variance. This

is the second potential problem with the EnKF.

The two problems of the EnKF are interwoven with each other. Wen and Chen (2005a,b) proposed an intuitive remedy to the implications. They suggested to add a so-called “conforming step” at each measurement time. So the process at one measurement time of their proposal is a three-step procedure: (1) a forecast step which propagates the state of the system from previous measurement time to current measurement time, which is the same with that in the EnKF (2) an update step which corrects only the model parameters with the EnKF update equation, and (3) a conforming step which re-initiates the system governing equations at the previous measurement time with the newly updated model parameters, conditional to data up to the current measurement time, and the state variables at the previous measurement time, then propagates the re-initialized equations to the current measurement time for the computation of the state variables. It can be easily seen that their scheme doubles the computing time comparing to that of the EnKF since there are two simulation runs for each simulation model in the forecast and conforming steps. They also suggested to iterate **Steps 2** and **3** when non-linearity of problems is strong.

In this dissertation, we showed that their scheme may not be correct even though it seems to provide plausible results. We initiated a new scheme, called **Ensemble Randomized Maximum Likelihood Filter** (EnRMLF), with solid theoretical background. It was proven to be robust at least with the illustration of linear and non-linear examples demonstrated in this work. Similarly as Wen and Chen’s remedy, only model parameters are corrected at the update step and one extra step is added after the correction of the model parameters to compute the state variables at the current measurement time being. Although the procedure of the newly proposed method has similarity to that of Wen and Chen’s, there are two major differences between the two. One is that we adopt the iterative Gauss-Newton formula to update the model parameters. Secondly, after the new model parameters are obtained, the

system governing equations are re-initialized at time 0 to compute the state variables for consistency. The disadvantage of this method is the intensive computational cost. However, it is not always necessary to apply the extra step. When the changes made to the variables in the state vectors are small at a measurement time, the general EnKF can be applied. By carefully setting up criteria for choosing whether to add the extra step, in the examples showed in the dissertation, we found that it is only applied when data carrying significant information arrive.

There are 7 chapters in this dissertation. Chapter 2 introduces the basics about the EnKF methodology in the background of reservoir applications. Chapter 3 addresses two primary concerns about using the EnKF to reservoir applications. Both one-dimensional and two-dimensional waterflood problems are selected to investigate the two issues. Chapter 4 applies the EnKF to a more realistic reservoir model, PUNQ-S3. Through the studies in Chapters 3 and 4, we found that the EnKF is most appropriate when the changes to the state variables are small. But when the changes made at some measurement times are big, possible problems arise with the EnKF. Aiming to eliminate the possible problems, we purposed a iterative scheme, called EnRMLF, in Chapter 5. Both linear and non-linear examples are used to demonstrate the performance of the newly purposed iterative scheme in comparison to that from the EnKF and Conforming EnKF (Wen and Chen, 2005a,b). Chapter 6 shows the application of the EnKF and Conforming EnKF to a data set from a real field. Finally, Chapter 7 concludes the study.

# CHAPTER II

## THE ENSEMBLE KALMAN FILTER

The chapter explains the basics about the EnKF methodology in the background of reservoir applications. In Section 2.1, the fluid flow and transport governing equations in porous media are reviewed and terminology commonly used in history matching is listed. Section 2.2 provides an outline of the Kalman filter algorithm. Sections 2.3 – 2.5 introduce the methodology of the EnKF and implementation procedures of the method on both PC and multiple computing clusters. The last Section gives some heuristics measures used throughout this work for evaluating the EnKF performance.

### 2.1 Background and Terminology

#### 2.1.1 Reservoir flow and transport equations

In reservoir applications, the dynamic model evolution equations governing the multi-phase flow and transport in porous media are derived from the basic laws of mass balance and momentum balance. Suppose that the reservoir under study can be modeled by the flow of three fluid phases (oil, water, and gas) as described by the following three partial differential equations (PDEs) for black oil models. The black oil model implies that (1) gas component can be in both gas phase and oil phase, but oil component can only be in oil phase, i.e., no oil vaporization; (2) water component appears only in water phase; (3) neither oil nor gas component is in water phase, i.e., water is immiscible with both gas and oil. As reservoir temperature is assumed to be constant in this application and we are not considering thermal or chemical process, energy balance is automatically honored and thus not explicitly shown in the

following equations.

$$C_1 \nabla \cdot \left[ \frac{k_{ro}}{\mu_o B_o} \vec{k} (\nabla p_o(x, y, z, t) - \gamma_o \nabla D(x, y, z)) \right] = \frac{\phi}{C_2} \frac{\partial}{\partial t} \left( \frac{S_o}{B_o} \right) + q_o(x, y, z, t), \quad (2.1)$$

$$C_1 \nabla \cdot \left[ \frac{k_{rw}}{\mu_w B_w} \vec{k} (\nabla p_w(x, y, z, t) - \gamma_w \nabla D(x, y, z)) \right] = \frac{\phi}{C_2} \frac{\partial}{\partial t} \left( \frac{S_w}{B_w} \right) + q_w(x, y, z, t), \quad (2.2)$$

$$\begin{aligned} C_1 \nabla \cdot & \left[ \frac{k_{rg}}{\mu_g B_g} \vec{k} (\nabla p_g(x, y, z, t) - \gamma_g \nabla D(x, y, z)) + \right. \\ & \left. \frac{R_{so} k_{ro}}{\mu_o B_o} \vec{k} (\nabla p_o(x, y, z, t) - \gamma_o \nabla D(x, y, z)) \right] \\ & = \frac{\phi}{C_2} \frac{\partial}{\partial t} \left( \frac{S_g}{B_g} + \frac{R_{so} S_o}{B_o} \right) + q_g(x, y, z, t). \end{aligned} \quad (2.3)$$

With an auxiliary equation

$$S_o + S_w + S_g = 1. \quad (2.4)$$

Throughout, the subscripts  $o$ ,  $w$ , and  $g$  stand for oil phase, water phase, and gas phase, respectively;  $C_1 = 1.127 \times 10^{-3}$  and  $C_2 = 5.615$  when oil field units are used; the oil and water formation volume factor ( $B_o$  and  $B_w$ ) are in units  $RB/STB$ , the gas formation volume factor is in  $RB/scf$ ;  $R_{so}$  is the dissolved gas-oil ratio in units of  $scf/STB$ ; the viscosity ( $\mu_m$ ) is in units of  $cp$ ;  $\gamma_m = \frac{\rho_m g}{144 g_c}$  is the specific density;  $D$  is the vertical distance from a datum level and in units of  $ft$ ;  $q_m$  is the source/sink term,  $q_o$  and  $q_w$  are in units of  $STB/(ft^3 \cdot Day)$ ,  $q_g$  is in units of  $scf/(ft^3 \cdot Day)$ , for production,  $q_m > 0$ , for injection  $q_m < 0$ ; the pressure ( $p_m$ ) is in units of  $psia$ , the pressure of phases are related to each other through capillary pressures, i.e.  $p_{cow} = p_o - p_w$  and  $p_{cgo} = p_g - p_o$  assuming water being the wetting phase in the presence of oil and water, and oil being the wetting phase in the presence of oil and gas; in the dissertation, the capillary pressure is assumed to be negligible so that  $p_o = p_w = p_g = p$ ; the saturation ( $S_m$ ) is dimensionless and varies between 0 and 1; porosity  $\phi \in [0, 1]$  is the fraction of pore space in the reservoir rock and dimensionless;  $\vec{k}$  is the diagonal permeability

tensor  $\vec{k} = \begin{bmatrix} k_x & 0 & 0 \\ 0 & k_y & 0 \\ 0 & 0 & k_z \end{bmatrix}$ , the non-zero entries are the absolute permeability of the reservoir rock along  $x$ ,  $y$ , and  $z$  directions; the permeabilities has the dimension of  $[L^2]$  and for oil field units it is in  $md$ ;  $k_{rm}$  is the relative permeability and a function

of saturation;  $\nabla$  is divergence operator, for Cartesian coordinate system,  $\nabla = \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix}$ .

With well determined initial and boundary conditions, the PDEs become into a well-defined Initial Boundary Value Problem (IBVP). Although solving the equations is theoretically appealing, it is formidably difficult even if a mild heterogeneous reservoir is considered. Therefore, leaning toward numerical solution is a natural choice. Usually, the reservoir is discretized into millions of gridblocks. In each gridblock, the three governing equations are still valid, but much simplified because the porosity, permeability, viscosity, and  $B_m$  are assumed to be constant within the grid at a single iterative step. Thus, the flow equations are linearized and equations from every gridblock are solved together to obtain pressure and phase saturations, which provide an insight of porous fluid distribution in subsurface. The program conducting this computation process is called reservoir simulator.

There are several different discretion methods to split the reservoir. In this dissertation, what we used is based on finite difference methods. As this study does not focus on reservoir simulation, more detailed discussion of the simulation equations are not involved. Some classical references can be found from extensive literature on this topic, for example, Aziz and Settari (1979).

### 2.1.2 History matching terminology

Before we introduce the methodology of the EnKF, let's first clarify some terminology that is commonly used in history matching for reservoir applications.

**Model parameters:** these are variables that are uncertain but not time varying. These variables are also called **static model variables**. They include rock properties such as porosity and absolute permeability. For sequential data assimilation methods, the estimate of these properties changes as data are incorporated, but the parameters themselves should not be interpreted to be changing with time.

**State variables:** these are uncertain, time-dependent variables that define the state of the system. Because of the time-dependence, they are also called **dynamic model variables**. The uncertainty in these variables is resulted from the uncertainty in the model parameters and some other uncertain factors, e.g. initial conditions. For petroleum reservoirs, state variables could include phase pressures, saturations of each fluid phase, or mass fraction. These variables are frequently solutions of systems of differential (or difference) equations. If the physical model is valid, and the model parameters are known, then it is possible to compute the state variables with given initial conditions.

**Data:** these are observable quantities related directly to the state variables and indirectly to the model parameters. For petroleum reservoirs, data might include surface flow rates, bottom-hole pressure (possibly at several locations in the well-bore), amplitude of seismic reflection, and etc. In reality, the observed data always have some unknown level of error or noise associated with them.

In reservoir history matching applications,  $m$  is used to denote the model parameters,  $f(m)$  is used to denote the state variables, and  $g(m)$  is used to denote the predictions of observations.  $Fm$  and  $Gm$  are used instead when the state variables and data are related to the model parameters through linear relationships.

We frequently write the relationship between the true model parameters and observed data as

$$d_{\text{obs}} = g(m^{\text{true}}) + \epsilon, \quad (2.5)$$

where  $\epsilon$  is the unknown measurement noise. It is assumed to be unbiased and Gaussian,  $\epsilon \sim N(0, C_D)$ , i.e.  $E[\epsilon] = 0$  and  $E[\epsilon\epsilon^T] = C_D$ .  $C_D$  is measurement error covariance matrix. It is a diagonal matrix if the measurement errors are uncorrelated, which is what we used throughout the applications in this work.

In the Kalman filter related literature, *the state vector for a dynamic system is composed of any set of quantities sufficient to completely describe the unforced motion of that system. Given the state vector at a particular point in time and a description of the system forcing and control functions from that points in time forward, the state at any other time can be computed* (Gelb, 1979, page 52).

We might define a state vector of the form

$$y = \begin{bmatrix} m \\ f(m) \\ g(m) \end{bmatrix} \text{ or } \begin{bmatrix} m \\ Fm \\ Gm \end{bmatrix}, \quad (2.6)$$

where  $y$  denotes the state vector. It consists of model parameters, state variables, and data. Using the state vector, the relationship between the observed data and the true state vector can be written as

$$d_{\text{obs}} = Hy^{\text{true}} + \epsilon, \quad (2.7)$$

where  $H$  is an operator matrix or row vector, depending on the number of observations, which relates the state vector to theoretical data. Because the data are part of the state vector as shown in Eq. 2.6,  $H$  is a trivial matrix whose elements are only ones or zeroes. We can always arrange it as

$$H = [ \mathbf{0} \mid \mathbf{I} ], \quad (2.8)$$

In practice, the construction of  $H$  is not essential. Pre-multiplied by  $H$  merely selects corresponding rows of a matrix. Similarly, post-multiplied by  $H^T$  selects the corresponding columns.

Note all of the above equations (Eq. 2.5 to Eq. 2.8) are written in a general form and the time-dependence of the variables is not shown. Later in the following sections, time indices will be added to the variables for dynamic systems.

## 2.2 Outline of the Kalman Filter Algorithm

The Kalman filter was first developed for problems of estimation for linear dynamic systems in the 60's by Kalman and Bucy (Kalman, 1960; Kalman and Bucy, 1961). Historically, it has been the most widely applied method for assimilating new measurements to continuously update the estimates for both the model parameters and state variables. Extensive applications can be found in the fields of submarine and aircraft navigation, radar tracking, and satellite orbit determination, to name a few.

When *filter* is mentioned, one may think of electrical circutes or network. However, in this dissertation, we use *filter* to simply mean a data processing algorithm. Because of the wide availability of the Kalman filter derivations from many resources (see, for example, Gelb (1979), Maybeck (1979), and Anderson (1979)), only an outline of the algorithm is provided here.

The Kalman filter is an optimal recursive data processing algorithm. At each of the recursive process, i.e. at one measurement time, there are two steps: a **forecast step** and an **update step**. We will elaborate on both steps at a general measurement time  $t_k$  as follows.

**Forecast step** is to evolve the state vector forward in time between two consecutive measurement times. Writing the dynamic system model in a mathematical form

$$y_k = \Psi_{k-1} y_{k-1} + \omega_{k-1}, \quad (2.9)$$

the evolution of the state vector is

$$y_k^p = \Psi_{k-1} y_{k-1}^u, \quad (2.10)$$

where the subscripts  $k$  and  $k - 1$  are integer time indices for measurement time  $t_k$  and  $t_{k-1}$ , respectively, where measured data are available; the superscript  $p$  represents *prior*, meaning that the values are direct output of the dynamic system before updating (Eq. 2.10);  $u$  represents *updated*, meaning that the values are after data assimilation;  $\Psi_{k-1}$  is a transition matrix that transits the state vector from time  $t_{k-1}$  to time  $t_k$ ;  $\omega_{k-1}$  is unbiased Gaussian model error with covariance  $Q_{k-1}$ , i.e.  $E[\omega_{k-1}] = 0$  and  $E[\omega_{k-1}\omega_{k-1}^T] = Q_{k-1}$ .

$y_{k-1}^u$  is the estimate of the state vector conditioned to measurements up to time  $t_{k-1}$ . Thus  $y_k^p$  is also regarded as conditional to observed data up to time  $t_{k-1}$ . Denote the collection of measurements up to time  $t_{k-1}$  by  $D_{\text{obs},k-1}$

$$D_{\text{obs},k-1} = \{d_{\text{obs},i} | 1 \leq i \leq k-1\}. \quad (2.11)$$

Assume that the probability density function (PDF) of  $y_k^p$ , called *prior* PDF, is Gaussian

$$p(y_k^p | D_{\text{obs},k-1}) \sim N(y_k^p, C_{Y,k}^p), \quad (2.12)$$

where  $y_k^p$  is calculated from Eq. 2.10;  $C_{Y,k}^p$  is the *prior* covariance matrix associated with the prior estimate. It is explicitly computed by propagating an assumed initial covariance matrix of the state vector at time 0,  $C_{Y,0}^u$ , through time

$$C_{Y,k}^p = \Psi_{k-1} C_{Y,k-1}^u \Psi_{k-1}^T + Q_{k-1}, \quad (2.13)$$

where  $C_{Y,k-1}^u$  is the *posterior* covariance matrix after data are assimilation at time  $t_{k-1}$ .

At the **update step**, with the arrival of the new data,  $d_{\text{obs},k}$ , the state vector is adjusted with

$$y_k^u = y_k^p + K_k(d_{\text{obs},k} - H_k y_k^p), \quad (2.14)$$

where  $K_k$  is the Kalman gain matrix

$$K_k = C_{Y,k}^p H_k^T (H_k C_{Y,k}^p H_k^T + C_{D,k})^{-1}, \quad (2.15)$$

where  $C_{D,k}$  is the data error covariance matrix at time  $t_k$ ,  $C_{D,k} = E[\epsilon_k \epsilon_k^T]$  and  $\epsilon_k$  is the noise contained in the measurement  $d_{\text{obs},k}$ . After the data assimilation, the PDF of  $y_k^u$ , called *posterior* PDF, now becomes

$$p(y_k^u | D_{\text{obs},k}) \sim N(y_k^u, C_{Y,k}^u), \quad (2.16)$$

where  $C_{Y,k}^u$  is the *posterior* covariance matrix associated with the updated state vector

$$C_{Y,k}^u = (I - K_k H_k) C_{Y,k}^p. \quad (2.17)$$

The above is an illustration of the two-step procedure at one measurement time. For convenience, the procedure is summarized in Table 2.1. Some of the equations are repeated in the table and are referred by their numbers appeared in the text. With the generation of the state vector at the initial time 0 (the initial state vector is generated by sampling from a multi-variate Gaussian distribution with the prior information about its mean and covariance matrix,  $C_{Y,0}^u$ ), the recursive process of the Kalman filter can be initiated. The two-step procedure is repeated until the last data are assimilated.

---

|                                                                                |                                                                                                                   |
|--------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|
| <b>1.</b> Propagate the state vector and its covariance matrix forward in time | Eq. 2.10: $y_k^p = \Psi_{k-1} y_{k-1}^u$<br>Eq. 2.13: $C_{Y,k}^p = \Psi_{k-1} C_{Y,k-1}^u \Psi_{k-1}^T + Q_{k-1}$ |
| <b>2.</b> Update the state vector and its covariance matrix using the new data | Eq. 2.14: $y_k^u = y_k^p + K_k (d_{\text{obs},k} - H_k y_k^p)$<br>Eq. 2.17: $C_{Y,k}^u = (I - K_k H_k) C_{Y,k}^p$ |

---

**Table 2.1:** Summary of the two-step procedure of the Kalman filter at measurement time  $t_k$ .

The solution that the Kalman filter offers is based on maximizing the *posterior* PDF of the state vector within the context of Bayesian statistics. It is equivalent

to minimizing the variances of the *posterior* covariance matrix in Eq. 2.17 with the assumption that the following variables are Gaussian (see Maybeck (1979, page 4), Anderson (1979, page 21)).

- model errors,  $\omega_{k-1}$ ,
- measurement errors,  $\epsilon_k$ , and
- the estimate of state vector at the initial time 0.

In addition to the Gaussian assumptions, the model and measurement errors are also assumed to be unbiased and *white*. Whiteness implies that the noises are not correlated in time. Note that if the dynamic model (Eq. 2.9) is linear and the estimate of the initial state vector is Gaussian, both the *prior* and *posterior* PDFs (Eqs. 2.12 and 2.16) would be Gaussian because the linear combination of Gaussian variables is still Gaussian. Both the non-linear dynamics and violation of the Gaussianity of the state variables themselves can cause the *prior* and *posterior* PDFs to lose the Gaussian statistics.

What if the Gaussian assumptions are dropped? We know that for Gaussian variables, mean and covariance are sufficient to describe a PDF. But for distributions that are not Gaussian, the mean and covariance provide incomplete knowledge to deduce a PDF. In such cases, although all the calculations of the Kalman filter can still be carried through, the resulted estimate would be no more optimal.

## 2.3 The Ensemble Kalman Filter

The Kalman filter has occasionally been applied to the problem of estimating values of petroleum model variables (Eisenmann et al., 1994; Corser et al., 2000), but they are most appropriate when the problems are characterized by relatively small numbers of variables (because the explicit computation of time evolution of the covariance matrix and the storage of the matrix) and when the variables to be estimated are linearly

related to the observations. However, most data assimilation problems in petroleum reservoir engineering are highly non-linear and are characterized by many variables, often two or more variables per simulator gridblock.

The Kalman filter algorithm has been extended for applications to non-linear dynamic models. The typical engineering approach is to linearize the dynamic model at some nominal point or trajectory, achieving a perturbation model or error model (Maybeck, 1979, page 7). Extended Kalman filter (ExKF) is one of these examples. Evensen (1994) showed that the ExKF failed in an application of a severe non-linear system and proposed the ensemble Kalman filter as a solution to resolve some of the problems that the ExKF has. There are at least two points about the ExKF that make it not preferable and inferior to the EnKF: (1) linearization of the dynamic model requires gradient of the non-linear dynamic model to the state vector. For efficient computation, adjoint equations may be needed. Derivation of the adjoint equations are complex and the deployment is time-consuming. The adjoint equations are also dependent on the dynamic model used thus not easily transferable from one dynamic model to another, and (2) the time evolution of covariance matrix neglects the higher order statistical moments, which may lead to unbounded error growth. Researchers have showed that the EnKF performs better than the ExKF for strong non-linear dynamic problems (Zang and Malanotte-Rizzoli, 2003; Bertino et al., 2003). Since its introduction in 1994 (Evensen, 1994), the EnKF has gained tremendous popularity with intensive applications to meteorology and oceanography (Evensen and van Leeuwen, 1996; Evensen, 1997, 2003; Houtekamer and Mitchell, 1998, 2001; Anderson and Anderson, 1999; Hamill et al., 2000). It is now beginning to be applied for data assimilation in groundwater hydrology (Reichle et al., 2002; Chen and Zhang, 2006) and in petroleum engineering (Nævdal et al., 2002, 2005; Gu and Oliver, 2006, 2005; Wen and Chen, 2005a,b; Liu and Oliver, 2005; Zafari and Reynolds, 2005; Gao et al., 2005; Lorentzen et al., 2005; Skjervheim et al., 2005; Dong et al., 2006).

The EnKF is a Monte Carlo approach, in which an ensemble of models is used. Different from the explicit evolution of the covariance matrix in the Kalman filter and ExKF, the covariance matrix of the state vector in the EnKF is estimated directly from the ensemble. By so doing, the higher order statistical moments are kept when the non-linear dynamics is propagated forward in time for the multiple models.

The framework of the EnKF is the same as that of the Kalman filter. It is also a recursive process and each of the recursive process contains two steps: a forecast step and an update step. The followings introduce the building blocks of the EnKF in the background of reservoir engineering.

### 2.3.1 State vectors for reservoir models

In our applications, the state vector for a reservoir model consists of all the variables that are uncertain, and that need to be specified in order to run a reservoir simulator. In a typical two-phase reservoir fluid flow and transport problem, we might have a state vector of the form,

$$y_{j,k} = \begin{bmatrix} \phi \\ \ln K \\ P \\ S_w \\ \text{WCT} \\ \vdots \end{bmatrix}_{j,k}, \quad (2.18)$$

where the first subscript  $j$  is the ensemble member index; it counts from 1 to the size of the ensemble,  $N_e$ ; the second subscript  $k$  is still the integer time index for measurement time  $t_k$ ;  $\phi$  is the porosity vector which is composed of the porosity values at each gridblock for a heterogenous reservoir;  $\ln K$  is the natural logarithm of the absolute permeability vector;  $P$  is the pressure vector;  $S_w$  is the water saturation vector; WCT is computed water cut. The state vector can be partitioned into model parameters,  $m$ , ( $\phi$  and  $\ln K$ ) that do not change with time, state variables,  $f(m)$ , ( $P$

and  $S_w$ ) that change substantially with time, and predictions of observations,  $g(m)$ , (WCT and etc), and written in its compact form as Eq. 2.6. The variables in the state vectors can be easily extended for more complex reservoir flow problems.

When  $k = 0$ , the initial pressure vector,  $P_0$ , and saturation vector,  $S_{w,0}$ , are given as reservoir initial conditions. In current EnKF applications, the initial conditions are not considered as uncertain, though there might be big uncertainties associated with them in reality. All the ensemble models use the same initial conditions. So at the initial time  $t = 0$ , only an ensemble of model parameters,  $m_j$  ( $j = 1, 2, \dots, N_e$ ), is generated by honoring the prior knowledge of these variables.

### 2.3.2 Observations for ensemble models

Burgers et al. (1998) showed that observations must be treated as random variables for the analysis scheme to be consistent and to avoid a too low variance after data assimilation. Random perturbations are needed to add into the measured data and create a suite of observation sets for the ensemble models. Denote the observation at time  $t_k$  of the  $j$ th ensemble model  $d_{\text{obs},j,k}$ . The relationship between the perturbed observation and the true state vector can be written as

$$d_{\text{obs},j,k} = H_k y^{\text{true}} + \epsilon_k + \nu_{j,k} = d_{\text{obs},k} + \nu_{j,k}, \quad (2.19)$$

where  $\epsilon_k$  is the unknown measurement error;  $\nu_{j,k}$  is the perturbation added to the noisy measured data,  $d_{\text{obs},k}$ , to form the observations for the  $j$ th ensemble member; both  $\epsilon_k$  and  $\nu_{j,k}$  are Gaussian distributions with mean 0 and covariance  $C_{D,k}$ , i.e.  $E[\epsilon_k \epsilon_k^T] = E[\nu_{j,k} \nu_{j,k}^T] = C_{D,k}$ ; the noise and perturbation are uncorrelated, i.e.  $E[\epsilon_k \nu_{j,k}^T] = 0$  for all  $j$  and  $k$ .

### 2.3.3 Forecast step for reservoir models

Section 2.1.1 describes the dynamic equations governing the three-phase hydrocarbon fluid flow and transport in porous media for black oil models. Reservoir simulators

are used to simulate the flow and transport processes. In this dissertation, both Schlumberger's commercial reservoir simulator **ECLIPSE** and Chevron's in-house reservoir simulator **CHEARS** (Chevron Extended Application Reservoir Simulator) are used.

As introduced in the Kalman filter section, the forecast step for the EnKF propagates the state vectors forward in time from a previous measurement time, using the estimates of the variables conditional to all the observed data up that time, to current measurement time

$$y_{j,k}^p = \psi(y_{j,k-1}^u) \quad (j = 1, 2, \dots, N_e), \quad (2.20)$$

where  $\psi$  represents solving the reservoir flow and transport equations. Note that in the forecast step, only the state variables (pressures and saturations) and the predictions of data vary with time, and the model parameters (porosities and permeabilities) do not change from the propagation, i.e.  $m_{j,k}^p = m_{j,k-1}^u$ .

In the EnKF, the propagation is for all ensemble models, which may take considerable time especially when the size of the reservoir is big and the fluid flow is complex. However, because of the independence of the ensemble models at the forecast step, the propagation of multiple models can be proceeded simultaneously using parallel computing techniques. The techniques implemented for the examples in this work will be explained in Section 2.5.

For non-linear dynamics, when the governing equations are solved from previous measurement time  $t_{k-1}$  to current measurement time  $t_k$ , it usually take a few sub-timesteps for the equations to be advance to time  $t_k$ . However, we do not need to store the states of the system at these intermediate steps. Only the state at the exact time  $t_k$  needs to be stored.

### 2.3.4 Update step

With the newly acquired data, the state vectors are updated using their prior quantities from the forecast step and a weighted innovation term. The innovation term is the difference between the observed and predicted data. The weighting matrix is the Kalman gain matrix and denoted by  $K_{e,k}$ .

$$y_{j,k}^u = y_{j,k}^p + K_{e,k}(d_{\text{obs},j,k} - H_k y_{j,k}^p) \quad (j = 1, 2, \dots, N_e), \quad (2.21)$$

where the subscript  $e$  on the Kalman gain matrix represent quantities that are computed from the ensemble. The Kalman gain matrix has exactly the same form as that in the Kalman filter (Eq. 2.15)

$$K_{e,k} = C_{Y,e,k}^p H_k^T (H_k C_{Y,e,k}^p H_k^T + C_{D,k})^{-1}. \quad (2.22)$$

In the Kalman filter, the *prior* covariance matrix,  $C_{Y,k}^p$ , is explicitly computed with the time evolution equation (Eq. 2.13). While in the EnKF, the covariance matrix,  $C_{Y,e,k}^p$ , is estimated from the ensemble at any time using the standard statistical formula:

$$C_{Y,e,k}^p = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (y_{j,k}^p - \bar{y}_k^p)(y_{j,k}^p - \bar{y}_k^p)^T, \quad (2.23)$$

where  $\bar{y}_k^p$  is the mean vector of the variables in the state vectors and computed from the ensemble,  $\bar{y}_k^p = \frac{1}{N_e} \sum_{j=1}^{N_e} y_{j,k}^p$ .

In more detail, any element  $c_{m,l}$  in the covariance matrix  $C_{Y,e,k}^p$  can be computed as following

$$c_{m,l} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (x_{m,j} - \bar{x}_m)(x_{l,j} - \bar{x}_l) \quad (m, l = 1, 2, 3, \dots, N_y), \quad (2.24)$$

where  $x_{m,j}$  and  $x_{l,j}$  are the  $m$ th and  $l$ th variables, respectively, in the  $j$ th state vectors;  $\bar{x}_m$  and  $\bar{x}_l$  are the means of the  $m$ th and  $l$ th variables;  $c_{m,l}$  is the computed covariance between the  $m$ th and  $l$ th variables in state vectors.

In practice, it is not necessary to compute an approximation of the covariance matrix, because only the product of  $C_{Y,e,k}^p H_k^T$  is required to compute the Kalman gain matrix (Eq. 2.22). The covariance matrix can have fairly large dimensions,  $N_{y,k} \times N_{y,k}$ , for field-scale problems, while  $C_{Y,e,k}^p H_k^T$  has reduced dimensions,  $N_{y,k} \times N_{d,k}$ .  $N_{y,k}$  is the length of a state vector and  $N_{d,k}$  is the number of observed data, both at time  $t_k$ .

## 2.4 Implementation of the EnKF

The procedure in the following is intended to provide a general idea of the basic computations and the reasons for computational efficiency of the EnKF algorithm.

1. Propagate the state vectors forward in time using a numerical reservoir simulator (Eq. 2.20).
2. Assemble the state vectors into matrix  $Y$

$$Y = [y_1 \ y_2 \ \dots \ y_j \ \dots \ y_{N_e}] . \quad (2.25)$$

3. Compute the vector of mean values for variables in the state vectors.

$$\bar{y} = \frac{1}{N_e} \sum_{j=1}^{N_e} y_j . \quad (2.26)$$

4. Compute the matrix of deviations,  $\Delta Y$ , of the variables in the state vectors from the mean

$$\Delta Y = [\Delta y_1 \ \Delta y_2 \ \dots \ \Delta y_j \ \dots \ \Delta y_{N_e}] . \quad (2.27)$$

The  $j$ th column of  $\Delta Y$  is

$$\Delta y_j = y_j - \bar{y} . \quad (2.28)$$

5. Compute the product of the transpose of the deviation matrix,  $\Delta Y^T$ , with the transpose of the operator matrix,  $H^T$ . As mentioned in Section 2.1.2, this requires only the selection of a few columns of  $\Delta Y^T$ .

$$A = \Delta Y^T H^T . \quad (2.29)$$

6. Compute the Kalman gain matrix

$$K_e = \frac{1}{N_e - 1} \Delta Y A \left( \frac{1}{N_e - 1} A^T A + C_D \right)^{-1}. \quad (2.30)$$

7. Perturb the observed data (Eq. 2.19).

8. Update the state vectors (Eq. 2.21).

9. If there are additional data, return to **Step 1**.

The procedure provided above is simplified as the calculations of matrices are not given. Below explains the details of some of the important terms involved in the computation of the Kalman gain matrix. Adopting the general form of a state vector in Eq. 2.6 for non-linear systems, we write  $y_j$  as

$$y_j = \begin{bmatrix} m_j \\ f(m_j) \\ g(m_j) \end{bmatrix} \quad (j = 1, 2, \dots, N_e). \quad (2.31)$$

For discretized gridblocks, each term itself in the state vector  $y_j$  is a vector. Define each of them as

$$m_j = \begin{bmatrix} (m_j)_1 \\ (m_j)_2 \\ \vdots \\ (m_j)_i \\ \vdots \\ (m_j)_{N_m} \end{bmatrix} \quad (j = 1, 2, \dots, N_e), \quad (2.32)$$

$$f(m_j) = \begin{bmatrix} (f(m_j))_1 \\ (f(m_j))_2 \\ \vdots \\ (f(m_j))_i \\ \vdots \\ (f(m_j))_{N_f} \end{bmatrix} \quad (j = 1, 2, \dots, N_e), \quad (2.33)$$

$$g(m_j) = \begin{bmatrix} (g(m_j))_1 \\ (g(m_j))_2 \\ \vdots \\ (g(m_j))_i \\ \vdots \\ (g(m_j))_{N_d} \end{bmatrix} \quad (j = 1, 2, \dots, N_e), \quad (2.34)$$

where the subscript  $i$  outside the parentheses represents that the value is the  $i$ th component of the vector;  $N_m$  is the number of model parameters;  $N_f$  is the number of state variables;  $N_d$  is the number of data. The length of a state vector then is  $N_y = N_m + N_f + N_d$ .

The mean vector of variables in the state vectors is

$$\bar{y} = \frac{1}{N_e} \sum_{j=1}^{N_e} y_j = \begin{bmatrix} \bar{m} \\ \overline{f(\cdot)} \\ \overline{g(\cdot)} \end{bmatrix}, \quad (2.35)$$

where  $\bar{m} = \frac{1}{N_e} \sum_{j=1}^{N_e} m_j$ ,  $\overline{f(\cdot)} = \frac{1}{N_e} \sum_{j=1}^{N_e} f(m_j)$ , and  $\overline{g(\cdot)} = \frac{1}{N_e} \sum_{j=1}^{N_e} g(m_j)$ .

The deviation matrix  $\Delta Y$  is a matrix with dimensions of  $N_y \times N_e$

$$\Delta Y = \begin{bmatrix} m_1 - \bar{m} & \dots & m_{N_e} - \bar{m} \\ f(m_1) - \overline{f(\cdot)} & \dots & f(m_{N_e}) - \overline{f(\cdot)} \\ g(m_1) - \overline{g(\cdot)} & \dots & g(m_{N_e}) - \overline{g(\cdot)} \end{bmatrix}. \quad (2.36)$$

The matrix  $A = \Delta Y^T H^T$  has dimensions of  $N_e \times N_d$

$$\begin{aligned} A &= \Delta Y^T H^T \\ &= \begin{bmatrix} (g(m_1) - \overline{g(\cdot)})^T \\ \vdots \\ (g(m_{N_e}) - \overline{g(\cdot)})^T \end{bmatrix}. \end{aligned} \quad (2.37)$$

The product of  $\frac{1}{N_e-1}\Delta Y A$  has dimensions of  $N_y \times N_d$

$$\begin{aligned}
\frac{1}{N_e-1}\Delta Y A &= \frac{1}{N_e-1}\Delta Y \Delta Y^T H^T = C_Y H^T \\
&= \frac{1}{N_e-1} \begin{bmatrix} m_1 - \bar{m} & \dots & m_{N_e} - \bar{m} \\ f(m_1) - \overline{f(\cdot)} & \dots & f(m_{N_e}) - \overline{f(\cdot)} \\ g(m_1) - \overline{g(\cdot)} & \dots & g(m_{N_e}) - \overline{g(\cdot)} \end{bmatrix} \begin{bmatrix} (g(m_1) - \overline{g(\cdot)})^T \\ \vdots \\ (g(m_{N_e}) - \overline{g(\cdot)})^T \end{bmatrix} \\
&= \frac{1}{N_e-1} \begin{bmatrix} \sum_{j=1}^{N_e} (m_j - \bar{m})(g(m_j) - \overline{g(\cdot)})^T \\ \sum_{j=1}^{N_e} (f(m_j) - \overline{f(\cdot)})(g(m_j) - \overline{g(\cdot)})^T \\ \sum_{j=1}^{N_e} (g(m_j) - \overline{g(\cdot)})(g(m_j) - \overline{g(\cdot)})^T \end{bmatrix}. \tag{2.38}
\end{aligned}$$

The product of  $\frac{1}{N_e-1}A^T A$  has dimensions of  $N_d \times N_d$

$$\begin{aligned}
\frac{1}{N_e-1}A^T A &= \frac{1}{N_e-1}H\Delta Y \Delta Y^T H^T = HC_Y H^T \\
&= \frac{1}{N_e-1} \left[ \sum_{j=1}^{N_e} (g(m_j) - \overline{g(\cdot)})(g(m_j) - \overline{g(\cdot)})^T \right]. \tag{2.39}
\end{aligned}$$

With Eqs. 2.38 and 2.39, the Kalman gain matrix can be written as

$$K_e = \frac{\frac{1}{N_e-1}}{\left[ \frac{1}{N_e-1} \sum_{j=1}^{N_e} (g(m_j) - \overline{g(\cdot)})(g(m_j) - \overline{g(\cdot)})^T + C_D \right]^{-1}} \begin{bmatrix} \sum_{j=1}^{N_e} (m_j - \bar{m})(g(m_j) - \overline{g(\cdot)})^T \\ \sum_{j=1}^{N_e} (f(m_j) - \overline{f(\cdot)})(g(m_j) - \overline{g(\cdot)})^T \\ \sum_{j=1}^{N_e} (g(m_j) - \overline{g(\cdot)})(g(m_j) - \overline{g(\cdot)})^T \end{bmatrix}. \tag{2.40}$$

The dimensions of the Kalman gain matrix are  $N_y \times N_d$ .

There are three numbers related to the dimensions of matrices:  $N_y$ ,  $N_d$ , and  $N_e$ .

For petroleum reservoirs,  $N_y$  can be easily gone over  $10^6$  for a field-scale problem;  $N_d$  is the number of observed data available at one measurement time; usually the number of production data measured from wells at one time is moderate, a few tens, for example; considering the plausibility of computational time to perform history matching using the EnKF,  $N_e$  has to be kept as an affordable number, around the order of 100. Generally for our applications,  $N_y \gg N_e$ ,  $N_y \gg N_d$ , and  $N_e > N_d$ . For

this case, the propagation of the state vectors dominates the computational cost in the EnKF. The total computation time equals to the simulation run time for all of the ensemble models plus some overhead time involved in matrix calculation at the update steps.

However if the number of data exceeds the size of ensemble, i.e.  $N_d > N_e$ , special techniques must be employed to handle the assimilation of large amount of data (Evensen, 2004; Kepert, 2004; Skjervheim et al., 2005).

## 2.5 Parallel Implementation of the EnKF

The independence of the ensemble members (Eqs. 2.20 and 2.21) allows the benefit of parallelism for the EnKF implementation so that the computational cost is significantly reduced.

In this dissertation, depending on the infrastructure of computing network available at hand, the parallelism was implemented on both distributed and shared memory networks. The distributed memory network used is the PENTIUM4 XEON LINUX cluster at Oklahoma University Supercomputing Center for Education and Research (<http://oscer.ou.edu>). The shared memory network used is the 64-node IBM REGATTA workstation in Chevron's Energy Technology Company.

When multiple processors are used to perform a task, loading balance on each of the multiple processors is an important factor that affects the computation efficiency. However, for the EnKF, it is straightforward to divide the task for all ensemble members into subtasks. With the assumption that every member of the ensemble takes more or less equal time for computations involved, we assigned each processor the same amount of local ensemble members for balancing the overall workload.

Firstly the implementation on the distributed memory network will be introduced. Both the propagation (Eq. 2.20) and the update (Eq. 2.21) of multiple state vectors are done simultaneously with multiple processors. However, the computation of the

Kalman gain matrix (Eq. 2.22) needs collective information of the whole ensemble. Thus communications of different processors are needed for the calculation.

The communications are achieved through **Message-Passing Interface (MPI)**. Currently, it is the most widely used method of programming many types of parallel computers. It contains a library of functions that explicitly transmit data from one processor to another. The principal drawback of using **MPI** is that it requires programmers to design and develop their programs with great care (Pacheco, 1997).

For explanation purposes, let's call one of the assemble processors master processor, and the rest of them slave processors. The application starts by distributing equal amount of state vectors to every processor of the assemble processors (**MPI\_Scatterv**). The following is a procedure for the parallel implementation at one measurement time and **MPI** functions used to achieve the desired communications.

1. Each processor propagates the state vectors assigned to it forward in time. The multiple processors perform the tasks simultaneously.
2. Wait until all processors finish the tasks of the propagation for all their local members. Bring the new state vectors at all processors together to the master processor (**MPI\_Gatherv**).
3. Compute the Kalman gain matrix on the master processor (Eq. 2.22).
4. Send the Kalman gain matrix from the master processor to all of the slave processors (**MPI\_Bcast**).
5. Each processor updates its associated local state vectors (Eq. 2.21). The tasks are also done simultaneously.

The procedure above needs to be repeated until the last data are assimilated.

For the implementation on the shared memory network, only the forecast steps are paralleled. Multiple tasks for propagating the state vectors are submitted to the

IBM REGATTA workstation at the same time. In Chevron, the network was set up in a way that when a simulation run on a computing node is finished, the output files of the simulation are copied to public storage devices accessible by any computer in the network. Therefore, communication is not concerned by users.

It is worth noting that due to the requirement of the EnKF algorithm, the simulation time is broken into a number of time intervals between measurement times. To finish the run for the total simulation time for a single ensemble member, multiple submissions are needed. The number of submissions needed for one ensemble member to finish an entire simulation run is equal to the number of the measurement times. Different from using a PC, a submitted job to a multiple-processor cluster forms a queue before executing. Depending on the load at the cluster, sometimes, the submitted job can be delayed in the queue. When estimating the total computation expense of performing history matching using EnKF, the queue time should be taken into consideration.

## 2.6 Heuristics Measures of the EnKF Performance

At each measurement time and after the update step, the EnKF produces multiple numbers of models conditional to observed data up to that time. These ensemble models composed by a relatively small sample size is only a tiny fraction of a wide space of possible solutions. For the EnKF to perform well, the samples need to be representative of the probability space throughout all measurement times. There are some heuristics measures used in the EnKF literature to evaluate the performance of the method.

With the ensemble models, we can compute the spread of the samples at each gridblock

$$\sigma_i = \sqrt{\frac{1}{N_e} \sum_{j=1}^{N_e} (x_{i,j} - \bar{x}_i)^2} \quad (i = 1, 2, \dots, N), \quad (2.41)$$

where  $i$  is the index for gridblock and  $N$  is the number of gridblocks;  $x$  can be porosity,

permeability, and any other variables of interest;  $\bar{x}$  is the ensemble mean. A scalar measure of the spread would be

$$\begin{aligned}\sigma &= \sqrt{\frac{1}{N} \sum_{i=1}^N \sigma_i^2} \\ &= \sqrt{\frac{1}{N} \sum_{i=1}^N \frac{1}{N_e} \sum_{j=1}^{N_e} (x_{i,j} - \bar{x}_i)^2}. \\ &= \sqrt{\frac{1}{N} \frac{1}{N_e} \sum_{i=1}^N \sum_{j=1}^{N_e} (x_{i,j} - \bar{x}_i)^2}\end{aligned}\quad (2.42)$$

For synthetic cases, with known true model variables, we can compute the grid-block root mean square error (RMSE) of the estimates

$$\text{RMSE}_i = \sqrt{\frac{1}{N_e} \sum_{j=1}^{N_e} (x_{i,j} - x_i^{\text{true}})^2} \quad (i = 1, 2, \dots, N), \quad (2.43)$$

where  $x^{\text{true}}$  is the true values of variables of interest. A scalar measure of the error would be

$$\begin{aligned}\text{RMSE} &= \sqrt{\frac{1}{N} \sum_{i=1}^N \text{RMSE}_i^2} \\ &= \sqrt{\frac{1}{N} \sum_{i=1}^N \frac{1}{N_e} \sum_{j=1}^{N_e} (x_{i,j} - x_i^{\text{true}})^2} \\ &= \sqrt{\frac{1}{N} \frac{1}{N_e} \sum_{i=1}^N \sum_{j=1}^{N_e} (x_{i,j} - x_i^{\text{true}})^2}\end{aligned}\quad (2.44)$$

With data assimilation, we would expect that both the spread and error of the ensemble estimates decrease. An decreasing error reflects the improvement of the estimates. The spread is regarded as the uncertainty associated with the estimates. However, we found through some examples in this work, the spread in the estimates generally become smaller with time, but at some measurement time, the error in the estimates can jump up again even it has been decreasing at all previous times, which is referred as “divergence” in the traditional Kalman filter. When the variance of

the ensemble become too small, the uncertainty estimated from them is no longer representative of the true uncertainty.

# **CHAPTER III**

## **INVESTIGATION OF THE ENSEMBLE KALMAN FILTER TO APPLICATIONS OF CONTINUOUS RESERVOIR SIMULATION MODELS UPDATING**

### **3.1 Introduction**

There are at least two primary concerns with the application of the EnKF to the problem of updating reservoir flow models. One is that whether it is possible to use a Kalman filter to make corrections to state variables in a problem for which the covariance almost certainly provides a poor representation of the distribution of variables. The second concern should be the representation of the covariance via a relatively small ensemble of state vectors may be inadequate.

Two synthetic problems are chosen to investigate the potential difficulties. The first one is a one-dimensional, two-phase waterflood problem with the first data assimilation occurring at a fairly late time. The water saturations take large values behind the water flood front, and small values ahead of the front. Its distribution is bi-modal in this case and is not well modeled by the mean and variance. The key issue in this case is the ability to update the water saturations realistically. In the second problem, the reservoir model is two-dimensional so that the number of state variables is increased substantially. The number of ensemble members is kept the same as for the one-dimensional problem, and the main concern is that the number of ensemble members used to create the covariance matrix is far less than the number

of state variables.

Besides the two primary concerns described above, we also checked two other issues related to the methodology of the EnKF which are different from traditional history matching methods: (1) EnKF updates model parameters (porosity and permeability) and state variables (pressure and saturation) simultaneously at update steps. In essence a linearized approximation to the reservoir simulator is used to make a prediction of the saturation changes and pressure changes that would result from the porosity and permeability changes. In this process, it is not clear if the constraints imposed by the dynamical equations are honored during the Kalman correction, and (2) EnKF updates model parameters continuously with data assimilation along the time line, it seems quite possible that the estimation of the model parameters at later times might no longer honor the observed data from earlier times. Both issues are tested on the two-dimensional problem.

In both test problems, the wells are constrained by bottom-hole pressures. The producers are produced at a constant pressure of 1900 *psig* (1 *psi* = 6,894.76 *pascal*) and injectors are injected at a constant pressure of 4500 *psig*. The reservoir gridblocks are uniform. The dimension of one gridblock is  $60 \times 60 \times 40 \text{ ft}^3$ . We are to utilize the observed data to estimate  $\phi$  and  $K$  at each gridblock. Since both cases are two-phase (oil and water) flow problems, the state vectors are in the same form

$$\begin{aligned} y &= [\phi_1, \dots, \phi_N, \ln K_1, \dots, \ln K_N, \\ &\quad P_1, \dots, P_N, S_{w,1}, \dots, S_{w,N}, d_1, \dots, d_{N_d}]^T. \end{aligned} \tag{3.1}$$

where  $N$  is number of gridblock. The ensemble size used for both cases is 40.

### 3.2 Example 1: One-dimensional Waterflood Problem

The reservoir in this test case is 32 gridblocks in length ( $N = 32$ ). Water is injected in gridblock 1 and fluids are produced from gridblock 32. Fig. 3.1 sketches the setup

for the problem. The data to be assimilated are water injection rate at the injector, water and oil production rates at the producer, and water saturation in gridblock 21. They are obtained every 10 days starting at day 110. Measurement errors are assumed to be Gaussian with mean 0 and standard deviation of 3% of the magnitude of the observed data in the rates, 0.01 in the water saturations.

| <b>1</b>     | <b>... ... ... ...</b> | <b>21</b>     | <b>... ... ...</b> | <b>32</b>            |
|--------------|------------------------|---------------|--------------------|----------------------|
| Injector     |                        | Obsv. well    |                    | Producer             |
| constant BHP |                        |               |                    | constant BHP         |
| measure WIR  |                        | measure $S_w$ |                    | measure OPR<br>& WPR |

**Figure 3.1:** Schematic setup for the one-dimensional waterflood problem. Note: BHP means bottom-hole pressure, WIR means water injection rate, OPR/WPR means oil/water production rate.

### 3.2.1 Generation of initial reservoir models

Table 3.1 summarizes the parameters used to generate the initial reservoir models.

|                                      | <b>Porosity</b> | <b>LnK (K in md)</b> |
|--------------------------------------|-----------------|----------------------|
| <b>Prior mean</b>                    | 0.2             | 5.5                  |
| <b>STD</b>                           | 0.04            | 0.5                  |
| <b>Covariance</b>                    | Exponential     | Exponential          |
| <b>Range (grid)</b>                  | 18              | 18                   |
| <b>Cross-correlation coefficient</b> | 0.5             | 0.5                  |

**Table 3.1:** Parameters used for generating initial reservoir models for the one-dimensional waterflood problem.

The exponential covariance function for three-dimensional random fields (Kitanidis, 1997) is

$$C(h_1, h_2, h_3) = \sigma^2 \exp\left(-3\sqrt{\left(\frac{h_1}{a_1}\right)^2 + \left(\frac{h_2}{a_2}\right)^2 + \left(\frac{h_3}{a_3}\right)^2}\right). \quad (3.2)$$

where  $\sigma$  is the standard deviation of a random field;  $h_1$ ,  $h_2$ , and  $h_3$  are the projections of lag distance between any two points at the field on the  $x$ -,  $y$ -, and  $z$ -direction, respectively;  $h_1 = x_2 - x_1$ ,  $h_2 = y_2 - y_1$ ,  $h_3 = z_2 - z_1$ ; with  $(x_1, y_1, z_1)$  being one point

on the three-dimensional field, and  $(x_2, y_2, z_2)$  being another point;  $a_1$ ,  $a_2$ , and  $a_3$  are the ranges of the random field on the three directions.

We can easily write the exponential covariance function for one-dimensional random fields,

$$C(h_1) = \sigma^2 \exp\left(-\frac{3|h_1|}{a_1}\right). \quad (3.3)$$

The initial covariance function for porosity in the one-dimensional discretized grid system is

$$\begin{aligned} C_{\text{Phi}}(i_1, i_2) &= \sigma_{\text{Phi}}^2 \exp\left(-\frac{3|i_2 - i_1|}{a_1}\right) \quad i_1, i_2 = 1, 2, \dots, 32; \\ &= \sigma_{\text{Phi}}^2 C \end{aligned} \quad (3.4)$$

similarly, the initial covariance function for  $\ln K$  is

$$\begin{aligned} C_{\text{LnK}}(i_1, i_2) &= \sigma_{\text{LnK}}^2 \exp\left(-\frac{3|i_2 - i_1|}{a_1}\right) \quad i_1, i_2 = 1, 2, \dots, 32; \\ &= \sigma_{\text{LnK}}^2 C \end{aligned} \quad (3.5)$$

the initial cross-covariance function for porosity and  $\ln K$  is

$$\begin{aligned} C_{\Phi, \text{LnK}}(i_1, i_2) &= \rho \sigma_{\Phi} \sigma_{\text{LnK}} \exp\left(-\frac{3|i_2 - i_1|}{a_1}\right) \quad i_1, i_2 = 1, 2, \dots, 32, \\ &= \rho \sigma_{\Phi} \sigma_{\text{LnK}} C \end{aligned} \quad (3.6)$$

where  $a_1$  is the range (correlation length), in units of grid, of the covariance function for both porosity and  $\ln K$ ;  $\sigma_{\Phi}$  and  $\sigma_{\text{LnK}}$  are the STDs of porosity and  $\ln K$ , respectively;  $\rho$  is the cross-correlation coefficient;  $C$  is covariance function for the one-dimensional random field with  $\sigma = 1$

$$C(i_1, i_2) = \exp\left(-\frac{3|i_2 - i_1|}{a_1}\right) \quad i_1, i_2 = 1, 2, \dots, 32. \quad (3.7)$$

For this small-scale problem, without concerning about the computational expense, Cholesky decomposition is applied to generate the initial reservoir models. The procedure for the generation is as followings:

1. Form the covariance matrix  $C$  of  $32 \times 32$  shown in Eq. 3.7.

2. Decompose it by Cholesky decomposition

$$C = LL^T. \quad (3.8)$$

where  $L$  is the unique “square root” of the covariance matrix  $C$  with Cholesky decomposition and an upper triangular matrix.

3. Generate the porosity fields

$$\Phi = \mu_\Phi + \sigma_\Phi L Z_1 \quad (3.9)$$

where  $\mu_\Phi$  is the prior mean of porosity;  $Z_1$  is a vector of uncorrelated random normal deviates,  $Z_1 \in N(0, I)$ .

4. Generate the  $\ln K$  fields

$$\text{LnK} = \mu_{\text{LnK}} + \sigma_{\text{LnK}} L(\rho Z_1 + \sqrt{1 - \rho^2} Z_2) \quad (3.10)$$

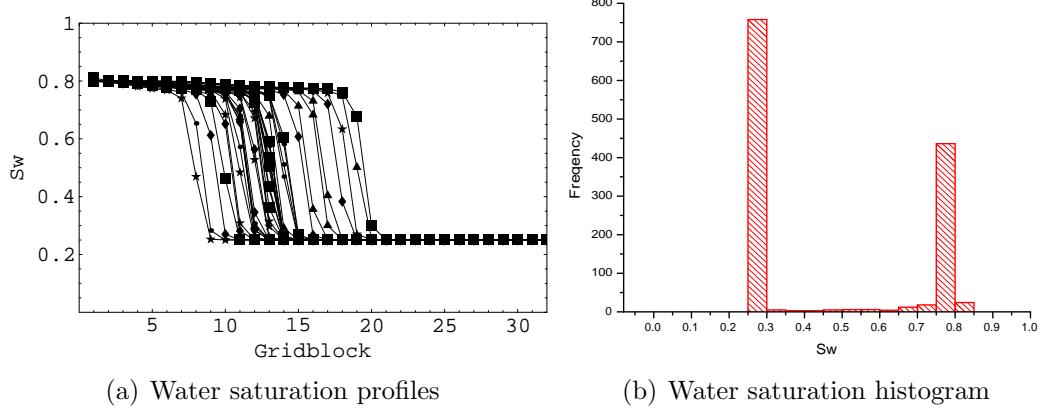
where  $\mu_{\text{LnK}}$  is the prior mean of  $\ln K$ ;  $Z_2$  is a vector of uncorrelated random normal deviates,  $Z_2 \in N(0, I)$ ;  $Z_1$  and  $Z_2$  are uncorrelated.

It can be easily shown that the generated porosity and  $\ln K$  fields in Eqs. 3.9 and 3.10 honor the covariance and cross-covariance functions in Eqs. 3.4 – 3.6.

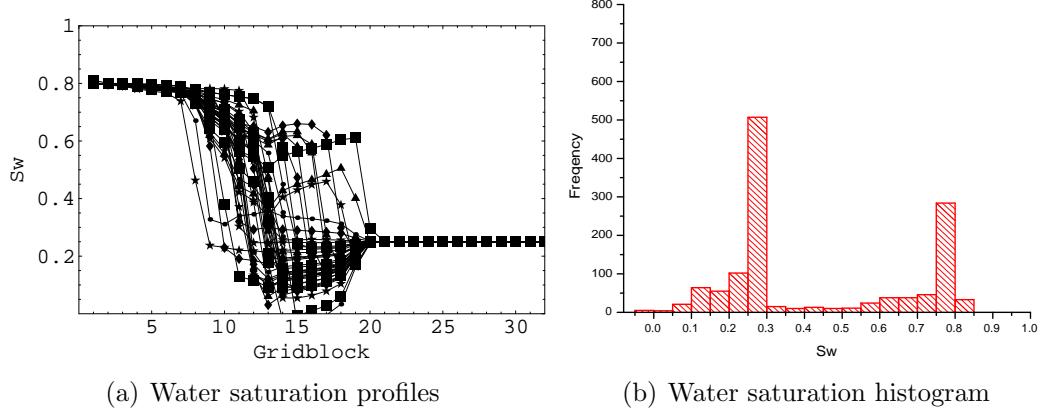
### 3.2.2 Solving problems with updated saturations

From Fig. 3.2(a), we can see that prior to the first application of Kalman correction, there is considerable variation in the location of the saturation fronts from the ensemble. The fastest model has the water moved to grid 20, while the slowest model only moves the water to grid 9. Fig. 3.2(b) shows the bi-modal distribution of the water saturation constructed based on the saturations in Fig. 3.2(a). After the correction (Fig. 3.3(a)), two obvious problems with the methodology are apparent. One is that some water saturations obtain nonphysical values, for example, water saturations at some gridblocks after water fronts are smaller than the connate water saturation,

$S_{wc} = 0.3$ , and some of them even go below 0. The other is that the water saturation profiles do not always decrease monotonically from high value at the injector to low value at the producer. The distribution of water saturation after the correction around  $S_{wc}$  is more like Gaussian than the original distribution before the correction, compare Fig. 3.3(b) to Fig. 3.2(b).



**Figure 3.2:** Water saturation profiles and histogram before the first application of the Kalman correction at 110 days.

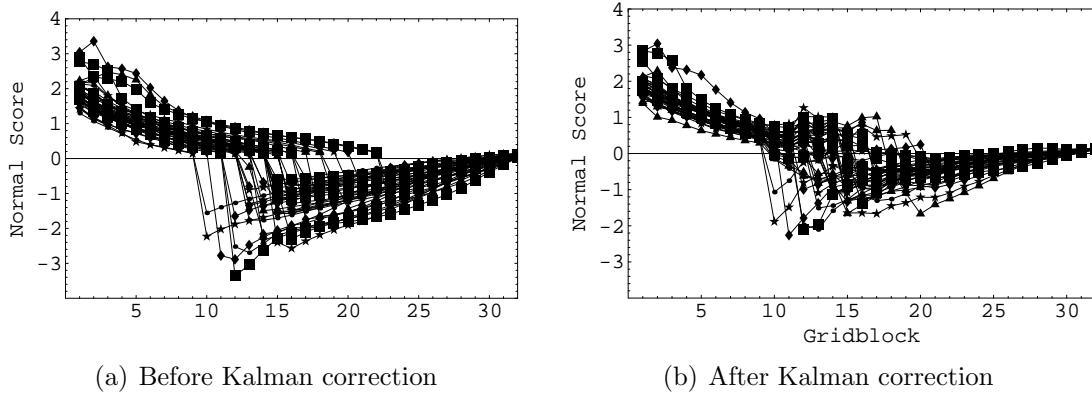


**Figure 3.3:** Water saturation profiles and histogram after the first application of the Kalman correction at 110 days.

Three methods for resolving the problems with the water saturations are considered: (1) transformation of the saturation to a variable whose univariate distribution is normal, (2) use of the location of saturation front, instead of saturation values as a state variable, and (3) iterating the update.

### 3.2.2.1 Normal score transform

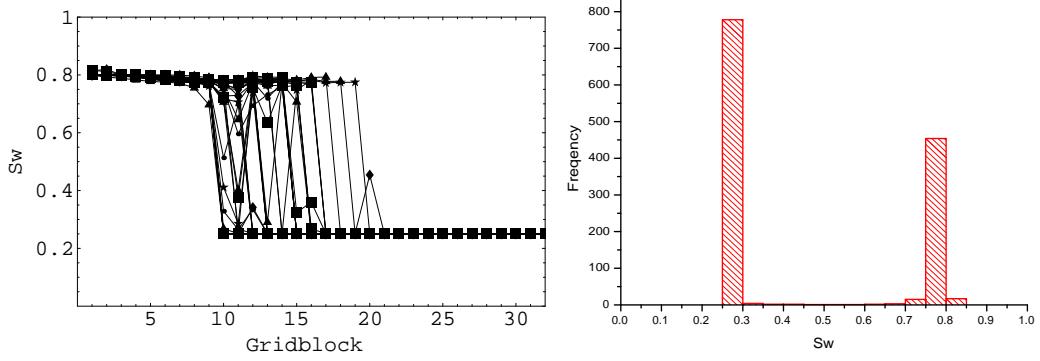
One obvious way to avoid the problem with nonphysical values of saturation is to use the normal score transform values (Goovaerts, 1997) of the saturations instead of the saturations themselves as state variables. The non-parametric transformation is constructed from the empirical cumulative density function (CDF) of gridblock saturations from the reservoir simulator. The variables after transformation are normally distributed. Figure 3.4 shows the results of before and after applying the EnKF to the profiles of the transformed variables at day 110. Fig. 3.5 plots the water saturation profiles after transforming back the updated normal score values and the histogram of these saturations. After applying the normal score transform, saturations do not go outside of the reasonable bounds any more. The distribution of the updated saturations is still bi-modal. Unfortunately, the saturations oscillate spatially between high and low values. The oscillations are not very obvious in Fig. 3.5(a), but become more pronounced at later times.



**Figure 3.4:** Normal score transformed profiles before and after the first application of the Kalman correction at 110 days.

### 3.2.2.2 Saturation front location as a state variable

Instead of applying an explicit transformation to the water saturations, we replaced the saturations in the state vector (Eq. 3.1) with a single variable which locates the



(a) Water saturation profiles after back transform

(b) Water saturation histogram

**Figure 3.5:** Water saturation profiles and histogram after transforming back the updated normal scores in Fig. 3.4(b) at 110 days.

water saturation front. The idea is to improve the linearity of the relationship between variables. Denotes the saturation front by  $X_{wf}$ , the state vector now becomes

$$\begin{aligned} y &= [\phi_1, \dots, \phi_N, \ln K_1, \dots, \ln K_N, \\ &P_1, \dots, P_N, X_{wf}, d_1 \dots, d_{N_d}]^T. \end{aligned} \quad (3.11)$$

The Kalman update formula is then used to adjust the water front locations. Saturation values at gridblock locations are computed by interpolation based on a table of saturation versus distance, calculated for each reservoir model of the ensemble. The following explains the interpolation in detail.

Adopting the same superscripts used in Chapter 2,  $p$  is used to represent prior values that are directly from the reservoir simulator and  $u$  is used to represent the updated values. In this notation,  $X_{wf}^p$  is the saturation front location calculated from prior saturations.  $X_{wf}^u$  is updated location after Kalman correction. Define  $r = X_{wf}^u / X_{sf}^p$ . Table 3.2 lists the prior saturation values versus distance. When the saturation front locations are updated, a new table is formed and shown in Table 3.3. It still has the same saturation values as Table 3.2, but the distance is updated. Based on the new table, the saturation at the regular grids can be interpolated, see Table 3.4.

| Grid         | $S_w$        |
|--------------|--------------|
| <b>0</b>     | $1 - S_{or}$ |
| <b>1</b>     | $S_{w,1}^p$  |
| <b>2</b>     | $S_{w,2}^p$  |
| <b>3</b>     | $S_{w,3}^p$  |
| $\vdots$     | $\vdots$     |
| <b>N</b>     | $S_{w,N}^p$  |
| <b>N + 1</b> | $S_{wc}$     |

**Table 3.2:** Prior water saturation versus distance.

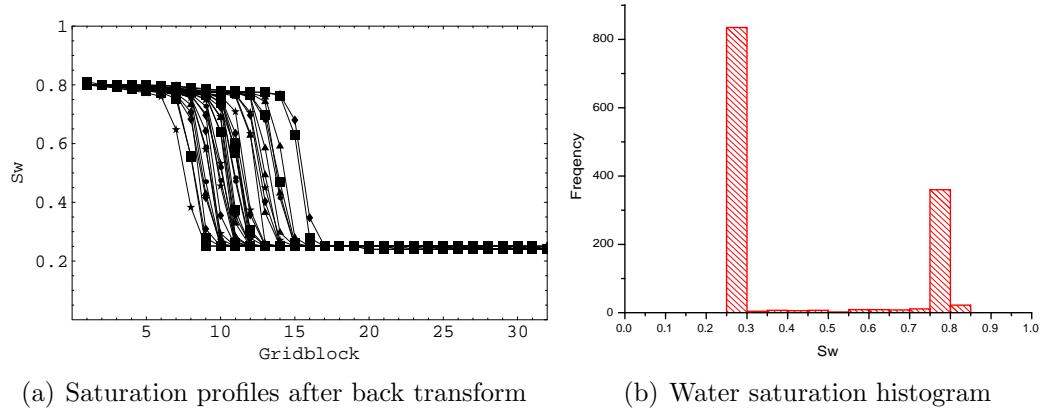
| Grid               | $S_w$        |
|--------------------|--------------|
| <b>0</b>           | $1 - S_{or}$ |
| <b>1 × r</b>       | $S_{w,1}^p$  |
| <b>2 × r</b>       | $S_{w,2}^p$  |
| <b>3 × r</b>       | $S_{w,3}^p$  |
| $\vdots$           | $\vdots$     |
| <b>N × r</b>       | $S_{w,N}^p$  |
| <b>(N + 1) × r</b> | $S_{wc}$     |

**Table 3.3:** Prior water saturation versus updated distance.

| Grid     | $S_w$       |
|----------|-------------|
| <b>1</b> | $S_{w,1}^u$ |
| <b>2</b> | $S_{w,2}^u$ |
| <b>3</b> | $S_{w,3}^u$ |
| $\vdots$ | $\vdots$    |
| <b>N</b> | $S_{w,N}^u$ |

**Table 3.4:** Interpolated water saturation versus distance. The interpolation is based on Table 3.3.

Fig. 3.6 shows the updated water saturation profiles and histogram after the first Kalman correction at day 110. The saturation values are within reasonable bounds and change monotonically from high at the injector to low at the producer. Comparing to the saturation profiles before Kalman correction in Fig. 3.2(a), the variation of water front locations from the ensemble is smaller now, while the bi-modal characteristic is still kept, see Fig. 3.6(b).



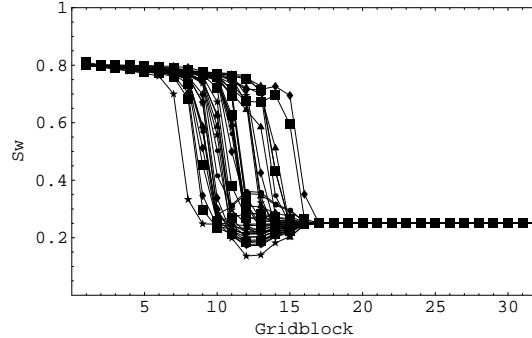
**Figure 3.6:** Water saturation profiles and histogram after the first application of the Kalman correction at 110 days, with saturation front location as a state variable.

### 3.2.2.3 Iterating the update

The front location method has been easily applied to the one-dimensional problem as illustrated above. However, for two- and three-dimensional problems, it is not straightforward to describe the saturation front locations. We attempted to use a more general approach. Wen and Chen (2005a) used a “conforming” check in their application to assure the consistency between the updated model parameters and state variables. Using a similar idea, whenever the updated saturations are detected out of their physical bound ( $S_{wc} < S_w < 1.0 - S_{or}$ ,  $S_{or}$  is residual oil saturation), we rerun the simulator from previous measurement time to re-compute the state variables (pressure and saturation) using the updated model parameters (porosity and permeability) obtained at current measurement time. The difference of the new

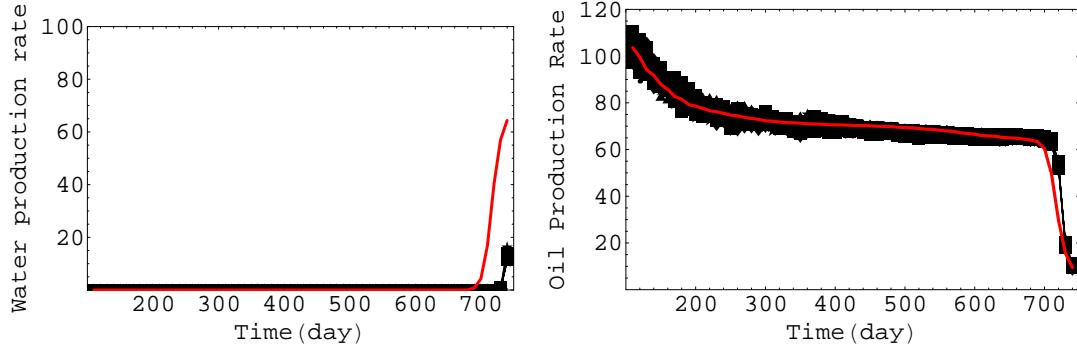
computational data and the observations is then used to update the model variables again. Iterations continue to be applied until the updated saturation profiles satisfy the physical bound, or the number of iterations exceeds a pre-set maximum number (three is used here). We noticed that, in this problem, the extra iterations are required only when the first data are assimilated (at 110 days), when the water front reaches the observation well in gridblock 21 (at 340 days), and some times following the water break-through at the observation well.

Fig. 3.7 displays the water saturation profiles at 110 days after correction with one extra iteration. Comparing with Fig. 3.3(a), the saturation profiles are improved but still problematic.

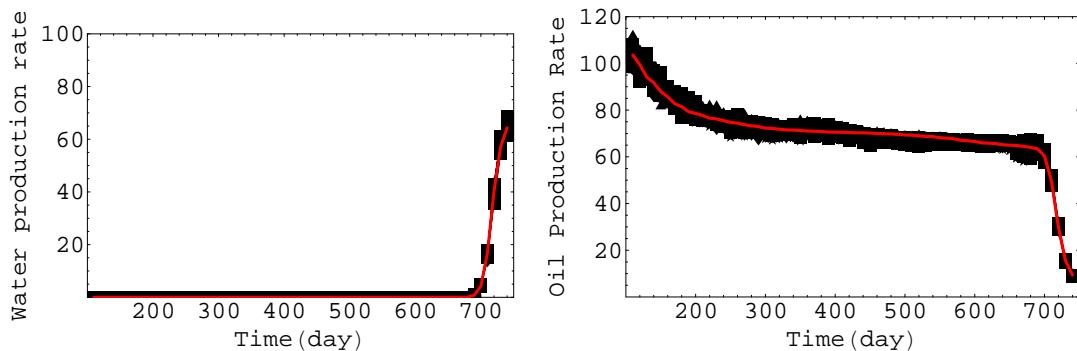


**Figure 3.7:** Saturation profiles after one extra iteration of Kalman correction at 110 days.

Figs. 3.8 and 3.9 compare the computed water and oil production rates (black color) with the corresponding observed values (red color), without and with iteration, respectively. From Figure 3.8, we can see that the timing of water break-through is not captured without iteration. However, with iteration, shown in Figure 3.9, not only do the saturation profiles get more realistic (see Fig. 3.7), but also the data are better honored.



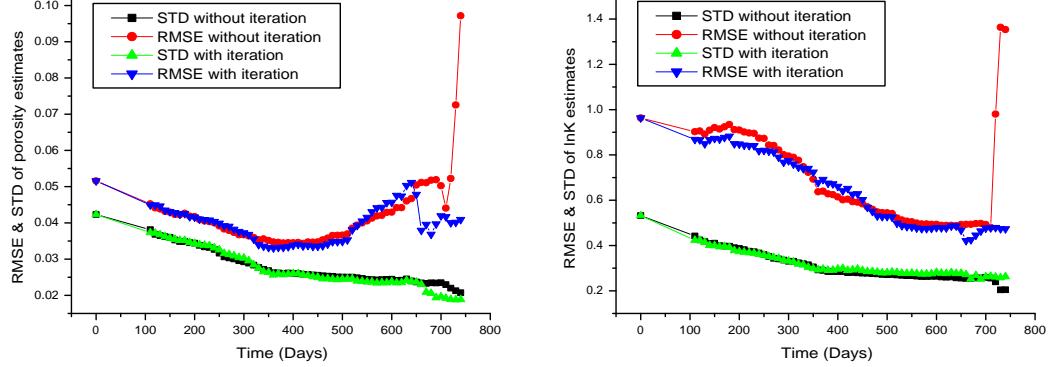
**Figure 3.8:** The observed and computed water and oil production rates from the ensemble after Kalman correction, without iteration.



**Figure 3.9:** The observed and computed water and oil production rates from the ensemble after Kalman correction, with one extra iteration.

### 3.2.3 Goodness of the model parameter estimates

Fig. 3.10 plots the RMSEs and spreads of both estimates of porosity and natural logarithm permeability. Because of the boundary conditions on flow (fixed pressure), and the type of data (rates), the observations do not inform about porosity until the rate of water advance can be observed. As permeability is adjusted to honor the rate measurements, the estimates of the permeability improve steadily and the porosity changes slightly because of its correlation with the permeability. Unfortunately, with the data assimilation at day 340 where the movement of water was first felt in the observation well, the porosity starts to move away from the truth in the example. The RMSEs of both estimates of the porosity and permeability jump up around the water break-through time at the producer (day 700) when no iteration is applied, which demonstrates the failure of assimilation at that time. However, with iteration, the large errors in both estimates are quenched.



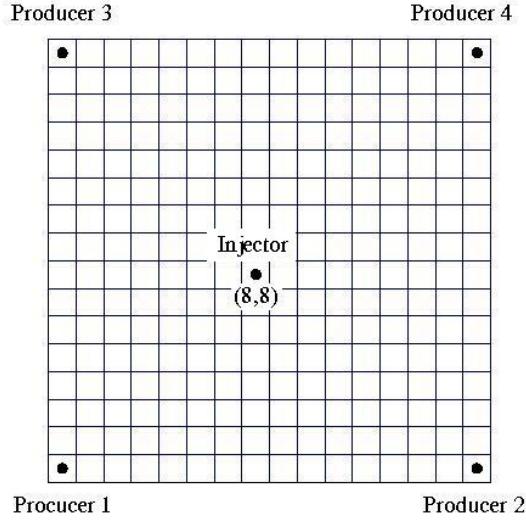
**Figure 3.10:** RMSEs and spreads of both the porosity (left) and natural logarithm permeability (right) estimates with and without iteration.

## 3.3 Example 2: Two-dimensional Waterflood Problem

A larger test is created to investigate the performance of the ensemble Kalman filter on a problem for which the number of state variables is far greater than the number

of the ensemble members. The grid for this test problem is  $16 \times 16$ , so the number of elements in each state vector is  $4 \times 256$  plus the number of data. We again used 40 ensemble members to represent the covariance between variables.

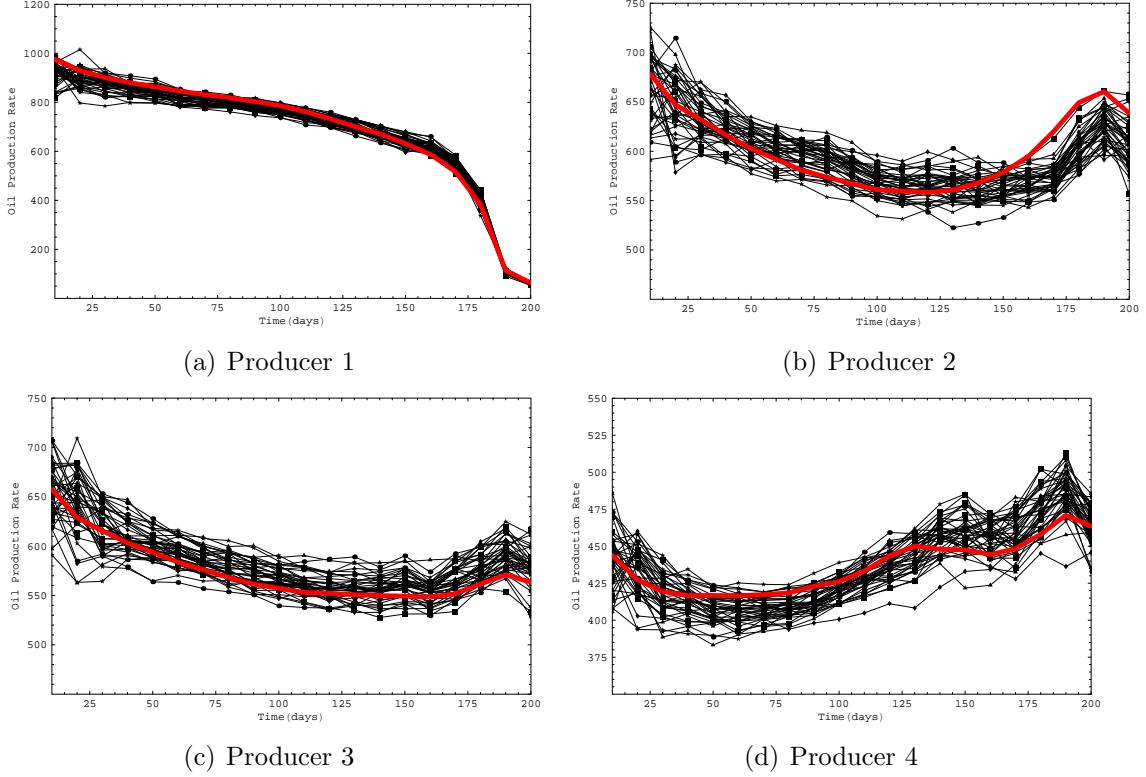
An injector is located approximately in the center of the reservoir, and four producers are located at the corners (Fig. 3.11). Oil and water rates at the producers and water injection rates at the injector are acquired every 10 days beginning at day 10. The standard deviations of the unbiased Gaussian measurement errors in the rates are assumed to be 10% of the magnitude of the observed data.



**Figure 3.11:** The locations of the injector and four producers for the two-dimensional waterflood problem.

### 3.3.1 History matching results

Fig. 3.12 shows the comparisons of the oil rates at the four production wells used for assimilation (red color), and the computed rates from all the ensemble members (black color) after Kalman correction. The oil rate at producer 1 drops fairly rapidly after 170 days because water breaks through the well at that time. The agreement between observations and predictions appear to be good, and the spread of the realizations is reasonable.

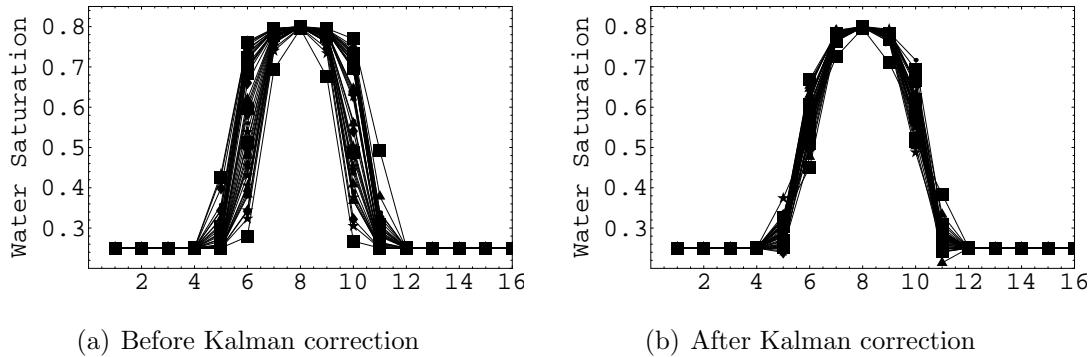


**Figure 3.12:** Oil production rates at the four producers after Kalman correction.

In the two-dimensional test, the water saturations are included in state vectors, even though they caused problems in the one-dimensional test example. Fig. 3.13 shows the saturation profiles on a cross-section through the injector at 10 days. The saturations after the update are mostly reasonable – we do not see profiles that are not decreasing away from the injector. The only indication of potential problem is one saturation profile is slightly below the connate water saturation. In this case, it does not appear to be necessary to change variables or iterate. We believe that this will generally be the case when changes to the state variables at the update steps are relatively small.

### 3.3.2 Model parameters estimates

The true permeability and porosity fields can be seen in Fig. 3.14. Both fields are generated using an exponential variogram model with a range of 10 gridblocks in



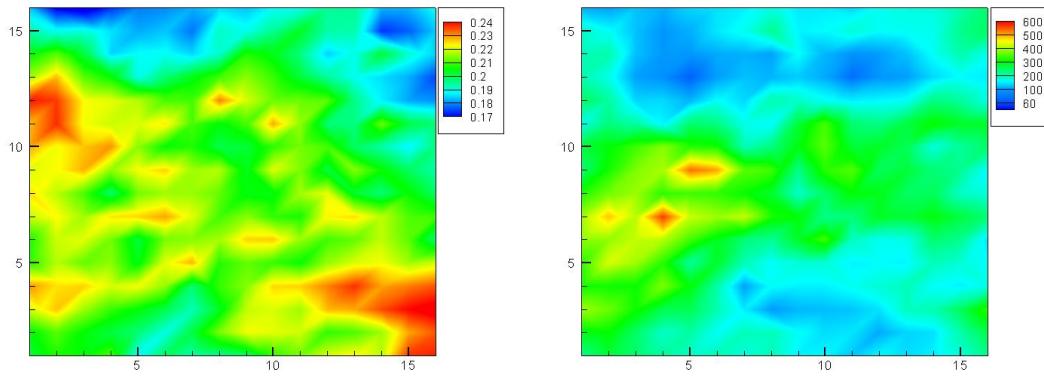
**Figure 3.13:** Water saturation profiles at the middle column before and after the first application of the Kalman correction at 10 days.

$x$ -direction and 5 gridblocks in  $y$ -direction. The mean values of the natural logarithm of permeability (md) and porosity are 5.5 and 0.2 respectively, and the standard deviations are 0.5 and 0.02 respectively. The correlation coefficient between log-permeability and porosity is 0.5.

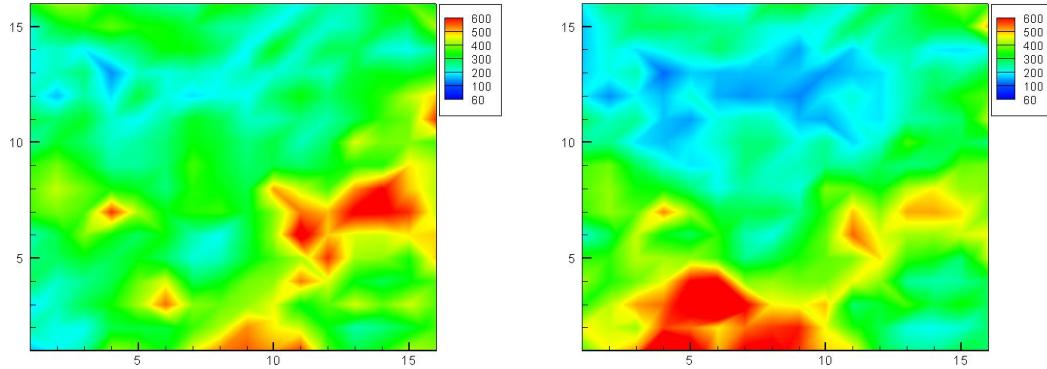
In the EnKF application, the permeability and porosity fields are continuously updated as data are assimilated. It is important that the corrections to the permeability and porosity fields be done in such a way that the plausibility of the fields, as indicated by the spatial auto-correlation, is maintained. Fig. 3.15 shows both the initial permeability field for model 1 (the first of the 40 ensemble members), and the permeability field after the first data assimilation at 10 days. Fig. 3.16 shows the porosity and permeability fields for model 1 after assimilation of 200 days of production data. These fields should be compared with the true fields in Fig. 3.14. Note in particular, the general reduction for permeability in model 1 occurs in the upper left quadrant. While the fields are not identical, it appears that the final property fields are not unrealistic, and that the changes are largely reasonable.

Fig. 3.17 shows the ensemble mean of permeability estimates at day 10 and day 200. Comparing to its counterparts of the individual model (model 1), the mean estimates are fairly smooth. However the characteristics of the mean and the individual model are similar.

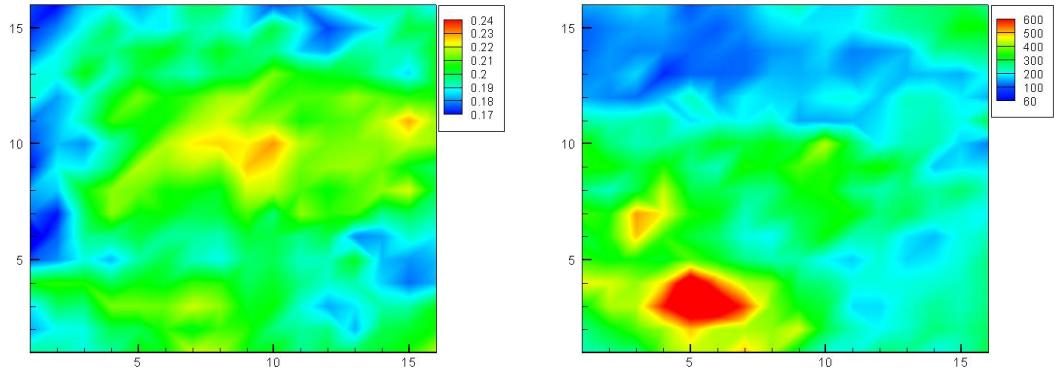
Fig. 3.18 plots the RMSEs and spreads of both estimates of porosity and natural logarithm permeability. Similarly as what has been seen in Fig. 3.10 for the one-dimensional example, both the errors in the estimates of porosity and permeability drops steadily with data assimilation. But the error in the porosity estimates starts to increase again after assimilating the data at day 170 when water breakthrough at Producer 1.



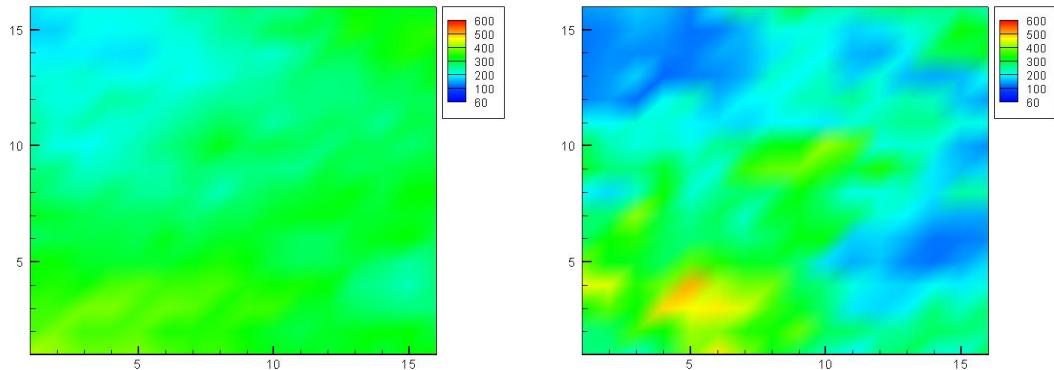
**Figure 3.14:** The true porosity (left) and permeability (right) fields.



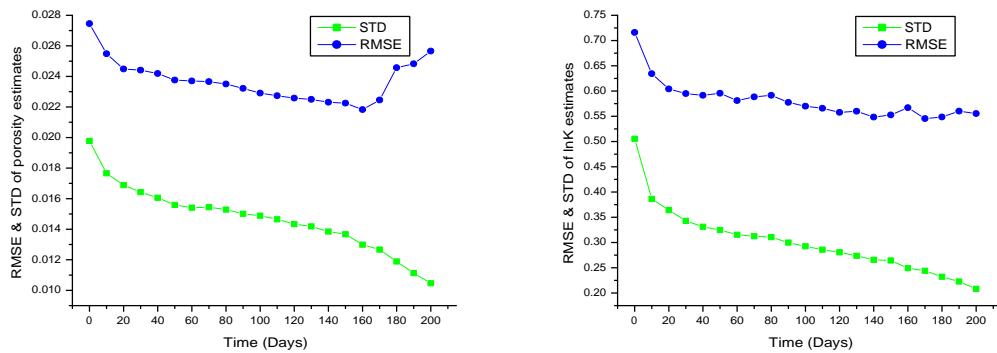
**Figure 3.15:** The permeability field for model 1 before (left) and after (right) the first application of Kalman correction at day 10.



**Figure 3.16:** The final porosity (left) and permeability (right) fields for model 1 after 200 days of Kalman correction.



**Figure 3.17:** The mean of the ensemble permeability estimates at day 10 (left) and day 200 (right) after Kalman correction.



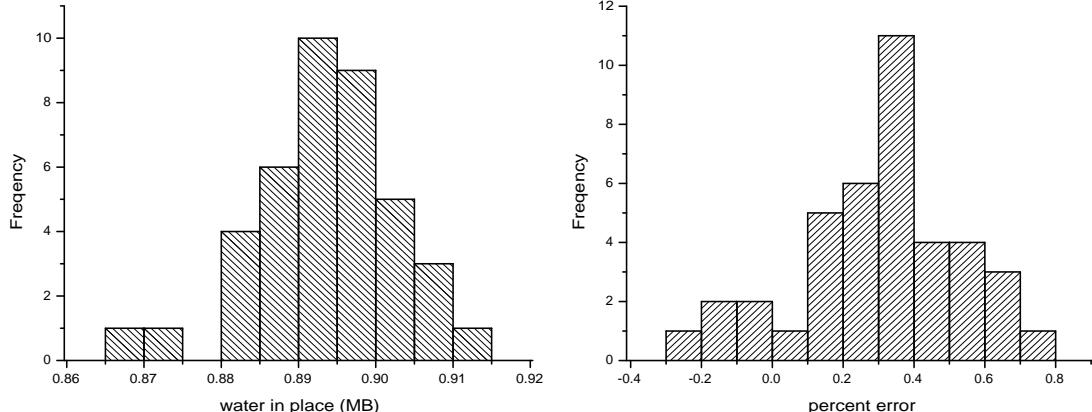
**Figure 3.18:** RMSEs and spreads of both the porosity (left) and natural logarithm permeability (right) estimates.

### 3.3.3 Consistency check

In traditional history matching, only the permeability and porosity fields would be updated using an optimizer, then the pressure and saturation fields that are consistent with the permeability and porosity fields would be computed by running the reservoir simulator. In the EnKF method, the permeability, porosity, pressure, and saturation are all updated simultaneously at update steps. In essence a linearized approximation to the simulator is used to make a prediction of the saturation changes and pressure changes that would result from the porosity and permeability changes.

It is not at all clear, however, that material balance is honored during the Kalman correction steps. In order to check the global material balance error, we computed the total water-in-place ( $\sum_{i=1}^N V_i \phi_i S_{w,i}$ ) for all of the final updated ensemble reservoir models at day 200 for two conditions: (1) using saturations after the Kalman correction at day 200, and (2) taking the final models to time 0 and rerun the reservoir simulation and using the re-computed saturations at day 200. If the difference in the computed water-in-place values, for each ensemble member, at these two conditions is small, it implies that the Kalman corrections are valid and material balance is generally honored. Fig. 3.19(a) plots the distribution of water-in-place in million barrel (1 MB = 158,987.3 m<sup>3</sup>) for all the final ensemble models computed at condition 2. Fig. 3.19(b) plots the distribution of the relative errors from both conditions. Most errors are less than 0.8% and regarded small enough. Of course the errors in individual gridblocks can be fairly large.

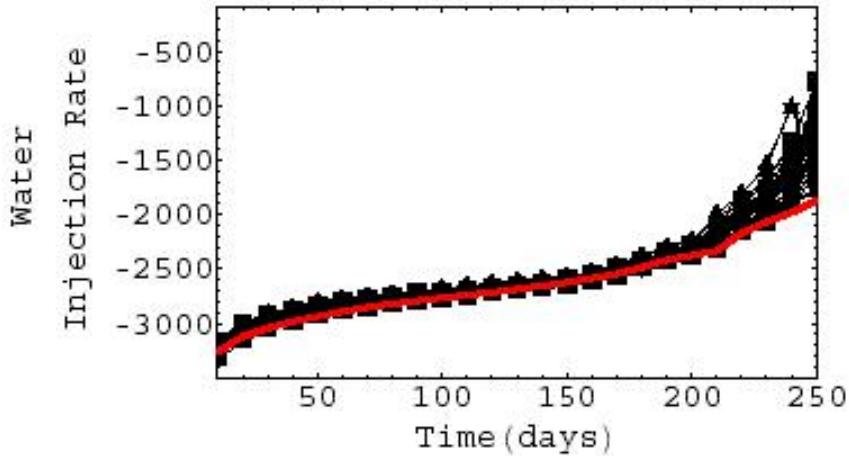
Another feature of the EnKF that can be disconcerting when used for reservoir models updating, is that the permeability and porosity fields are changing with time as additional data are assimilated. It seems quite possible that the final permeability and porosity fields that resulted from 200 days of data assimilation, might no longer honor the observed data at previous times. We have already seen that the rock property fields are much different after 200 days than they are after 10 days (compare



(a) Distribution of water-in-place computed at condition 2 (b) Relative errors in water-in-place computed from both conditions

**Figure 3.19:** Material balance checking after 200 days of data assimilation.

Fig. 3.16 to Fig. 3.15). In Fig. 3.20, we see a comparison of the actual data (red line) that are used in the data assimilation, with the computed values of the measurements (black lines) from each of the ensemble permeability and porosity fields at 200 days by taking them to time 0 and rerunning the reservoir simulator. In this case, the final updated models seem to produce data that honors the observations at all times.



**Figure 3.20:** Comparison of the observed water injection rates (red line) to its computed counterparts (black lines) from the final updated ensemble models at day 200 by rerunning the reservoir simulator from time 0.

## 3.4 Discussion

When the distribution of one of the state variables (saturation) is very far from normal, we find that it is possible to generate non-physical values after Kalman correction. If necessary, it could be resolved by a change of variables, or iteration. It is not necessary to make a change of variables for the two-dimensional problem, presumably because the changes in the state variables at any measurement time are relatively small.

The final models, at the end of data assimilation, while much different from the earlier models, are still consistent with the early data. The material balance errors is relatively small, at least compared to the uncertainty in the actual values. This is encouraging for the use of the method in reservoir history matching because we would want models that honored all data.

Because of the apparent need for small corrections at any measurement time, it seems that there will be some reservoir data assimilation problems that the EnKF will not be suitable for. It does, however, seem to be ideally suited for the assimilation of data from time series, for example, data from permanent sensors.

The results with a relatively small number of ensemble models are remarkably good. It seems that larger ensemble will be required for problems with larger amounts of data to be assimilated. At this time, we do not know if the results might begin to deteriorate if the assimilation period is much longer or if the models are much larger.

We also find from both examples that with the arrival of significant data at some measurement times, the error in the estimates of variables may grow after assimilating the data. This might be caused by the big changes in the state variables where the linearization used in the EnKF update formula is not acceptable.

The iterative scheme suggested by Wen and Chen (2005a) are used on the one-dimensional test problem and seems to produce improved history-matching results. However, it still needs further investigation. Chapter 5 will address the issue in detail.

# CHAPTER IV

## HISTORY MATCHING OF THE PUNQ-S3 RESERVOIR SIMULATION MODEL USING THE ENSEMBLE KALMAN FILTER

In Chapter 3, we have shown through two small-scale simple synthetic waterflood examples the plausibility of using the EnKF to history matching problems of reservoir applications. In this chapter, we will demonstrate the usage of this technique to a much bigger and more realistic reservoir model, PUNQ-S3.

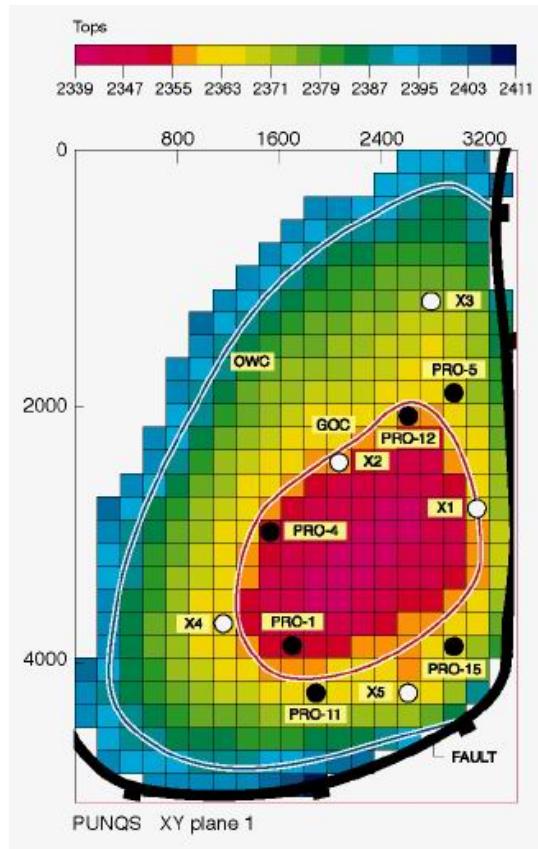
### 4.1 Problem Description

PUNQ-S3 is a synthetic reservoir engineering model constructed on the basis of a real field operated by Elf Exploration & Production. The **PUNQ** project is a joint effort of European companies, universities, and research centers supported by European Union to compare methods for quantifying uncertainty assessment in history matching. **PUNQ** is an acronym for **P**roduction forecasting with **U**Ncertainty **Q**uantification. Because the PUNQ-S3 model has been used as a test case for many inverse methods (Floris et al., 2001; Barker et al., 2001), it was chosen to evaluate the EnKF.

The detailed description of the PUNQ-S3 reservoir simulation model can be found in many resources (Floris et al., 2001; Barker et al., 2001). Here, only brief information related to the application in this dissertation is given. The reservoir simulation model contains  $19 \times 28 \times 5$  gridblocks, 1,761 of which are active. The gridblocks are uniform in areal sense,  $180 \times 180\text{ m}^2$ . The reservoir is bounded by a fault to the east and south,

and is in communication with a fairly strong aquifer on the west and north. Because of the strength of the aquifer, no injection wells are drilled. The top structural map of the reservoir is shown in Fig. 4.1. At the center of the map, there is a small gas cap in red color. The red closed line drawn near to the gas cap is the initial gas-oil contact. Six producing wells denoted by solid black dots are located close to the contact. Table 4.1 shows the well locations of the six producers and perforated layers of each well.

The PUNQ web page (<http://www.nitg.tno.nl/punq/cases/punqs3/index.htm>) reveals additional information to the public about the project, e.g. the true reservoir rock properties, the input deck file containing 16.5-year production schedule for a commercial reservoir simulator, **ECLIPSE**, and etc.



**Figure 4.1:** Top structural map of the PUNQ-S3 reservoir, from the PUNQ web page: <http://www.nitg.tno.nl/punq/cases/punqs3/index.htm>.

| Well name | Location: (x,y) | Perforated layers |
|-----------|-----------------|-------------------|
| PRO-1     | (10,22)         | 4, 5              |
| PRO-4     | (9,17)          | 4, 5              |
| PRO-5     | (17,11)         | 3, 4              |
| PRO-11    | (11,24)         | 3, 4              |
| PRO-12    | (15,12)         | 4, 5              |
| PRO-15    | (17,22)         | 4                 |

**Table 4.1:** Well locations of the six producers and perforated layers of each well.

## 4.2 Generation of Initial Realizations

The realizations for the five layers of the reservoir are generated independently using different spherical variogram models. Table 4.2 lists the parameters of the spherical variogram models used for different layers.

At the PUNQ web page, normalized porosity values at all the six well locations are provided. The normalized values are the normal scores when transforming non-Gaussian distributed values to a Gaussian distribution. Sequential Gaussian Simulation (Deutsch and Journel, 1998) is used to generate Gaussian random fields (GRFs) of the normalized porosity. The actual porosity fields are obtained by back transformation. Horizontal permeabilities are co-simulated assuming that the correlation coefficient between porosity and horizontal permeability is 0.8. The same correlation coefficient value is used between vertical and horizontal permeability. Table 4.3 summarizes the mean and standard deviation values of the generated realizations. The size of the realizations is 40. All the 40 realizations are conditional to hard data at the well locations.

From both Tables. 4.2 and 4.3, it can be seen that layers 1, 3, and 5 have similar pattern though the precise parameters vary, layer 2 has a much lower porosity/permeability values, and layer 4 has intermediate values.

| Layer | Principle angle | Longest range (Grid) | Shortest range (Grid) |
|-------|-----------------|----------------------|-----------------------|
| 1     | -60             | 19.45                | 5.56                  |
| 2     | -60             | 4.17                 | 4.17                  |
| 3     | -45             | 33.33                | 8.33                  |
| 4     | 60              | 8.33                 | 4.17                  |
| 5     | -30             | 20.8                 | 6.94                  |

**Table 4.2:** Parameters of spherical variogram models used to generate initial realizations for the five layers of the reservoir.

|           | Layer 1   | Layer 2   | Layer 3   | Layer 4   | Layer 5   |
|-----------|-----------|-----------|-----------|-----------|-----------|
| $\phi$    | 0.15/0.08 | 0.08/0.04 | 0.15/0.08 | 0.11/0.05 | 0.17/0.08 |
| $\ln K_h$ | 4.04/2.56 | 2.64/2.03 | 4.04/2.51 | 2.65/2.67 | 3.95/2.67 |
| $\ln K_v$ | 3.28/2.32 | 1.56/1.70 | 3.32/2.38 | 1.73/2.21 | 3.20/2.48 |

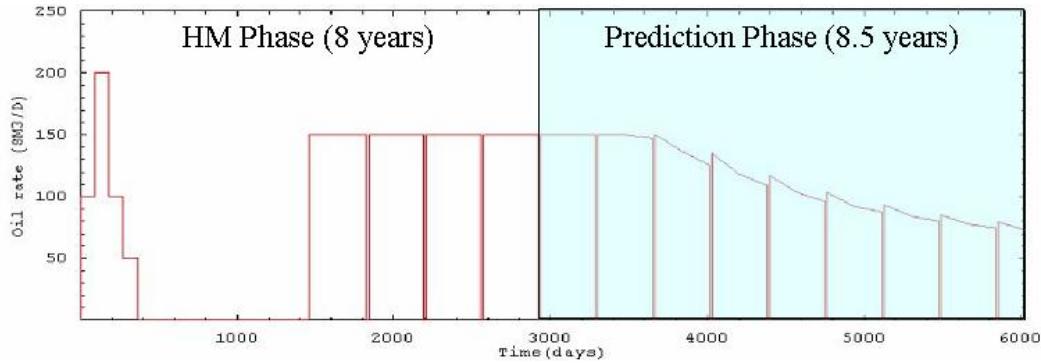
**Table 4.3:** Mean and standard deviation values of the initial realizations for different layers (all cells have same format: mean/std)

### 4.3 Production Data

The revealed true reservoir simulation model at the PUNQ web page was taken and run on the **ECLIPSE** black oil reservoir simulator to provide the true production data of 16.5 years. In order to check the validity of the Kalman corrections, only the production data of the first 8 years are used to correct the initial models and the final corrected models are used to predict the reservoir performance for the next 8.5 years. The prediction of the total oil production at the end of 16.5 years is compared to the results obtained by Floris et al. (2001).

Before the data for assimilation are introduced, let's first provide the information about how the wells are produced. All the six production wells are produced according to the same schedule: (1) an extended well testing period during the first year, and (2) a shut-in period for the next following 3 years, then (3) a 12-year production period. The one-year well-testing period consists of 4 time windows with 4 variable flow rates, each of which lasts 3 months long. Within the 12-year production period,

the oil production rate is fixed at  $150 \text{ sm}^3/\text{day}$ , and wells are shut-in for two weeks at the end of each year. All of the wells are primarily controlled by oil production rate; if the bottom-hole pressure falls below 120 bar ( $1 \text{ bar} = 100,000 \text{ pascal} \approx 14.5 \text{ psi}$ ), the bottom-hole pressure limit will be applied as constraint; also, if the gas-oil ratio is greater than  $200 \text{ sm}^3/\text{sm}^3$ , the oil production rate will be cut back by a factor of 0.75. Fig. 4.2 shows the oil production rate of the true reservoir simulation model versus time for one of the wells. This well maintains the specified rates until late in the prediction times.

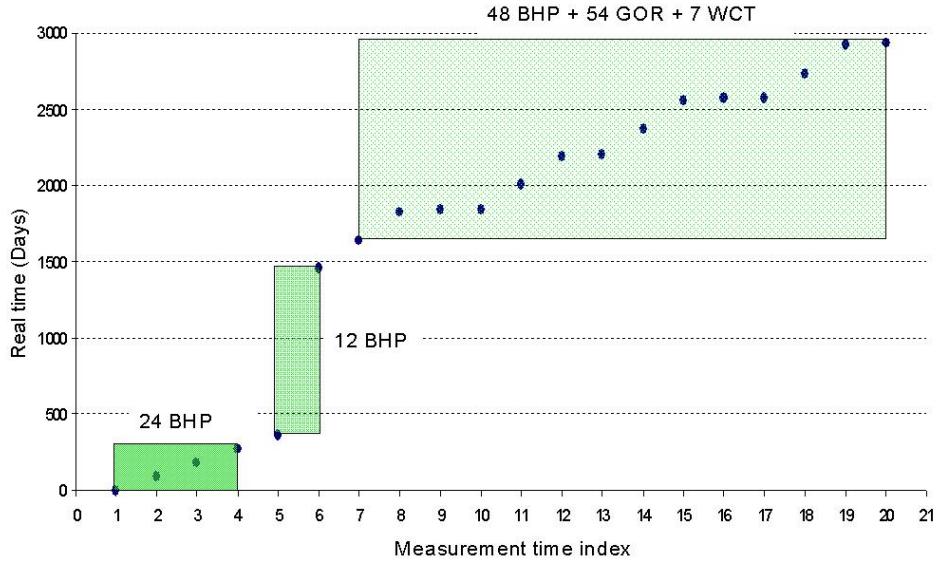


**Figure 4.2:** Oil production rate of the true reservoir simulation model versus time for one of the production wells.

Four kinds of data are used for data assimilation: well bottom-hole pressure (WBHP), well gas-oil ratio (WGOR), well water cut (WWCT) and well oil production rate (WOPR). Although the target oil production rate is identical for all the reservoir models, the actual oil rate values vary because wells in some simulation models are unable to meet the target. By using the oil production rate as additional data together with other measured data, we can adjust the simulation models under different production constraints and bring the failed models back up to the target oil production rate through the adjustment.

The kinds and numbers of data available at different times vary and are listed in Table 4.4. Shut-in pressures are marked in parentheses. Unmarked pressures are flowing pressures. There are totally 20 times in the first 8 years where data are

measured. The total amount of data for different kinds is: 84 for WBHP (36 shut-in and 48 flowing), 54 for WGOR, 7 for WWCT, and 120 WOPR. Fig. 4.3 is a summary plot of how often data are assimilated and what and how many data are available at the well testing, shut-in, and production periods of the 8 years. The  $x$ -axis is measurement time index and it is the same with the first column in Table 4.4. The  $y$ -axis is the real time for each measurement time index and corresponds to the second column in Table 4.4. From the vertical distance of two consecutive points, we can see that how often data are assimilated. Fig. 4.4 plots the time interval between two consecutive measurement times. The shortest measurement time interval is one day at the 10th (day 1841) and 17th (day 2572) measurement times. The longest interval is three years at the 6th (day 1461) measurement time. The rest of the intervals are around 2 weeks, three months, 5.5 months, and 6 months.

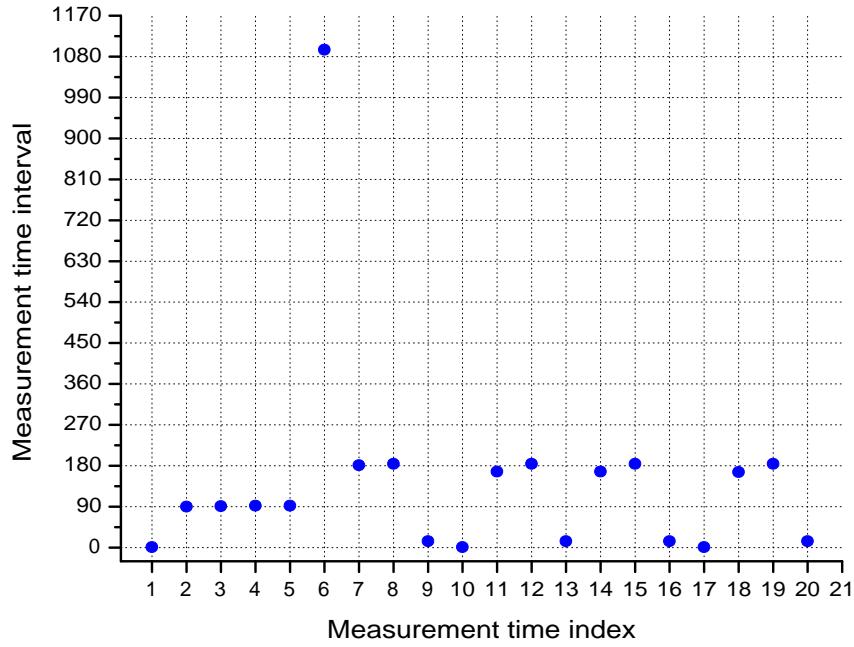


**Figure 4.3:** Summary plot of how often data are assimilated and what and how many data are available at the well testing, shut-in, and production periods of the first 8 years.

All the data used are corrupted with noise. The noise is assumed to be Gaussian distributed with mean 0. The standard deviations are (Floris et al., 2001) shown in Table 4.5.

| Measurement time index | Time (Days) | No. of WBHP | No. of WGOR | No. of WWCT | No. of WOPR |
|------------------------|-------------|-------------|-------------|-------------|-------------|
| 1                      | 1.01        | 6           | -           | -           | 6           |
| 2                      | 91          | 6           | -           | -           | 6           |
| 3                      | 182         | 6           | -           | -           | 6           |
| 4                      | 274         | 6           | -           | -           | 6           |
| 5                      | 366         | 6 (shut-in) | -           | -           | 6           |
| 6                      | 1461        | 6 (shut-in) | -           | -           | 6           |
| 7                      | 1642        | -           | 6           | -           | 6           |
| 8                      | 1826        | 6           | 6           | -           | 6           |
| 9                      | 1840        | 6 (shut-in) | -           | -           | 6           |
| 10                     | 1841        | -           | 6           | -           | 6           |
| 11                     | 2008        | -           | 6           | -           | 6           |
| 12                     | 2192        | 6           | 6           | -           | 6           |
| 13                     | 2206        | 6 (shut-in) | -           | -           | 6           |
| 14                     | 2373        | -           | 6           | -           | 6           |
| 15                     | 2557        | 6           | 6           | -           | 6           |
| 16                     | 2571        | 6 (shut-in) | -           | -           | 6           |
| 17                     | 2572        | -           | -           | 1           | 6           |
| 18                     | 2738        | -           | 6           | -           | 6           |
| 19                     | 2922        | 6           | 6           | 6           | 6           |
| 20                     | 2936        | 6 (shut-in) | -           | -           | 6           |
| Total                  | -           | 84          | 54          | 7           | 120         |

**Table 4.4:** The kinds and numbers of data available at different measurement times.



**Figure 4.4:** Measurement time intervals of the 20 measurement times.

During the history matching period of the first 8 years ( $0 \sim 2,936$  days), PRO-1 and PRO-4 have gas breakthrough at day 1826; PRO-11 has water breakthrough at day 2572; none of the other wells have neither gas nor water breakthrough.

| Data kind                     | STD of noise                      |
|-------------------------------|-----------------------------------|
| Shut-in pressure              | 1 bar                             |
| Flowing pressure              | 3 bar                             |
| Gas-oil ratio before gas b.t. | 10%                               |
| Gas-oil ratio after gas b.t.  | 25%                               |
| Water cut                     | 1%                                |
| Oil production rate           | $10^{-4} \text{ sm}^3/\text{day}$ |

**Table 4.5:** Standard deviations of the noises added to the synthetic data computed from the true reservoir simulation model.

The state vectors are in the form of

$$\begin{aligned}
y &= [\phi_1 \dots \phi_N \ \ln K_{h,1} \dots \ln K_{h,N} \ \ln K_{v,1} \dots \ln K_{v,N} \\
&\quad P_1 \dots P_N \ S_{w,1} \dots S_{w,N} \ S_{g,1} \dots S_{g,N} \ d_1 \dots d_{N_d} \ q_o]^T. \tag{4.1}
\end{aligned}$$

where  $N$  is the number of active cells,  $N = 1,761$ ;  $q_o$  is the oil production rate.

In regions for which the oil is under-saturated,  $S_g$  is not a valid state variable and it would be preferable to use  $R_s$  in the state vector as described by Nævdal et al. (2005). However, for the **PUNQ-S3** model, the magnitude of under-saturation is small and the use of  $S_g$  as a state variable did not result in significant problems.

In this application, the initial pressure and phase saturation distributions are not treated as random variables. All the ensemble models use the same prescribed distributions of initial pressure and saturations.

## 4.4 Production Data Matched

The comparison of well performance from the initial models and during history-matching (HM) process and prediction for all the six wells are shown from Fig. 4.5 to Fig. 4.10. On each plot, the black vertical line divides the time axis into two phases: the HM phase (0 to 2,936 days) and prediction phase (2,937 to 6,025 days); the red line is the simulated data from the true reservoir simulation model; the multiple black lines are the results from the ensemble models; the blue crosses represent the data used for assimilation.

Comparing the production data of all kinds at each of the six well from the initial models to their counterparts from the HM process and prediction, we can see that the former demonstrates greater variation in performance than the latter. Let us examine the comparisons by data kind in the following:

**WOPR:** Fig. 4.5(a) shows that some of the initial models fail to produce at the specified well production target rate while still in the well-testing period. After the correction, however, most of the models meet the target (Fig. 4.5(b)). Even in the prediction phase, most of the corrected models can produce at the desired rate and only a few uses the 120 bar bottom-hole pressure constraint (Fig. 4.5(d)). For all others wells, the models that have lower oil production than the true simulation

model are adjusted and are able to produce at the desired rate after the adjustment (compare Fig. 4.8(a) to 4.8(b), Fig. 4.9(a) to 4.9(b), and Fig. 4.10(a) to 4.10(b)).

**WBHP:** Fig. 4.5(d) shows that the true bottom-hole pressures are honored well by the corrected models in the HM phase, and the uncertainty in the prediction phase is reduced significantly comparing to that from the initial models. The same holds true for all the other wells except for PRO-12 where the prediction go astray during the later prediction phase, see Fig. 4.9(d).

**WGOR:** The comparison of Figs. 4.5(e) and 4.5(f) shows substantial improvements of the gas-oil ratio match. Some of the initial models produce much more free gas than the true model does, while the corrected models have their gas oil ratio distributed closely to the truth. Note during the first year of well testing period, some of the initial models have gas-oil ratio as high as  $400 \text{ sm}^3/\text{sm}^3$ , even without any gas-oil ratio data assimilated during that time, the adjustment is able to bring the high values down to  $150 \text{ sm}^3/\text{sm}^3$ . At well PRO-11, the free gas production is corrected during the history matching phase, however, shortly after the data assimilation is ended, the gas-oil ratio goes up again in the prediction phase, see Fig. 4.8(f). This might be because the solution gas-oil ratio data at this well do not provide much information.

**WWCT:** Well PRO-1 does not see water breakthrough for all the 16.5 years, while some of the initial and corrected models predict water breakthrough within the prediction period. However, the water breakthrough time predicted by the corrected models is further delayed than that by the initial models, compare Fig. 4.5(h) to Fig. 4.5(g). Let us look at the water cut at well PRO-11 where water breakthrough during the HM time period. The initial models do not give correct timing for the water breakthrough (Fig. 4.8(g)). After data assimilation, the timing is captured (Fig. 4.8(h)). However, the corrected model does not provide a good prediction for water cut at well PRO-12, see Fig. 4.9(h).

Data at all six wells are matched fairly well. The predictions based on the corrected models are generally better than those from the initial models. Only at well PRO-12, the predictions in both the bottom-hole pressure and water cut from the corrected models are not better than these from the initial models, see Figs. 4.9(d) and 4.9(h).

## 4.5 Prediction of Total Oil Production

We have seen the variation of the predicted reservoir performance in the corrected models. It can be used to assess uncertainty in the quantities of interest.

For this case, we are interested in comparing the variability of the cumulative oil production after 16.5 years from the ensemble of final corrected models using the EnKF to the results from other methods summarized by Floris et al. (2001). Fig. 4.11 shows such a comparison. In the plot, the  $x$ -axis is for different methods and the EnKF is shown on the right-most. The  $y$ -axis is the total oil production at the end of year 16.5. The horizontal white line crossing the entire plot is the values computed from the true reservoir simulation model. The bar plot for each method specifies values of  $q_{90}$  (upper bar),  $q_{50}$  (central dot), and  $q_{10}$  (lower bar). It can be seen that the  $q_{50}$  from the EnKF is very close to the truth, and the range between the  $q_{10}$  and  $q_{90}$  reflects an reasonable distribution of the ensemble of predictions in the oil production.

Fig. 4.12 plots the CDF of the cumulative oil production after 16.5 years from the corrected models along with the Gaussian CDF with mean equal to the true production value and STD of  $71,845 \text{ sm}^3$ . It shows the closeness of the empirical CDF to the Gaussian CDF.

## 4.6 Porosity and Permeability Estimates

Figs. 4.13 – 4.17 plot the true porosity field, ensemble mean, gridblock RMSE and STD of the porosity estimates of the five layers at the initial time 0, the 4th (274

days), 11th (2008 days), and 20th (2936 days) measurement times. The solid black dots denote the well locations. From these figures, we have some general observations: (1) the initial mean porosity fields are quite smooth. With data assimilation, the features of the true porosity field are able to be recovered gradually, for example, the alternating low and high porosity streaks in layers 1, 3, and 5, (2) because the hard data are used when generating initial realizations, we can see throughout the time, the gridblock RMSE and STD at wells have the smallest value, (3) even though the mean porosity estimates seem to capture the major characteristics of the true fields, the RMSEs show that the error in the porosity estimates grows larger after some time, especially at the regions lacking of constraints, e.g. regions far from wells, and (4) the gridblock STDs decrease firstly around the well, then with more data are assimilated, the STD of the entire field decreases steadily.

Fig. 4.18 plots the scalar RMSEs and STDs of the porosity,  $\ln K_h$ , and  $\ln K_v$  estimates. The RMSEs and STDs for all the three estimates behave quite similarly. As what has been observed from the gridblock RMSE in Figs. 4.13 – 4.17, the RMSEs of all the estimates decrease firstly with time, but jump suddenly after the data assimilation at the 18th measurement time, day 2738, while the STDs of the three estimates always decrease along the time. Fortunately, for this case, the error is reduced with the two more data assimilation. However, the level of the reduction is less than the level of the sudden rise.

Potential problems of the EnKF are indicated when the error in the estimates begins to grow dramatically while the spread of estimates goes down. The spread of the ensemble is not able to capture the variability which makes it harder to lower the error in the estimates to a reasonable level even with more data assimilation.

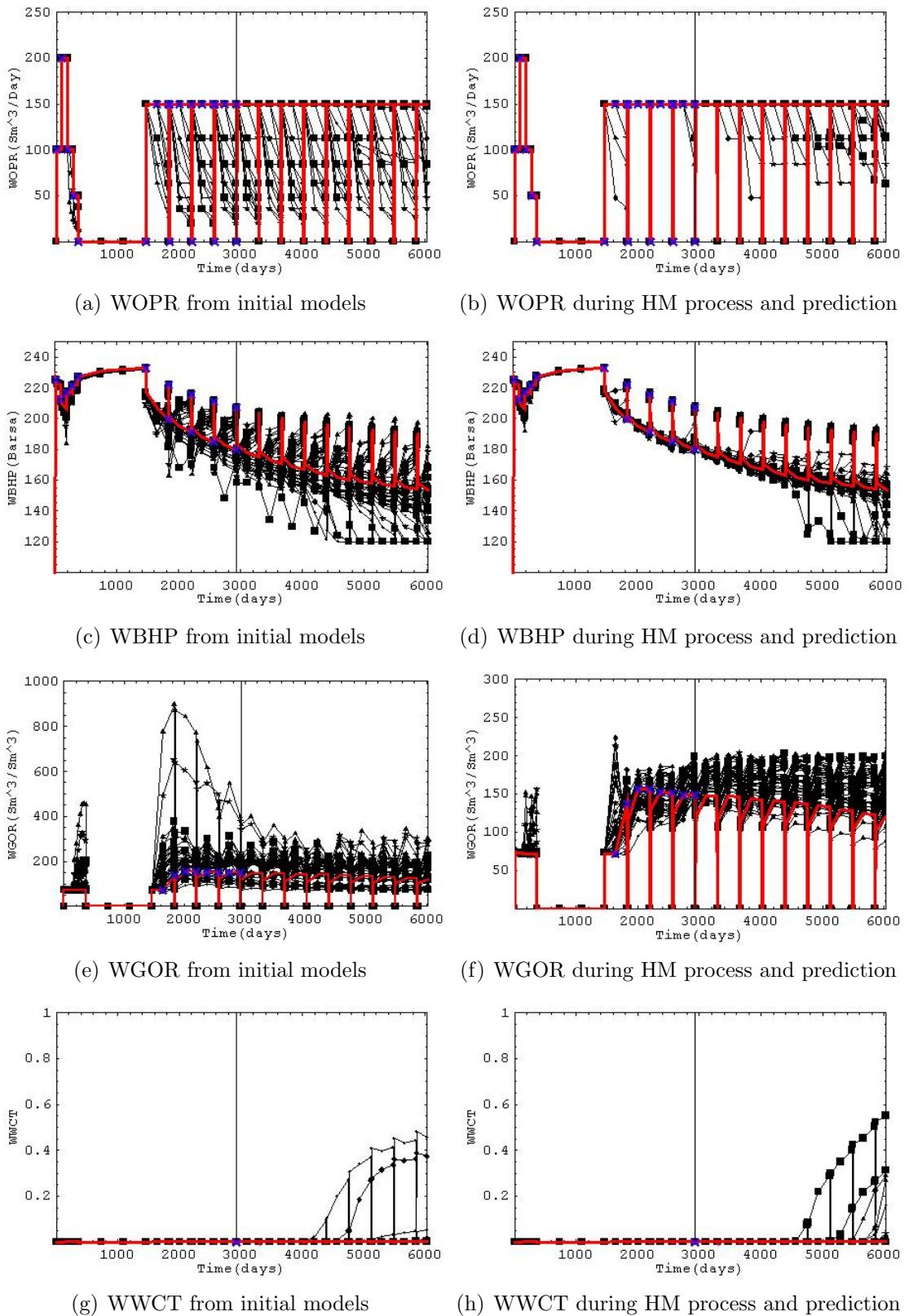
## 4.7 Discussion

We have successfully demonstrated the application of the EnKF to a more realistic history matching problem, also as a method to quantify uncertainty in predicted reservoir performance. The following summarizes the study:

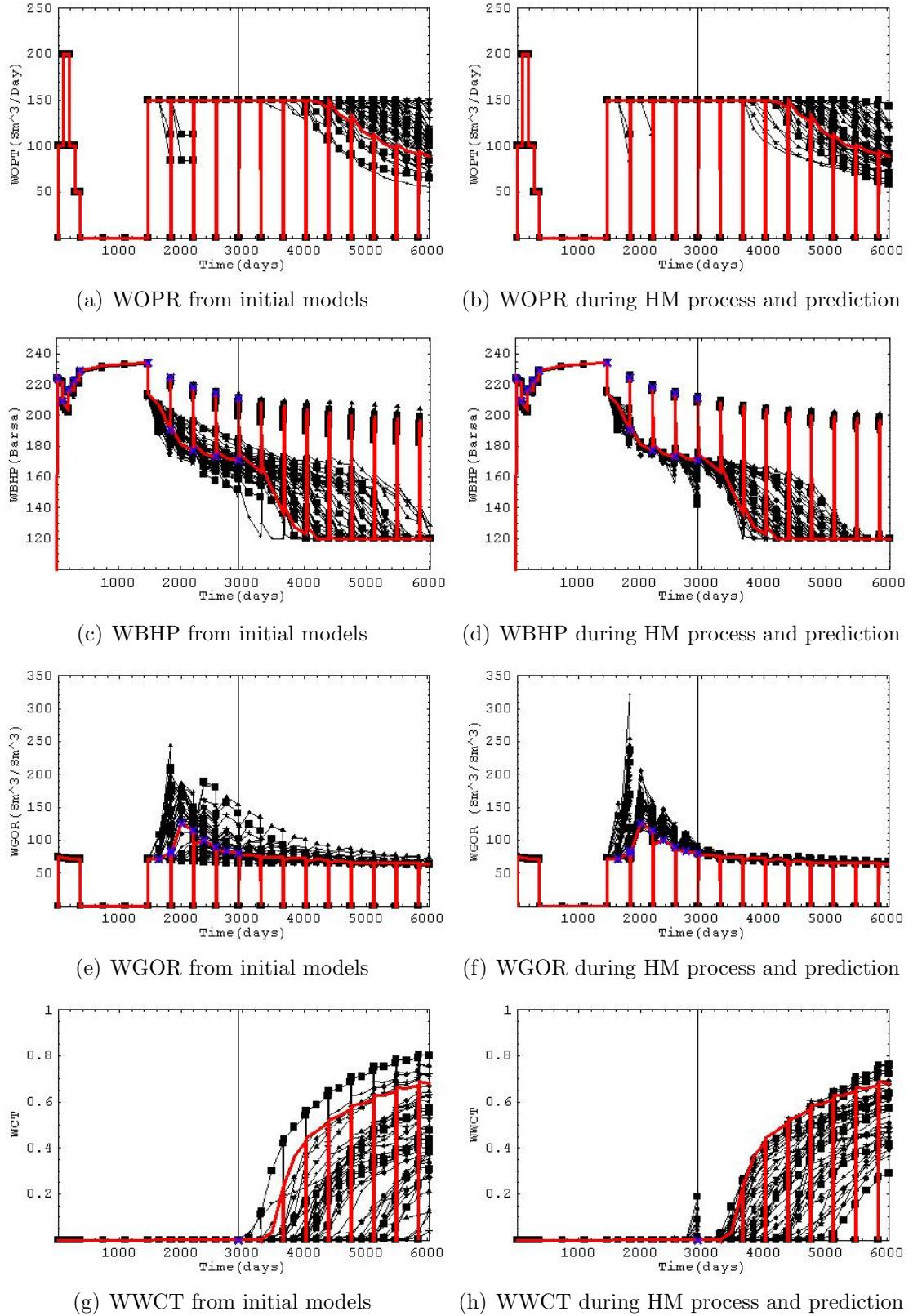
A fairly small ensemble (40) provides a quite good match of production history at all the six wells. The computational cost for generating 40 “history-matched” models is approximately 40 simulation runs plus the matrix computations in the update steps. It is clearly very efficient compared to other history-matching methods. Certainly, parallelism can further reduce the computation time.

The prediction of the total oil production at the end of 16.5 years from the ensemble of corrected models is in reasonable agreement with the truth and comparable to the results obtained by other methods. The variability provides a way to estimate the uncertainty.

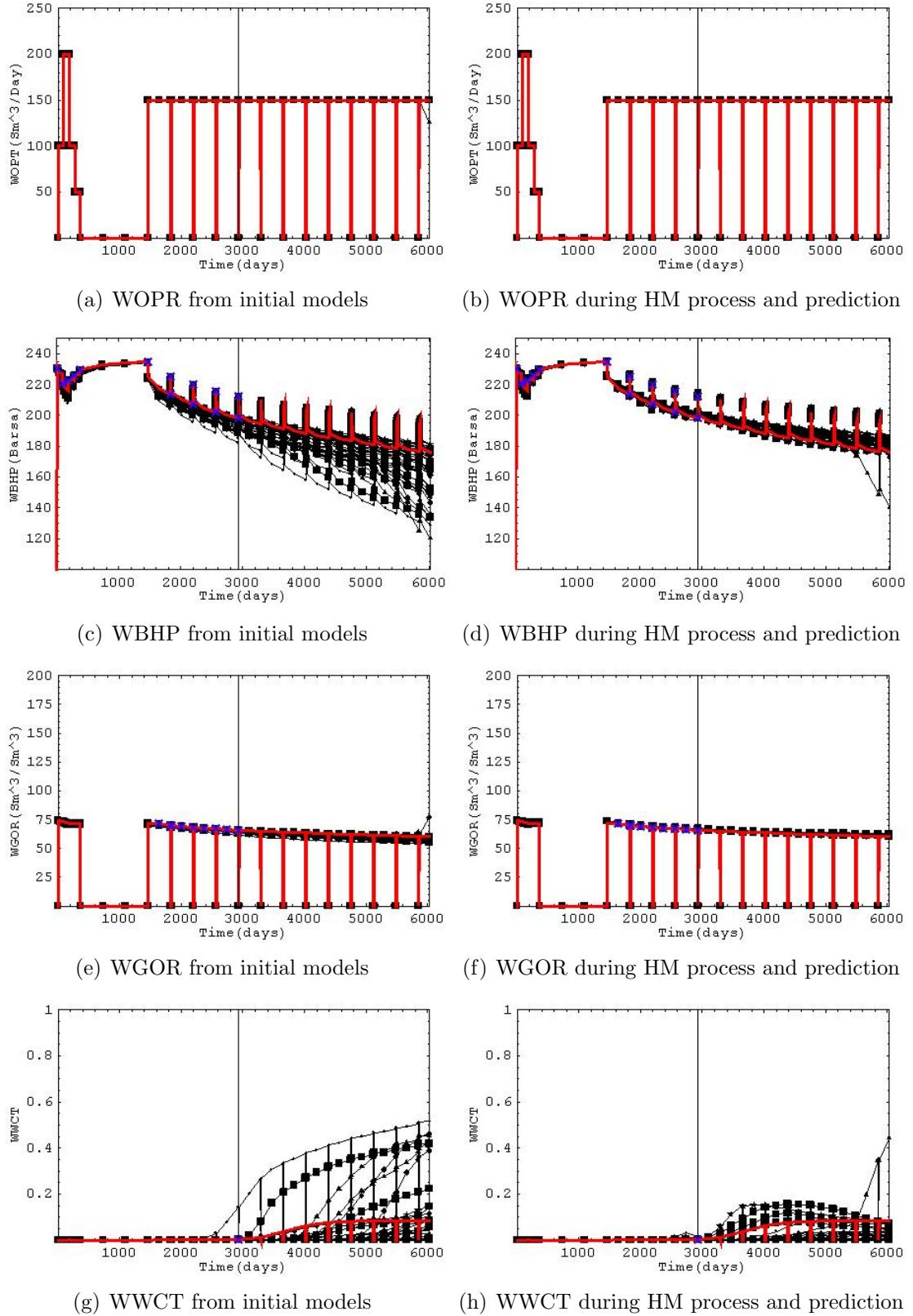
Problems with increasing error from the truth in estimates of porosity and permeability fields are apparent, which are indications of the potential problems with the EnKF.



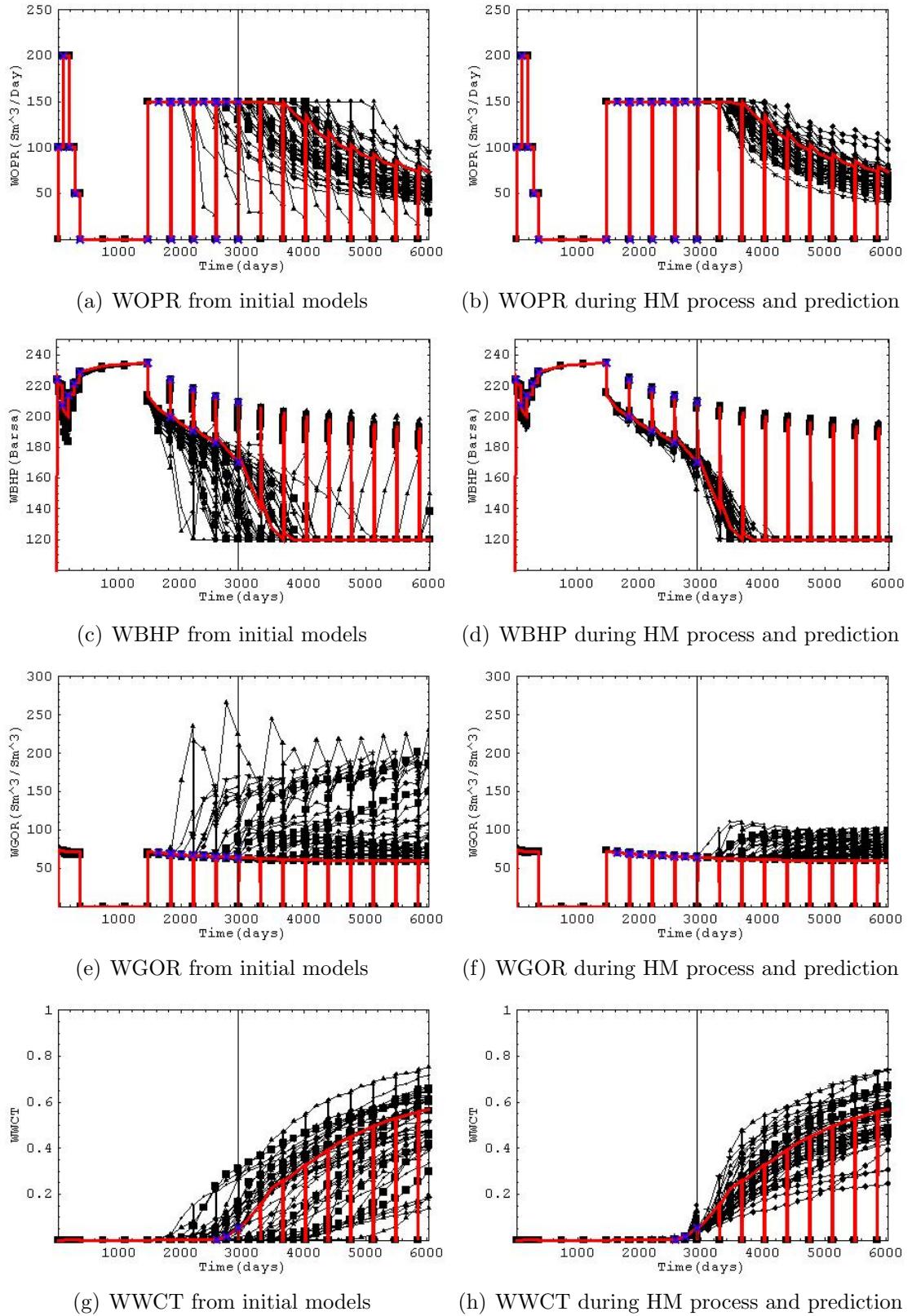
**Figure 4.5:** The production data at PRO-1 from the initial models and during the HM process & prediction. The vertical black line divides the time axis into history matching phase and prediction phase. The red line denotes the simulated data from the true simulation model. The black lines denote the results from different ensemble members. The blue crosses represent the data used.



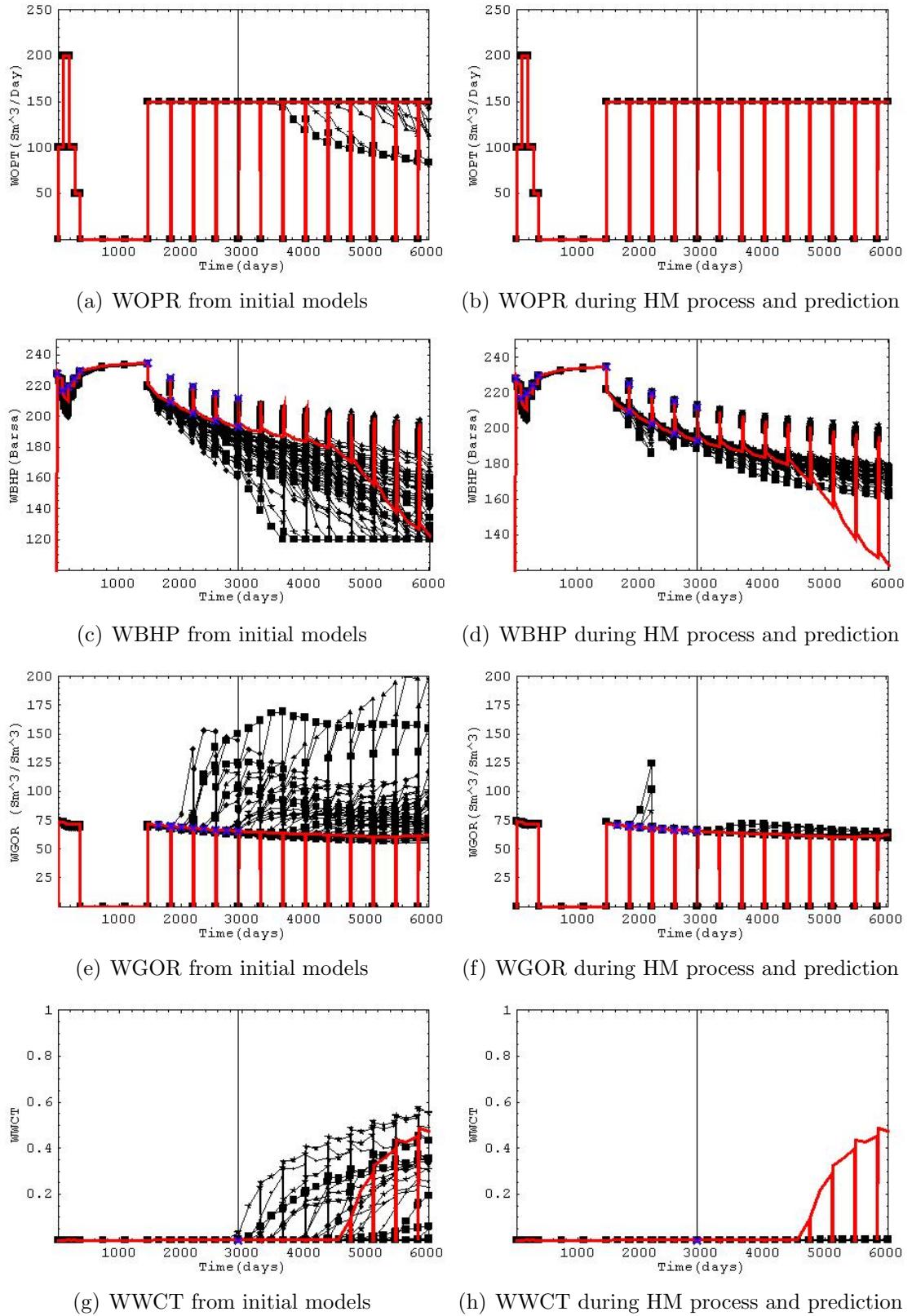
**Figure 4.6:** The production data at PRO-4 from the initial models and during the the HM process & prediction. The legends used are the same with Fig. 4.5.



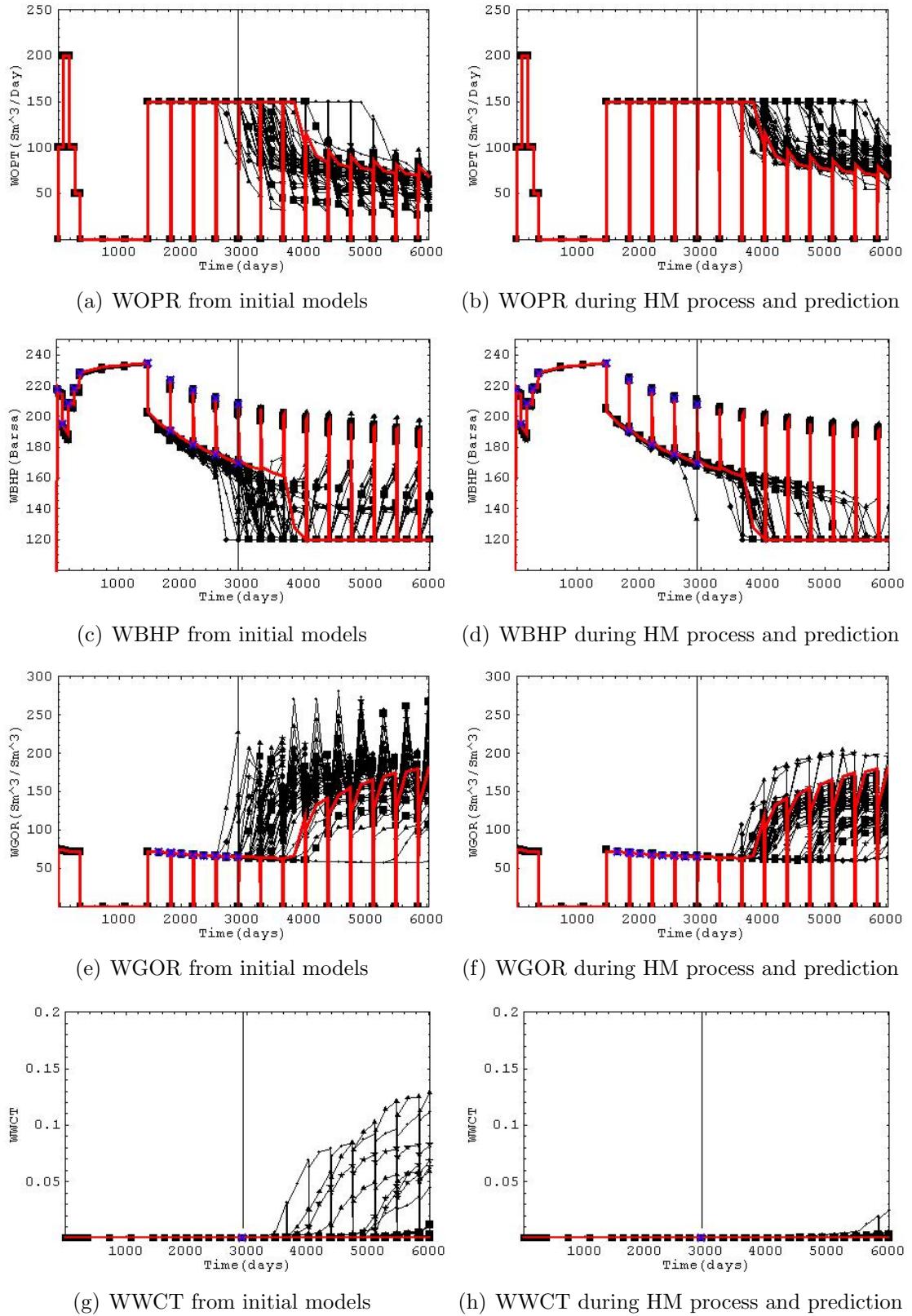
**Figure 4.7:** The production data at PRO-5 from the initial models and during the the HM process & prediction. The legends used are the same with Fig. 4.5.



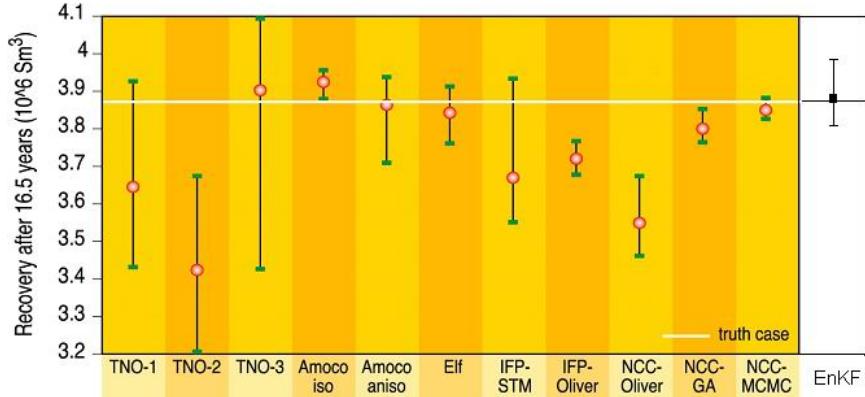
**Figure 4.8:** The production data at PRO-11 from the initial models and during the the HM process & prediction. The legends used are the same with Fig. 4.5.



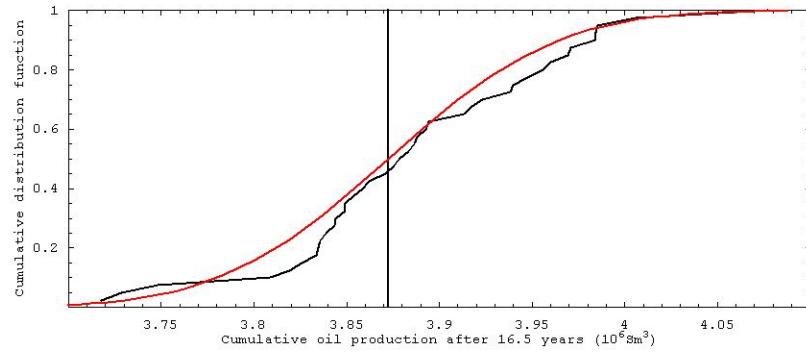
**Figure 4.9:** The production data at PRO-12 from the initial models and during the the HM process & prediction. The legends used are the same with Fig. 4.5.



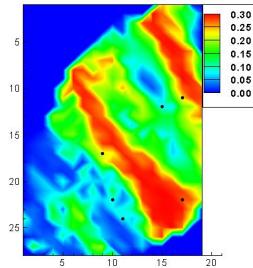
**Figure 4.10:** The production data at PRO-15 from the initial models and during the the HM process & prediction. The legends used are the same with Fig. 4.5.



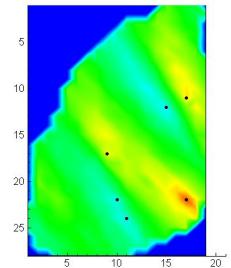
**Figure 4.11:** Quantiles of 10, 50, and 90, of the cumulative oil production after 16.5 years from the EnKF, and results summarized by Floris et al. (2001). The horizontal line crossing the entire plot denotes the computed total oil production value from the true reservoir simulation model.



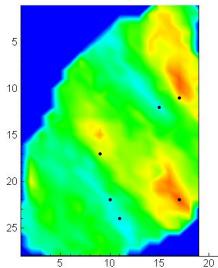
**Figure 4.12:** Cumulative distribution function (CDF) of the cumulative oil production after 16.5 years from the EnKF. The vertical black line stands for the true total oil production value. The smooth red curve is the referenced Gaussian CDF with mean equal to the true production value and STD of  $71,845 \text{ sm}^3$ .



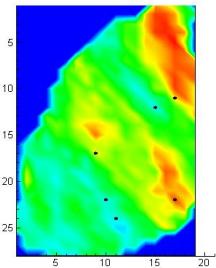
(a) Truth.



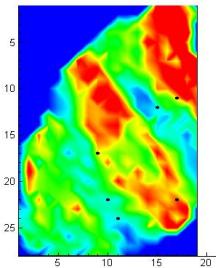
(b) Initial.



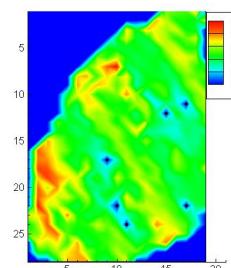
(c) Estimate at day 274.



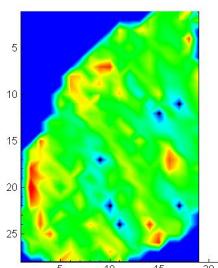
(d) Estimate at day 2008.



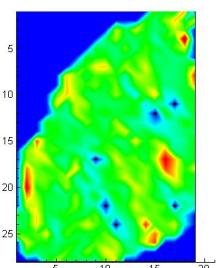
(e) Estimate at day 2936.



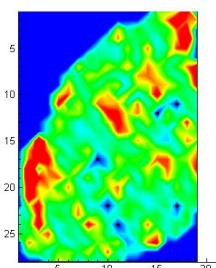
(f) Initial RMSE.



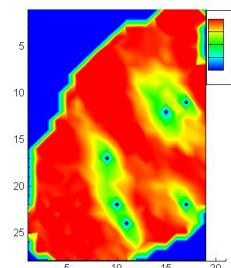
(g) RMSE at day 274.



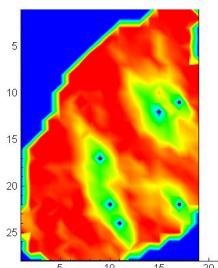
(h) RMSE at day 2008.



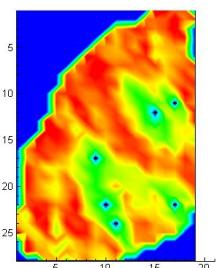
(i) RMSE at day 2936.



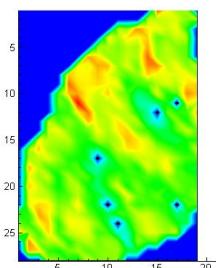
(j) initial STD.



(k) STD at day 274.

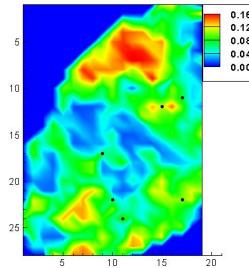


(l) STD at day 2008.

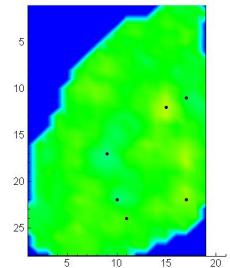


(m) STD at day 2936.

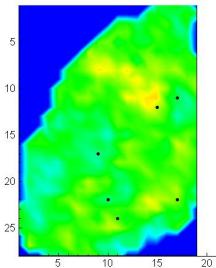
**Figure 4.13:** True porosity, evolution of the mean, gridblock RMSE, and STD of porosity estimates in layer 1.



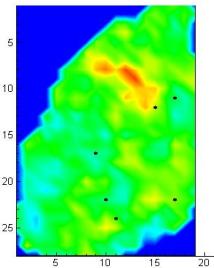
(a) Truth.



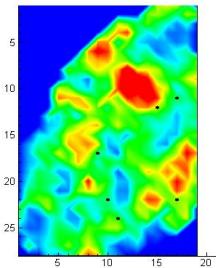
(b) Initial.



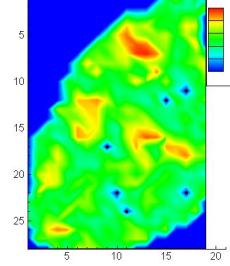
(c) Estimate at day 274.



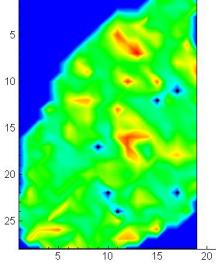
(d) Estimate at day 2008.



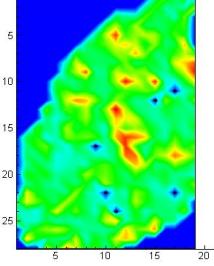
(e) Estimate at day 2936.



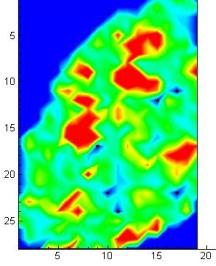
(f) Initial RMSE.



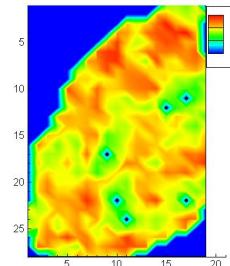
(g) RMSE at day 274.



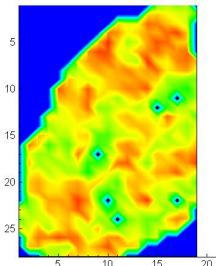
(h) RMSE at day 2008.



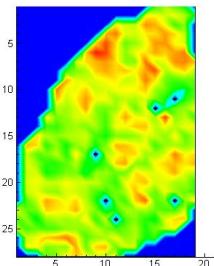
(i) RMSE at day 2936.



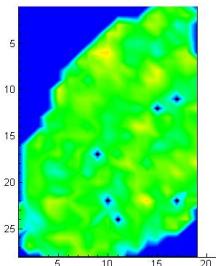
(j) initial STD.



(k) STD at day 274.

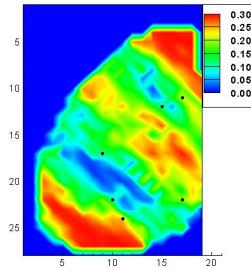


(l) STD at day 2008.

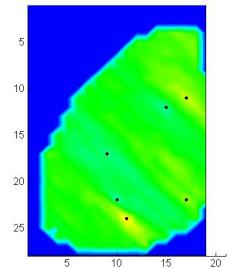


(m) STD at day 2936.

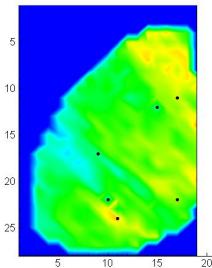
**Figure 4.14:** True porosity, evolution of the mean, gridblock RMSE, and STD of porosity estimates in layer 2.



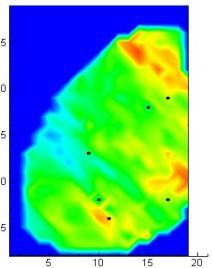
(a) Truth.



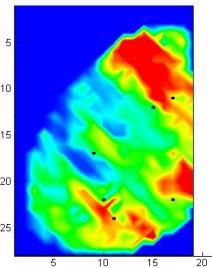
(b) Initial.



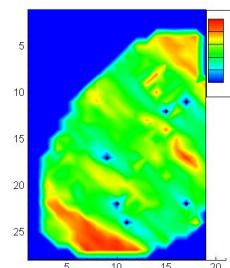
(c) Estimate at day 274.



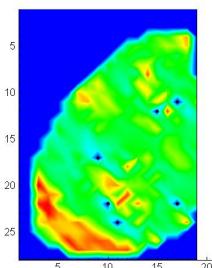
(d) Estimate at day 2008.



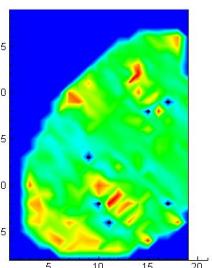
(e) Estimate at day 2936.



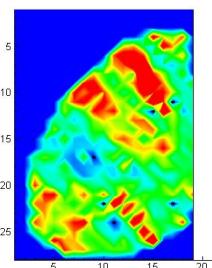
(f) Initial RMSE.



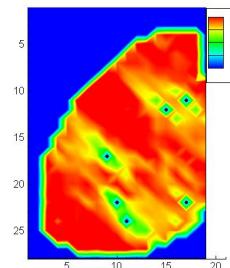
(g) RMSE at day 274.



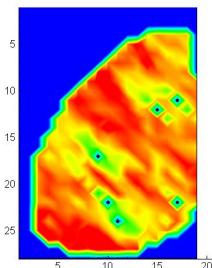
(h) RMSE at day 2008.



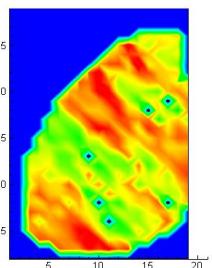
(i) RMSE at day 2936.



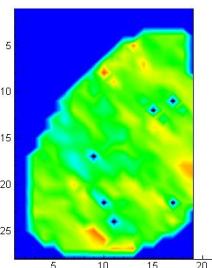
(j) initial STD.



(k) STD at day 274.

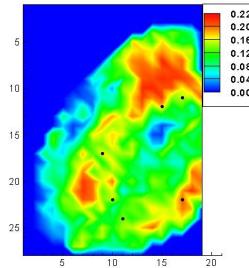


(l) STD at day 2008.

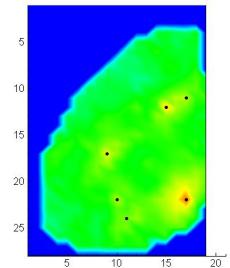


(m) STD at day 2936.

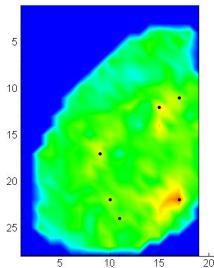
**Figure 4.15:** True porosity, evolution of the mean, gridblock RMSE, and STD of porosity estimates in layer 3.



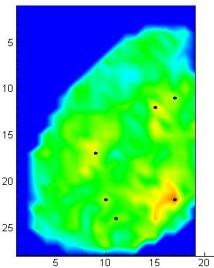
(a) Truth.



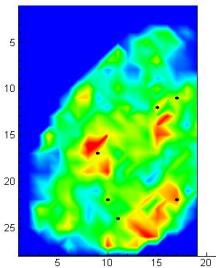
(b) Initial.



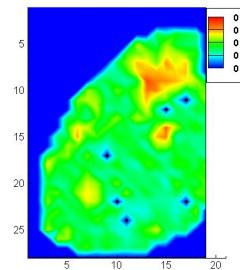
(c) Estimate at day 274.



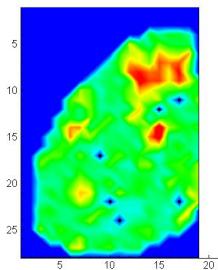
(d) Estimate at day 2008.



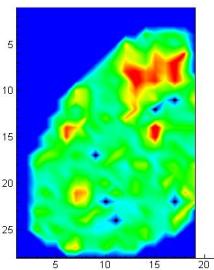
(e) Estimate at day 2936.



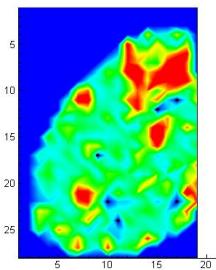
(f) Initial RMSE.



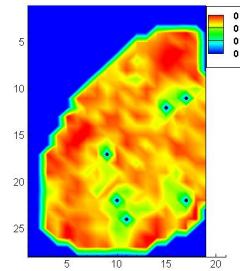
(g) RMSE at day 274.



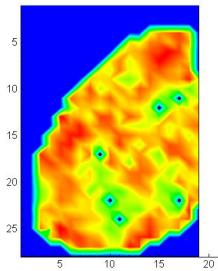
(h) RMSE at day 2008.



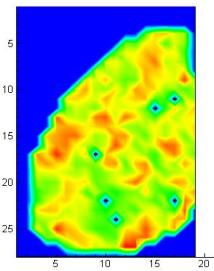
(i) RMSE at day 2936.



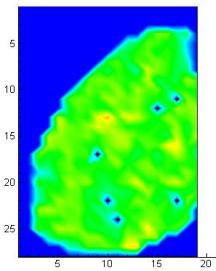
(j) initial STD.



(k) STD at day 274.

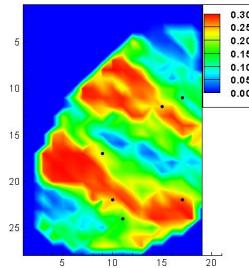


(l) STD at day 2008.

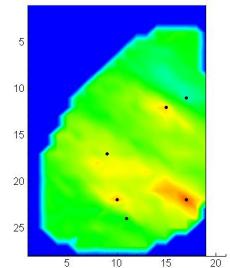


(m) STD at day 2936.

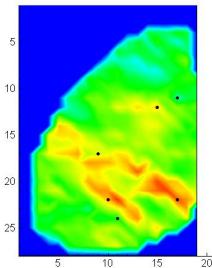
**Figure 4.16:** True porosity, evolution of the mean, gridblock RMSE, and STD of porosity estimates in layer 4.



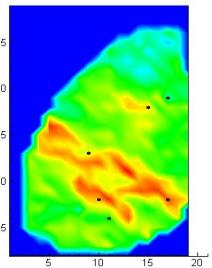
(a) Truth.



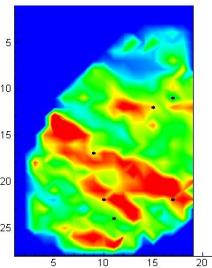
(b) Initial.



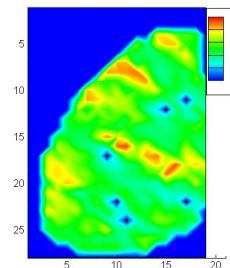
(c) Estimate at day 274.



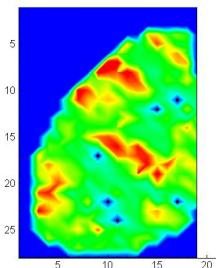
(d) Estimate at day 2008.



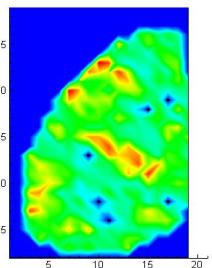
(e) Estimate at day 2936.



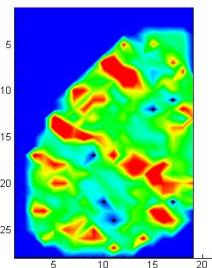
(f) Initial RMSE.



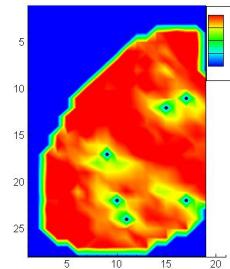
(g) RMSE at day 274.



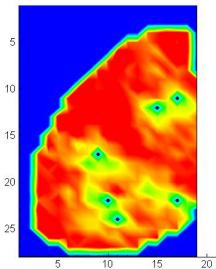
(h) RMSE at day 2008.



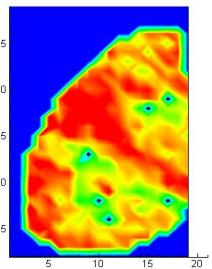
(i) RMSE at day 2936.



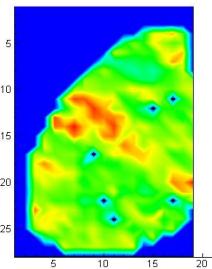
(j) initial STD.



(k) STD at day 274.

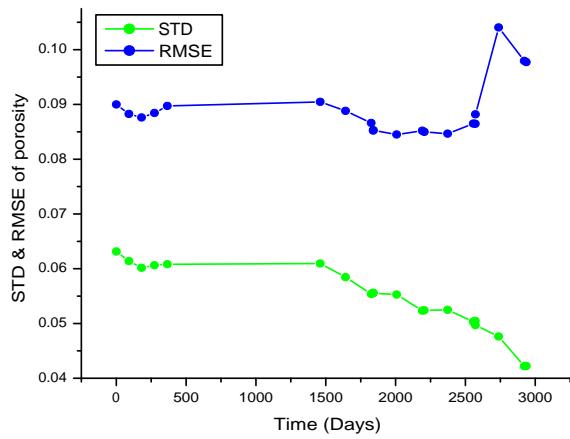


(l) STD at day 2008.

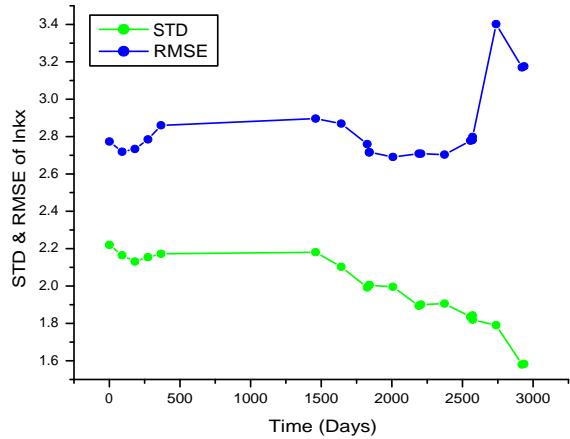


(m) STD at day 2936.

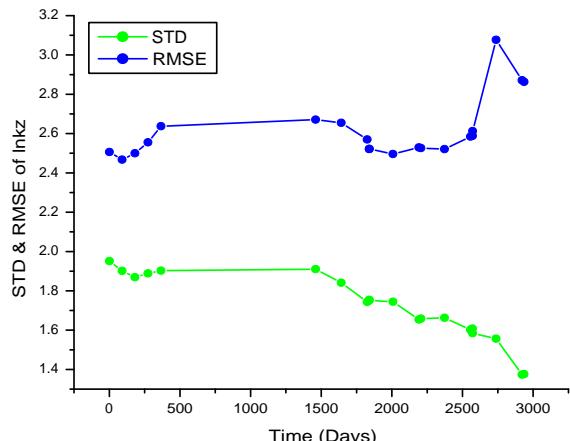
**Figure 4.17:** True porosity, evolution of the mean, gridblock RMSE, and STD of porosity estimates in layer 5.



(a) For porosity



(b) For  $\ln K_h$



(c) For  $\ln K_v$

**Figure 4.18:** RMSEs and STDs of for the porosity,  $\ln K_h$ , and  $\ln K_v$  estimates.

# CHAPTER V

## INVESTIGATION OF ITERATIVE SCHEMES HANDLING NON-LINEARITY BASED ON THE ENSEMBLE KALMAN FILTER

### 5.1 Possible Problems of the EnKF

In traditional history-matching approaches, the collection of variables to be estimated consists only of the model parameters (e.g. porosities and permeabilities). The state variables (e.g. pressures and saturations) are usually determined from the knowledge of the model parameters by solving the system governing equations. However, in the EnKF, model parameters and state variables are both updated at the update steps whenever measurements are available. In essence a linearized approximation to the reservoir fluid flow equations is used to make a prediction of the changes in the state variables that would result from the changes in the model parameters. When the changes in the state variables are small, the linearized approximation to the non-linear equations is acceptably reliable. However, if the changes are big, it may be impossible to update the state variables to be consistent with the updated model parameters without re-solving the non-linear forward equations. This is one potential problem with updating both the model parameters and state variables simultaneously in the EnKF. The update equation in the EnKF is based on Gaussian error statistics. The update of state variables whose density functions are bi-modal with the EnKF has been shown as problematic (Gu and Oliver, 2006). In reservoir applications, such state variables are water saturations at some circumstances. For a simple water flooding problem, the water saturations take large values behind the water flood front,

and small values ahead of the front. Its distribution is bi-modal in this case and is not well modeled by the mean and variance, see the one-dimensional waterflood problem in Chapter 3. This is the second potential problem with the EnKF.

### 5.1.1 Wen and Chen’s remedy

The two problems of the EnKF are interwoven with each other and both are resulted from the inclusion of the state variables in the state vectors. Intuitively, to impose proper constraints on the state variables, one can update the model parameters only with the EnKF update equation. Once the model parameters are corrected, the state variables can be computed by solving the reservoir flow governing equations. Wen and Chen (2005a) suggested to add a so-called “conforming step” to enforce physical constraints on the state variables at each measurement time. So the process at one measurement time of their proposal is a three-step procedure: (1) a forecast step which propagates the state of the system from previous measurement time to current measurement time, which is the same with that in the EnKF (2) an update step which corrects only the model parameters with the EnKF update equation, and (3) a conforming step which re-initiates the system governing equations at the previous measurement time with the newly updated model parameters, conditional to data up to the current measurement time, and the state variables at the previous measurement time, then propagates the re-initialized equations to the current measurement time. It can be easily seen that their scheme doubles the computing time comparing to that of the EnKF since there are two simulation runs for each simulation model in the forecast and conforming steps. They also suggested to iterate **Steps 2** and **3** when non-linearity of problems is strong (Wen and Chen, 2005b). Throughout the dissertation, Wen and Chen’s method is referred as Conforming EnKF.

Writing the **update** and **conforming** steps in equations at the  $k$ th measurement

time for the  $j$ th member ( $j = 1, 2, \dots, N_e$ )

$$m_{j,k}^{\ell+1} = m_{j,k}^{\ell} + \alpha_{\ell} K_{e,k,\ell} (d_{\text{obs},j,k} - g(m_{j,k}^{\ell})) , \quad (5.1)$$

$$\begin{bmatrix} f(m_{j,k}^{\ell+1}) \\ g(m_{j,k}^{\ell+1}) \end{bmatrix} = \psi \left( m_{j,k}^{\ell+1}, f(m_{j,k-1}^f), t : t_{k-1} \rightarrow t_k \right) , \quad (5.2)$$

where  $\ell$  denotes the iteration number; for the first iteration, i.e.  $\ell = 1$ ,  $m_{j,k}^{\ell} = m_{j,k}^p = m_{j,k-1}^f$ ; the superscript  $f$  means that the values are the final solution with data assimilation at a particular time;  $\alpha_{\ell} \in [0, 1]$  is a damping factor at the  $\ell$ th iteration;  $\psi(\cdot)$  denotes the reservoir simulator, the first two arguments specified in the parentheses are model parameters and state variables used to re-initialize the flow equations, and the time argument on the left of the arrow is the time of the re-initialization and on the right is the current measurement time.

From Eqs. 5.1 and 5.2, we have the following observations: (1) the **conforming** step has mismatched re-initialization variables: the model parameters are conditional to data up to the current measurement time, while the state variables are conditioned to data up to the previous measurement time (2) the iterative updating equation is almost equivalent to applying Kalman corrections iteratively over one measurement time.

## 5.2 Ensemble Randomized Maximum Likelihood Filter

A new form of the iterative scheme, called Ensemble Randomized Maximum Likelihood Filter (EnRMLF), aiming to solve the possible problems with the EnKF is proposed. There are two major differences between the newly proposed scheme and the Conforming EnKF: (1) the update equation of model parameters is the iterative Gauss-Newton formula derived from computing the maximum a posteriori model (2) the state variables are computed by re-initiating the system governing equations at time 0 with the updated model parameters. For linear dynamic systems, this method

would give equivalent solutions as the EnKF does. For non-linear dynamic systems, this method would outperform the EnKF.

### 5.2.1 Linear dynamic system

If the state variables and theoretical data are linearly related to the model parameters, the state vector has the general form as written in Eq. 2.6,

$$y = \begin{bmatrix} m \\ Fm \\ Gm \end{bmatrix} .$$

Suppose that the auto-covariance of the model parameters  $m$  is represented by the matrix  $C_M$ , then the auto-covariance of the state vector  $y$  is

$$C_Y = \begin{bmatrix} C_M & C_M F^T & C_M G^T \\ F C_M & F C_M F^T & F C_M G^T \\ G C_M & G C_M F^T & G C_M G^T \end{bmatrix} . \quad (5.3)$$

The “best” estimate of the state vector  $y$  is obtained by minimizing the objective function composed of two terms: a data mismatch term and a model mismatch term

$$S(y) = \frac{1}{2}(Hy - d_{\text{obs}})^T C_D^{-1}(Hy - d_{\text{obs}}) + \frac{1}{2}(y - y^p)^T C_Y^{-1}(y - y^p) . \quad (5.4)$$

If  $C_Y$  were nonsingular, the minimum of the objective function would be achieved at

$$\langle y \rangle = y^p + C_Y H^T (H C_Y H^T + C_D)^{-1} (d_{\text{obs}} - H y^p) , \quad (5.5)$$

where  $\langle \cdot \rangle$  is used to denote the best estimate;  $y^p$  is the prior estimate of the state vector, it is the same with that used in Eq. 2.10; all the other terms are the same with those introduced before.

Expanding Eq. 5.5 in terms of  $C_M$ ,  $F$  and  $G$ , it would become

$$\begin{bmatrix} \langle m \rangle \\ F \langle m \rangle \\ G \langle m \rangle \end{bmatrix} = \begin{bmatrix} m^p \\ Fm^p \\ Gm^p \end{bmatrix} + \begin{bmatrix} C_M G^T \\ F C_M G^T \\ G C_M G^T \end{bmatrix} (G C_M G^T + C_D)^{-1} (d_{\text{obs}} - H y^p) , \quad (5.6)$$

because

$$C_Y H^T = \begin{bmatrix} C_M G^T \\ F C_M G^T \\ G C_M G^T \end{bmatrix}$$

and  $H C_Y H^T = G C_M G^T$ . From Eq. 5.6, it can be seen that when the state vector is updated, the adjustment of model parameters, state variables, and data are consistent. Eq. 5.6 can be simplified, since the three sets of equations are redundant and only the first set is needed.

$$\langle m \rangle = m^p + C_M G^T (G C_M G^T + C_D)^{-1} (d_{\text{obs}} - H y^p). \quad (5.7)$$

In this form it is identical to the result in Tarantola (1987, Eq. 1.93 on page 70).

### 5.2.2 Non-linear dynamic system

The general form of the state vector is also written in Eq. 2.6 for non-linear dynamic systems,

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

We can consider a series of linear approximations to the non-linear functions,  $f(m)$  and  $g(m)$ , by linearizing them at point  $m^\ell$ . Suppose that

$$f(m) \approx f(m^\ell) + F_\ell(m - m_\ell)$$

and

$$g(m) \approx g(m^\ell) + G_\ell(m - m_\ell),$$

so the approximation to the auto-covariance of the state vector  $y$  based on the linearization at  $m^\ell$  is

$$C_Y = \begin{bmatrix} C_M & C_M F_\ell^T & C_M G_\ell^T \\ F_\ell C_M & F_\ell C_M F_\ell^T & F_\ell C_M G_\ell^T \\ G_\ell C_M & G_\ell C_M F_\ell^T & G_\ell C_M G_\ell^T \end{bmatrix}. \quad (5.8)$$

In this case, it is necessary to explicitly note that the objective function to be minimized is (different from that in linear dynamic system (Eq. 5.4)

$$S(m) = \frac{1}{2} (g(m) - d_{\text{obs}})^T C_D^{-1} (g(m) - d_{\text{obs}}) + \frac{1}{2} (m - m^p)^T C_M^{-1} (m - m^p). \quad (5.9)$$

The  $\ell$ th iteration of the Gauss-Newton method for finding the model parameters  $m$  that minimizes the objective function is

$$\begin{aligned} m^{\ell+1} &= m^p \\ &\quad - C_M G_\ell^T (C_D + G_\ell C_M G_\ell^T)^{-1} [g(m^\ell) - d_{\text{obs}} - G_\ell(m^\ell - m^p)]. \end{aligned} \quad (5.10)$$

The Gauss-Newton formula for iteration with a shorter step-length is

$$\begin{aligned} m^{\ell+1} &= \beta_\ell m^p + (1 - \beta_\ell)m^\ell \\ &\quad - \beta_\ell C_M G_\ell^T (C_D + G_\ell C_M G_\ell^T)^{-1} [g(m^\ell) - d_{\text{obs}} - G_\ell(m^\ell - m^p)], \end{aligned} \quad (5.11)$$

where  $\beta_\ell$  is an adjustment to the step length and takes a value between 0 and 1; in the EnKF framework,  $m^p$  is the estimate for the model parameters after assimilation of all data before the current measurement time, and prior to assimilation of the current data; for the first iteration, i.e.  $\ell = 1$ ,  $m^\ell = m^p$ , and with full step length,  $\beta = 1$ , Eq. 5.10 is the same with the EnKF update equation.

Note that  $C_M$  is the covariance matrix for the prior model parameters  $m^p$ . It should not change during the Gauss-Newton iteration. However, the estimate of  $G_\ell$  changes with iteration.

For linear dynamic systems, when new data are obtained, both the model parameters and the state variables can be adjusted simultaneously with consistency, i.e. the system equations are honored by the updated model parameters and state variables. The result after data assimilation contains an improved estimate of the (non-varying) model parameters and also an improved estimate of the current value of the state variables.

However, for non-linear systems, it may be impossible to update the state variables to be consistent with the updated model parameters without re-solving the non-linear forward problem.

### 5.2.3 Implementation and computational cost of the EnRMLF

Like the EnKF, the EnRMLF is also a recursive process. We use the first step of the recursive process to illustrate the EnRMLF procedure in the following.

1. Compute the reservoir states variables using the updated model parameters,  $m_{j,0}^\ell$ , from the initial time 0 to the first measurement time  $t_1$

$$\psi(m_{j,0}^\ell, t : 0 \rightarrow t_1) \quad j = 1, 2, \dots, N_e. \quad (5.12)$$

where  $\psi(\cdot)$  denotes the reservoir simulator, the first argument specified in the parentheses are model parameters used to re-initialize the flow equations at initial time 0. The state variables for the re-initialization are determined by the initial conditions.

2. Update the model parameters using Eq. 5.11 to get  $m_{j,0}^{\ell+1}$  ( $j = 1, 2, \dots, N_e$ ).
3. Evaluate the data mismatch term for both  $m_{j,0}^\ell$  and  $m_{j,0}^{\ell+1}$

$$S(M_0^\ell) = \sum_{j=1}^{N_e} (g(m_{j,0}^\ell) - d_{\text{obs},j,1})^T C_D^{-1} (g(m_{j,0}^\ell) - d_{\text{obs},j,1}). \quad (5.13)$$

$$S(M_0^{\ell+1}) = \sum_{j=1}^{N_e} (g(m_{j,0}^{\ell+1}) - d_{\text{obs},j,1})^T C_D^{-1} (g(m_{j,0}^{\ell+1}) - d_{\text{obs},j,1}). \quad (5.14)$$

Note the computation of  $g(m_{j,0}^{\ell+1})$  involves solving the forward flow equations again from initial time 0 with  $m_{j,0}^{\ell+1}$

$$\psi(m_{j,0}^{\ell+1}, t : 0 \rightarrow t_1) \quad j = 1, 2, \dots, N_e. \quad (5.15)$$

4. If  $S(M_0^{\ell+1}) < S(M_0^\ell)$ , overwrite  $m^\ell$  with  $m^{\ell+1}$  and increase  $\beta_\ell$ ; otherwise, keep  $m^\ell$  and decrease  $\beta_\ell$ .

5. Check if the pre-set converging criteria are satisfied. If not, go to **Step 1** and iterate the procedure; otherwise, finish the first update step and go to the next.

Denote the final model parameters at the end of the first measurement time by  $m_j^f$  ( $j = 1, 2, \dots, N_e$ ). Note that at **Step 3**, the forward equations have been advanced from time 0 to  $t_1$  (Eq. 5.15) using the final model parameters, therefore for the first time at the second measurement time, we only need to continue advancing the forward equations from the first measurement time  $t_1$ , instead of 0, to  $t_2$ .

With the above illustration, it is obvious to see that the cost of solving the forward non-linear equations for the whole EnRMLF process is

$$\begin{aligned} T &= N_e (C_{0 \rightarrow t_1} + C_{t_1 \rightarrow t_2} + \dots + C_{t_{N_t-1} \rightarrow t_{N_t}}) + N_e \sum_{k=1}^{N_t} (2I_k - 1) C_{0 \rightarrow t_k} \\ &= N_e C_{0 \rightarrow t_k} \left( 1 + \sum_{k=1}^{N_t} (2I_k - 1) \right) \end{aligned}, \quad (5.16)$$

where  $I_k \geq 1$  is the number of Gauss-Newton iterations at measurement time  $t_k$ ;  $N_t$  is the total number of measurement times;  $C$  represents the cost for solving the forward equations for a particular time interval shown in its subscript;  $C_{0 \rightarrow t_1} + C_{t_1 \rightarrow t_2} + \dots + C_{t_{N_t-1} \rightarrow t_{N_t}} = C_{0 \rightarrow t_{N_t}}$ , which equals to one forward run from time 0 to the end of measurement time  $t_{N_t}$  for one simulation model.

In Eq. 5.16, the first term is equivalent to the computation cost of the forecast steps in the traditional EnKF; the second term is the cost incurred by re-solving the reservoir flow equations from the initial time 0 and iterations for the solution to converge. Even with only minimum iterations, one iteration at each measurement time, the extra cost is substantial. The computation cost is the major disadvantage of this method.

There are at least two steps, **Step 2** and **Step 5**, in the EnRMLF procedure worthy of further explanation for clarification. For Step 2, we need to answer how to compute the sensitivity coefficient matrix  $G_\ell$  and how to choose  $\beta_\ell$ . For Step 5, we need to explain what the converging criteria are.

There is no standard guidelines yet for choosing  $\beta$ . However, we have some rules of thumb other than trial and error.

- To start at every measurement time, if the simulated data are close to the observed data, choosing a bigger  $\beta$ ; otherwise, a smaller  $\beta$ .
- During the Gauss-Newton iteration, if  $S(M^{\ell+1}) < S(M^\ell)$ , increase  $\beta$  by 2; otherwise decrease by 2.

To answer how to determine the solution is converged or exit the Gauss-Newton iteration, we follow the criteria below.

- $\text{MAX}_{1 \leq i \leq N_m; 1 \leq j \leq N_e} |m_{i,j}^{\ell+1} - m_{i,j}^\ell| < \epsilon_1$  or
- $S(M^{\ell+1}) - S(M^\ell) < \epsilon_2 S(M^\ell)$  or
- $S(M^{\ell+1}) \leq n_D$  or
- Iteration exceeds the pre-set maximum number of iterations,  $I_{\text{MAX}}$ .

We used  $\epsilon_1 = 10^{-5}$ ,  $\epsilon_2 = 10^{-4}$ , and  $I_{\text{MAX}} = 6$  for the non-linear example showed in Section 5.4.

#### 5.2.4 Computation of sensitivity coefficient matrix

One feature that makes the implementation of the traditional EnKF so efficient is that it is never necessary to compute  $C_M$ , only the products of  $C_M G^T$  and  $G C_M G^T$  are needed. It is not as straightforward in the iterative method, because it is important to maintain the distinction between the model parameters covariance matrix estimate, which should be based on the prior models, and the sensitivity matrix, which should be based on the current values of model parameters. Let  $M^P$  be the ensemble of model parameters after assimilation of all previous data. Denote the mean of the prior model parameters by  $\bar{m}^P$  and the deviation from the mean by  $\Delta M^P$ . The ensemble estimate

of the prior model parameters covariance (after assimilation of all previous data) is  $C_M = \frac{\Delta M^P (\Delta M^P)^T}{(N_e - 1)}$ . At the  $\ell$ th iteration, let  $\Delta D^\ell$  represent the deviation of each vector of computed data from the mean vector of computed data. The sensitivity coefficient matrix,  $G_\ell$ , relates the changes in model parameters to the changes in computed data.

$$\Delta D^\ell = G_\ell \Delta M^\ell \quad (5.17)$$

The dimensions for matrix  $\Delta D^\ell$  are  $N_d \times N_e$ ; for  $\Delta M^\ell$  are  $N_m \times N_e$ ; for  $G_\ell$  are  $N_d \times N_m$ . The sensitivity matrix can be computed by

$$G_\ell = \Delta D^\ell (\Delta M^\ell)^\# , \quad (5.18)$$

where  $(\Delta M^\ell)^\#$  is the pseudo-inverse of  $\Delta M^\ell$  since it is not necessarily a square matrix. The pseudo-inverse matrix has a dimension of  $N_e \times N_m$ .

The following lists the step-by-step procedure for computing  $G_\ell$ .

1. Assemble model parameters for each ensemble member,  $m_{j,k}$  ( $j = 1, \dots, N_e$ ); Compute the vector of mean values of the model parameters,  $\bar{m}_k$ ; Compute the matrix of deviations of the model parameters from the mean,  $\Delta M_k$ .

$$m_{j,k} = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_N \\ \ln K_1 \\ \vdots \\ \ln K_N \end{bmatrix}_{j,k} , \quad (5.19)$$

$$\bar{m}_k = \frac{1}{N_e} \sum_{j=1}^{N_e} m_{j,k} . \quad (5.20)$$

The  $j$ th column of  $\Delta M_k$  is

$$\Delta m_{j,k} = m_{j,k} - \bar{m}_k , \quad (5.21)$$

where  $N$  is number of total roadblocks of the reservoir.

2. Assemble the simulated data for each ensemble member,  $d_{j,k}$  ( $j = 1, \dots, N_e$ );

Compute the vector of mean values of the simulated data,  $\bar{d}_k$ ; Compute the matrix of deviations of the simulated data from the mean,  $\Delta D_k$ .

$$d_{j,k} = \begin{bmatrix} d_{k,j,1} \\ \vdots \\ d_{k,j,N_d} \end{bmatrix}, \quad (5.22)$$

$$\bar{d}_k = \frac{1}{N_e} \sum_{j=1}^{N_e} d_{j,k}. \quad (5.23)$$

The  $j$ th column of  $\Delta D_k$  is

$$\Delta d_{j,k} = d_{j,k} - \bar{d}_k. \quad (5.24)$$

3. Compute  $G_l$

$$G_l = \Delta D.(\Delta M)^\# . \quad (5.25)$$

where the pseudo-inverse  $\Delta M^\#$  is computed by Singular Value Decomposition (Press et al., 1992).

### 5.2.5 EnRMLF with iteration when needed

When iterating at every measurement time, we have shown in Eq. 5.16 that the EnRMLF is fairly expensive and thus prohibitive for large scale problems. On the other hand, the traditional EnKF has been shown through some synthetic case studies that it handles non-linearity well when the non-linear relationship is not so severe. Combining these information, we suggest to only iterate using the Gauss-Newton formula when the traditional EnKF does not work well.

In this dissertation, we used the maximum change in saturation vectors to decide whether or not to apply iteration. If the maximum change of saturation exceeds 0.2, then iterate; otherwise, trust the EnKF solution. By so doing, extra computation is imposed only when informative data are assimilated and the total cost is significantly reduced without compromising the results, see Section 5.4.

## 5.3 A Linear Example

As shown in Section 5.2.1, the EnKF will work well for linear dynamic systems if the ensemble is large enough. Results from the EnRMLF and Conforming EnKF will be identical to results from the EnKF if the proposed iterative schemes are correct. In this section, a simple linear dynamic example is used to analyze performance of the two iterative schemes: Conforming EnKF and the newly proposed EnRMLF. The purposes of this test example are two-fold. First is to see if results from the two iterative methods are identical to results from the EnKF. Second, to compare the exact sensitivity matrix,  $G$ , with the approximation computed by Eq. 5.18 when the EnRMLF is applied.

### 5.3.1 Problem description

Assume that a passive tracer is injected at one end of a core and that the times for transport of the tracer to different locations of the core are measured. The flow is single phase and the fluid is incompressible; the length of the core is  $L$ ; the cross-sectional area is  $A$ ; the permeability is  $k(x)$ ; the porosity is  $\phi(x)$ ; and the viscosity of the fluid is  $\mu$ . For simplicity, assume a consistent set of units, so Darcy's Law is  $q = uA = \frac{kA}{\mu} \frac{\Delta p(x)}{x}$ , where  $q$  is the flow rate,  $u$  is the superficial velocity, and  $\Delta p(x)$  is the pressure drop from the inlet to location  $x$ . Incidentally, the velocity of the tracer front is  $v = \frac{u}{\phi} = \frac{q}{A\phi}$ .

Because the volumetric flow rate  $q$  throughout the core is constant, for a given location  $x$ , the arrival time of an ideal tracer that is injected at time 0 is

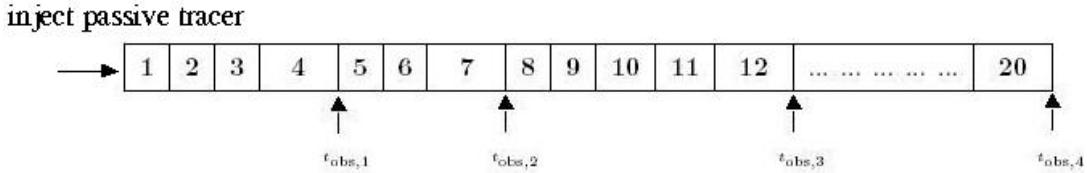
$$t_x = \frac{x}{v} = \frac{A}{q} \int_0^x \phi(x') dx'. \quad (5.26)$$

For discretized grids, the integration in Eq. 5.26 becomes summation

$$\begin{aligned} t_x &= \frac{A}{q} \sum_{i=1}^{i_x} \phi_i \Delta x_i \\ &= \frac{A \Delta x}{q} \sum_{i=1}^{i_x} \phi_i \quad \text{assuming } \Delta x_i = \Delta x, i = 1, \dots, N., \end{aligned} \tag{5.27}$$

where  $i_x$  is the discretized grid index corresponding to the location  $x$ ;  $N$  is the total number of the uniformly discretized grids. We can see from Eq. 5.27 that the travel time of the tracer to the grid  $i_x$  depends on the average porosity of the core from the inlet up to that grid. It does not depend on the porosity elsewhere, neither on the permeability distribution of the core, nor on the fluid viscosity.

The measured tracer arrival times to different locations are used to infer the porosity distribution of the core. The core is divided into 20 uniform grids. There are four locations where the arrival time is measured. Fig. 5.1 shows the four locations. Table 5.1 lists the parameters that are constant and related to the calculation of the tracer's travel time.



**Figure 5.1:** Locations where the tracer arrival times are measured for the linear example.

| Parameter                      | Value |
|--------------------------------|-------|
| $L$ (cm)                       | 10    |
| $\Delta x$ (cm)                | 0.5   |
| $A$ ( $\text{cm}^2$ )          | 1.0   |
| $q$ ( $\text{cm}^3/\text{s}$ ) | 0.02  |

**Table 5.1:** Constant parameters related to the calculation of the tracer arrival time for the linear example.

In this example, the data are assimilated sequentially beginning with the arrival time at  $x = 2$ , then arrival time at  $x = 3.5$ , then  $x = 6$ , and finally at  $x = 10$ . To

propagate the system from the previous measurement time (i.e. previous measured location) to the new measurement time, the computation only requires porosities from the previous location to the new location:

$$t_{x_k} = t_{x_{k-1}} + \frac{A\Delta x}{q} \sum_{i=1+i_{x_{k-1}}}^{i_{x_k}} \phi_i, \quad (5.28)$$

and to propagate the system from time 0 (i.e. location 0) is

$$t_{x_k} = \frac{A\Delta x}{q} \sum_{i=1}^{i_{x_k}} \phi_i, \quad (5.29)$$

where  $i_{x_{k-1}}$  and  $i_{x_k}$  are the integer grid indices corresponding to measurement location  $x_{k-1}$  and  $x_k$ , respectively.

The state vectors for this problem consist of the following variables

$$y_j = [\phi_1, \dots, \phi_N, t_x]_j^T \quad j = 1, \dots, N_e., \quad (5.30)$$

where  $t_x$  is the state variable in this case that would be re-computed with Eq. 5.28 and Eq. 5.29, respectively, when the Conforming EnKF and EnRMLF are used.

### 5.3.2 Observations

Table 5.2 lists the parameters used to generate the initial realizations of the porosity distribution for this linear example. The procedure to generate these porosity realizations is the same with that described in Section 3.2.1.

|                     |             |
|---------------------|-------------|
| <b>Prior mean</b>   | 0.25        |
| <b>STD</b>          | 0.05        |
| <b>Covariance</b>   | Exponential |
| <b>Range (grid)</b> | 4           |

**Table 5.2:** Parameters used to generate the initial realizations of the porosity distribution for the linear example.

The synthetic true porosity distribution is generated using the same parameters as those for the initial realizations. The four observations are computed with the

generated true porosity values using a similar equation as Eq. 5.27

$$\begin{aligned}
t_{\text{obs},k} &= t_{\text{true},k} + \epsilon_k \\
&= \frac{A\Delta x}{q} \sum_{i=1}^{i_k} \phi_{\text{true},i} + \epsilon_k \\
&= 25 \sum_{i=1}^{i_k} \phi_{\text{true},i} + \epsilon_k \quad k = 1, 2, 3, 4. ,
\end{aligned} \tag{5.31}$$

where the subscript  $k$  is the measurement index;  $i_k$  represents the grid at which the  $k$ th measurement is made;  $\epsilon_k$  is the measurement error drawn from uni-variate unbiased Gaussian distributions independently for each location. Table 5.3 shows the uncontaminated synthetic observations,  $t_{\text{true},k}$ , and the standard deviations of the Gaussian measurement errors,  $\epsilon_k$ , added to the synthetic data, at the four measurement times/locations.

| Measurement time index | Measurement time /location: $i_k$ (grid) | Arrival time: $t_{\text{true},k}$ (s) | STD of measurement error: $\sigma(\epsilon_k)$ (s) |
|------------------------|------------------------------------------|---------------------------------------|----------------------------------------------------|
| 1                      | 4                                        | 24.0553                               | 100.00                                             |
| 2                      | 7                                        | 43.9585                               | 0.25                                               |
| 3                      | 12                                       | 70.8425                               | 0.25                                               |
| 4                      | 20                                       | 119.9635                              | 0.25                                               |

**Table 5.3:** The uncontaminated synthetic observations, computed using the true porosity values, and the standard deviations of the unbiased Gaussian measurement errors, at the four measurement times/locations.

It should be noted that the measurement error at the first measurement time is intentionally given as a fairly large value. By assuming very inaccurate data at the first measurement time, the updated estimate of porosity is essentially unchanged from the initial porosity, as is the updated estimate of the state variable, tracer travel time. Setting a smaller step length, both iterative methods apply their update equations several times at a single measurement time. We can see the differences that would be resulted from the two iterative formulae.

At the second measurement time with accurate data, results from the Conforming

EnKF and EnRMLF will differ greatly because in one method the state variables are computed from previous measurement time/location, and in the second method, the state variables are re-computed from measurement time/location 0. With the Conforming EnKF, the estimate of porosity is updated using the EnKF. It will get the average porosity correct between the origin and the location of the measurement. However, the tracer travel times will be obtained by rerunning the computation of total porosity-length from the location of the first measurement to the location of the second measurement. The travel time to the first measurement location simply uses the old state variable (travel time from the original ensemble of porosities). With the EnRMLF, the estimate of porosity is updated using the Gauss-Newton formula. The tracer travel time will be obtained by computing the total porosity-length from the origin to the second measurement location. The result will be different from that obtained using the conforming step.

At the third and fourth measurement times, with the cumulative errors from the first two times, the distinction between the Conforming EnKF and EnRMLF will be more pronounced.

### 5.3.3 Criteria used to stop iterations

For the two iterative methods, both the full step length, i.e.  $\alpha = 1.0$ ,  $\beta = 1.0$ , and half step length, i.e.  $\alpha = 0.5$ ,  $\beta = 0.5$ , are used. At each measurement time, the following criteria are used to stop the iterations. If any of the criteria is satisfied, the iteration is terminated.

1.  $\text{MAX}_{1 \leq i \leq N; 1 \leq j \leq N_e} |\phi_{ij}^{\ell+1} - \phi_{ij}^\ell| < 1.0 \times 10^{-5}$ , or
2.  $S(M^{\ell+1}) - S(M^\ell) < 1.0 \times 10^{-6} S(M^\ell)$ , or
3.  $S(M^{\ell+1}) \leq 1$ , or
4. Iteration exceeds 20.

### 5.3.4 Comparison of the two iterative methods

Table 5.4 lists the final data mismatch ( $S(M) = \sum_{j=1}^{N_e} (d_{\text{obs},j} - Hy_j)^T C_D^{-1} (d_{\text{obs},j} - Hy_j)$ ), the number of iterations applied, and the criteria satisfied when iterations were terminated, for different methods at all measurement times with 30 ensemble members. First, compare the results from the EnRMLF with full and half step lengths to the results from the EnKF. Both formulations of EnRMLF obtain similar data mismatch with the EnKF at all measurement times. At the first measurement time, the EnRMLFs terminated their iterations when both the data mismatch and model parameters stopped changing. At all other measurement times with accurate data, iterations terminated because the data mismatch was sufficiently small. The EnRMLF with half step length has a slightly larger data mismatch, but we believe with more iterations, it would obtain the same value as the EnKF and the EnRMLF with full step length. The differences in the data mismatch and iteration criteria satisfied between the Conforming EnKF and the EnRMLF are apparent at all four measurement times. At the first measurement time, the Conforming EnKF methods used 20 iterations with the inaccurate data, while the EnRMLF used a smaller number. Recall that at the first assimilation time, both the Conforming EnKF and the EnRMLF go back to location 0 to re-compute the tracer travel time when the porosity is updated during iterations. They use different formulae, however, for the porosity correction, e.g. the Conforming EnKF changes the model covariance matrix,  $C_M$ , for different iterations, while the EnRMLF keeps the same  $C_M$  computed from prior estimates. Also, the EnRMLF keeps the prior term,  $m^p$  in the update formula, while the Conforming EnKF does not. At the second measurement time, as stated in Section 5.3.2, the large data mismatch results from the conforming step. At the two other measurement times, the Conforming EnKF still has large data mismatch even though it decreases with data assimilation.

Fig. 5.2 plots the histogram of the computed data from the 30 initial ensemble

| Methods                                  | Time | $S(M)/\text{No. of iterations}$ | Criteria satisfied |
|------------------------------------------|------|---------------------------------|--------------------|
| <b>EnKF</b>                              | 1    | 17.97/0                         | /                  |
|                                          | 2    | 0.12/0                          | /                  |
|                                          | 3    | 0.34/0                          | /                  |
|                                          | 4    | 0.11/0                          | /                  |
| <b>EnRMLF</b><br>$\beta = 1.0$           | 1    | 17.97/2                         | 1 & 2              |
|                                          | 2    | 0.11/1                          | 3                  |
|                                          | 3    | 0.34/1                          | 3                  |
|                                          | 4    | 0.11/1                          | 3                  |
| <b>EnRMLF</b><br>$\beta = 0.5$           | 1    | 17.97/9                         | 1 & 2              |
|                                          | 2    | 0.49/8                          | 3                  |
|                                          | 3    | 0.71/9                          | 3                  |
|                                          | 4    | 0.47/8                          | 3                  |
| <b>Conforming EnKF</b><br>$\alpha = 1.0$ | 1    | 17.07/20                        | 4                  |
|                                          | 2    | 5197.24/4                       | 1                  |
|                                          | 3    | 4549.85/4                       | 1                  |
|                                          | 4    | 573.44/5                        | 1                  |
| <b>Conforming EnKF</b><br>$\alpha = 0.5$ | 1    | 17.57/20                        | 4                  |
|                                          | 2    | 4051.56/13                      | 1                  |
|                                          | 3    | 3166.03/16                      | 1                  |
|                                          | 4    | 228.47/20                       | 1 & 4              |

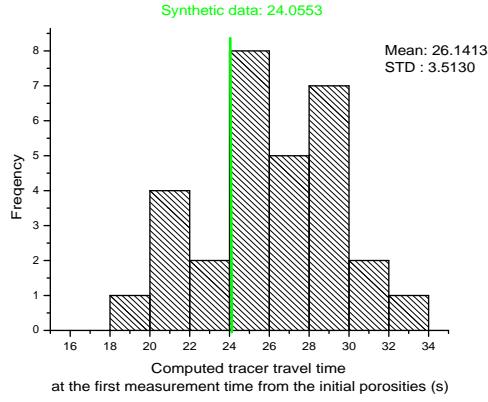
**Table 5.4:** Final data mismatch, number of iterations applied, and criteria satisfied when iterations are terminated, for different methods at all measurement times with 30 ensemble members for the linear example.

porosities at the first measurement time prior to data assimilation. Fig. 5.3 compares the histograms of the computed data from the Conforming EnKF and EnRMLF both with half step length at the end of their iterations for all measurement times. The green vertical line on each plot represents the value of the synthetic data. At the first measurement time with the inaccurate datum, the EnRMLF keeps the characteristic of the distribution of the computed data from the initial ensemble with a slightly lower ensemble data standard deviation, while the distribution from the Conforming EnKF is different and the data standard deviation is increased comparing to the results from the initial realizations. At all other measurement times, the computed data are centered around the observed data and the distribution of these data are Gaussian-like with the EnRMLF. With the Conforming EnKF, at the second measurement time, the mean of the computed data is about right, but at the third measurement time, the distribution of the computed data is biased, then at the last measurement time, the mean is nearly right again. Note at all times, the ensemble data standard deviation is higher from the Conforming EnKF than that from the EnRMLF.

Figs. 5.4 and 5.5 compare the mean and gridblock STD of the ensemble porosity estimates with 30 members from the EnKF, Conforming EnKF and EnRMLF both with half step length at the end of their iterations for all measurement times. The vertical green line on each plot indicates the measurement location. The results from the EnKF and EnRMLF are overlapped with each other. At the first two measurement times, the results from both iterative methods differ slightly, see Figs. 5.4(a), 5.4(b), 5.5(a), and 5.5(b). However, at the third and fourth measurement times, the differences become pronounced and the gridblock STD values are lower from the EnRMLF than those from the Conforming EnKF, see Figs. 5.4(c), 5.4(d), 5.5(c), and 5.5(d).

Fig. 5.6 shows the changes of the RMSE and STD of the ensemble porosity estimates with 30 ensemble members from the Conforming EnKF and EnRMLF both

with half step length during iterations at all measurement times. Note at the first measurement time, the STD from the Conforming EnKF is obviously increasing with iteration and the RMSE is also increasing with iteration at the third measurement time, which are indications of mistakes in the method.



**Figure 5.2:** Histogram of the computed tracer travel times with 30 ensemble members at the first measurement times prior to data assimilation.

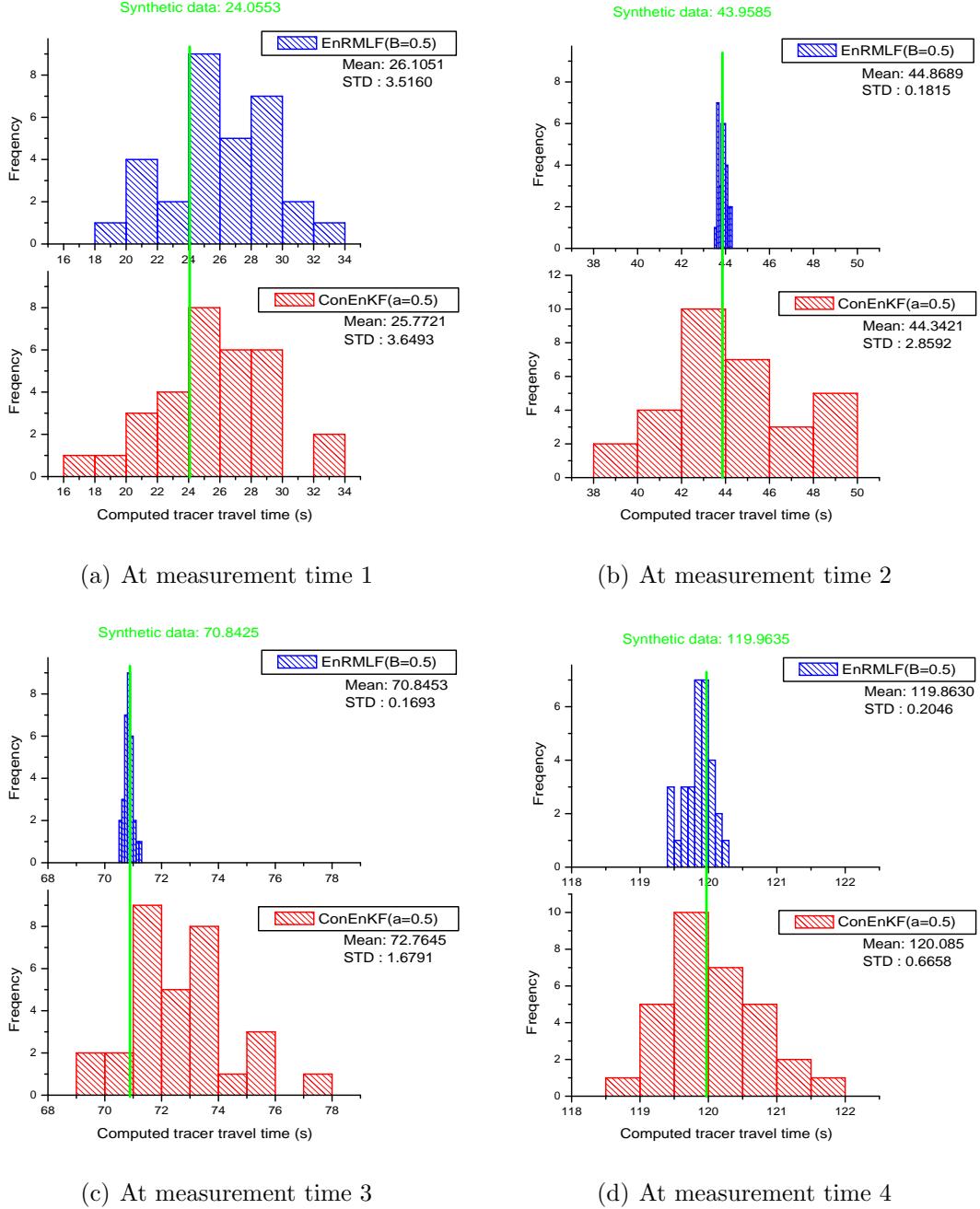
### 5.3.5 Computed sensitivity coefficients

The sensitivity coefficients are obvious for this linear case by examining Eq. 5.27. They form a step function

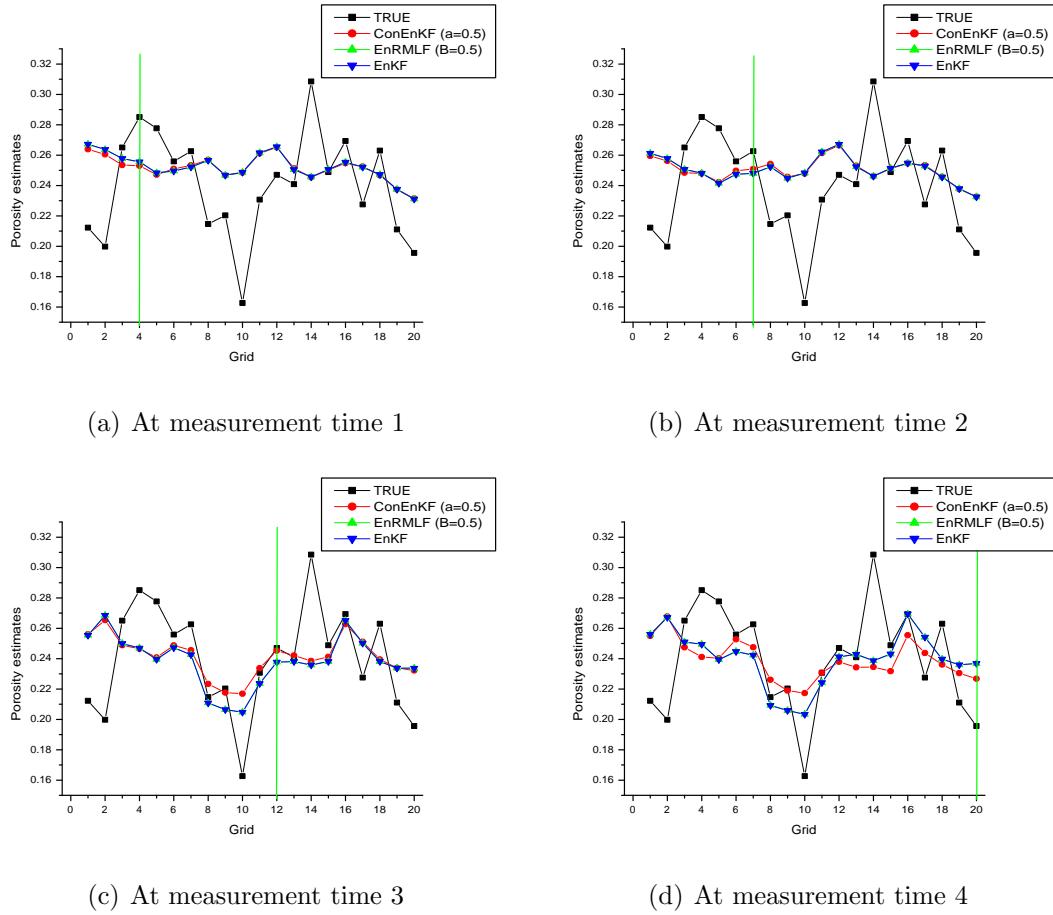
$$G_{t_x, \phi}(i) = \begin{cases} \frac{A\Delta x}{q} & \text{if } i \leq i_x \\ 0 & \text{if } i_x < i \leq N \end{cases}. \quad (5.32)$$

The sensitivity coefficients computed with different ensemble size at the third measurement time (location=12) are plotted in Fig. 5.7. It can be seen that when the size of the ensemble is larger than 21 (i.e.  $N + 1$ ), the computed sensitivities align with the analytical result (Eq. 5.32); when small-sized ensemble is used, the computed coefficients only provides an approximation to the real sensitivity values. With less members, the approximation becomes rougher.

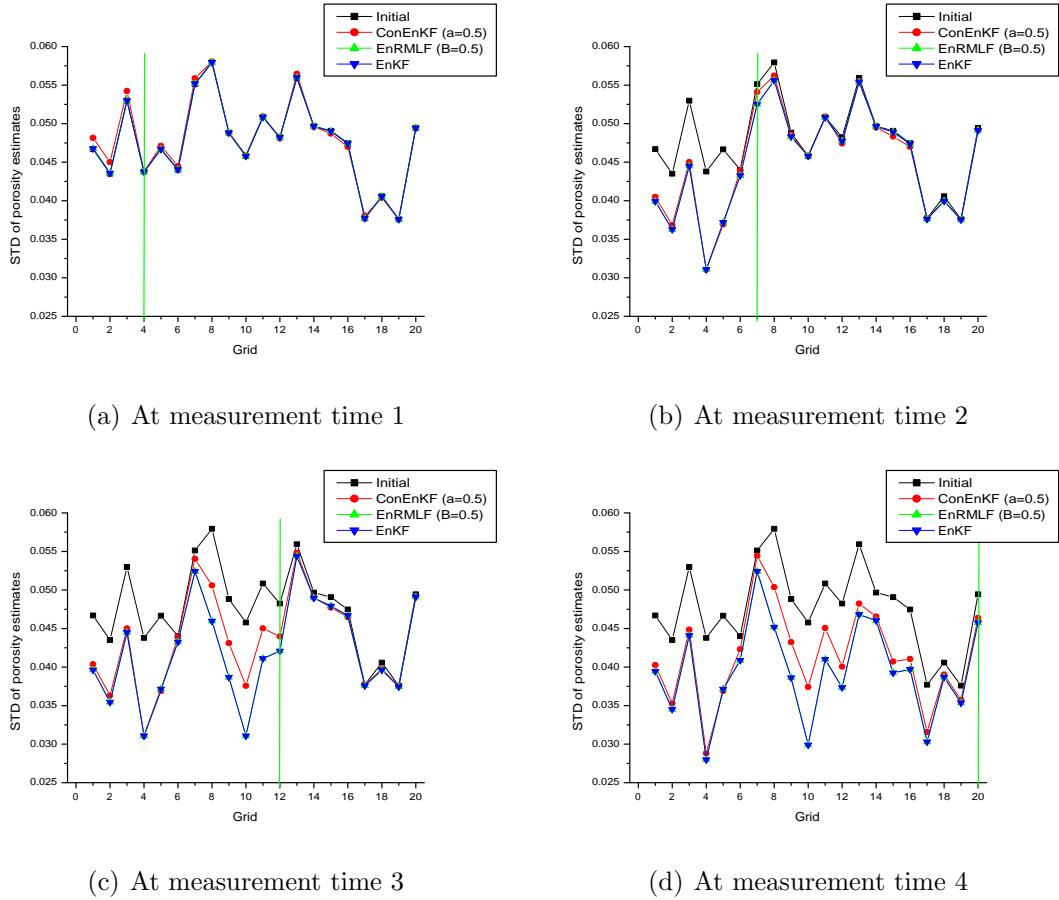
Fig. 5.8 plots the mean and gridblock STD of the ensemble porosity estimates



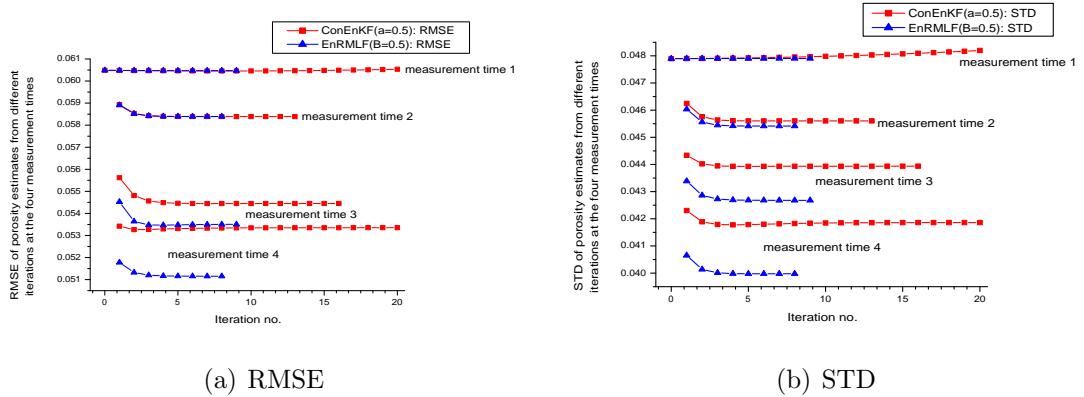
**Figure 5.3:** Histograms of the computed tracer travel times with 30 ensemble members from the Conforming EnKF and EnRMLF both with half step length at the end of their iterations for all measurement times.



**Figure 5.4:** Ensemble mean porosity with 30 ensemble members from the EnKF, Conforming EnKF and EnRMLF both with half step length at the end of their iterations for all measurement times.

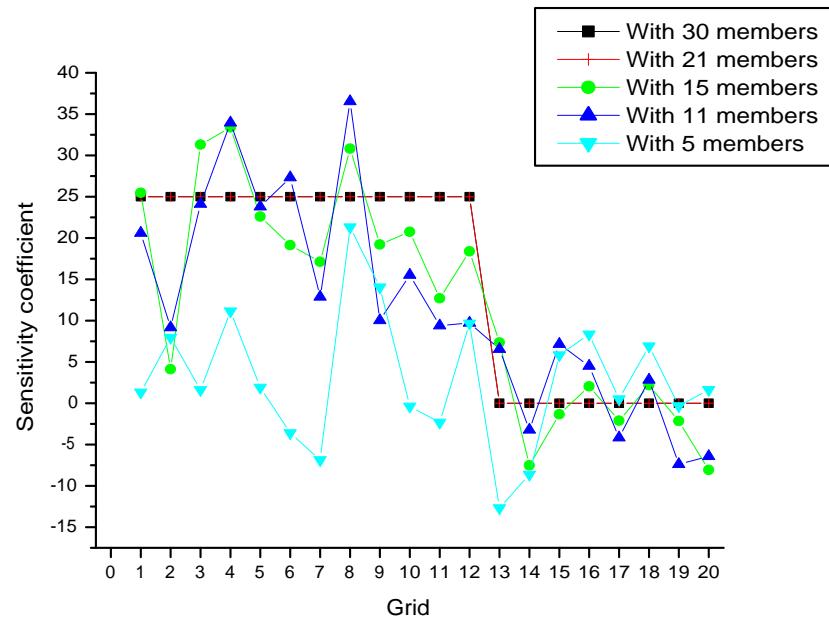


**Figure 5.5:** Gridblock STD of the porosity estimates with 30 ensemble members from the EnKF, Conforming EnKF and EnRMLF both with half step length at the end of their iterations for all measurement times.

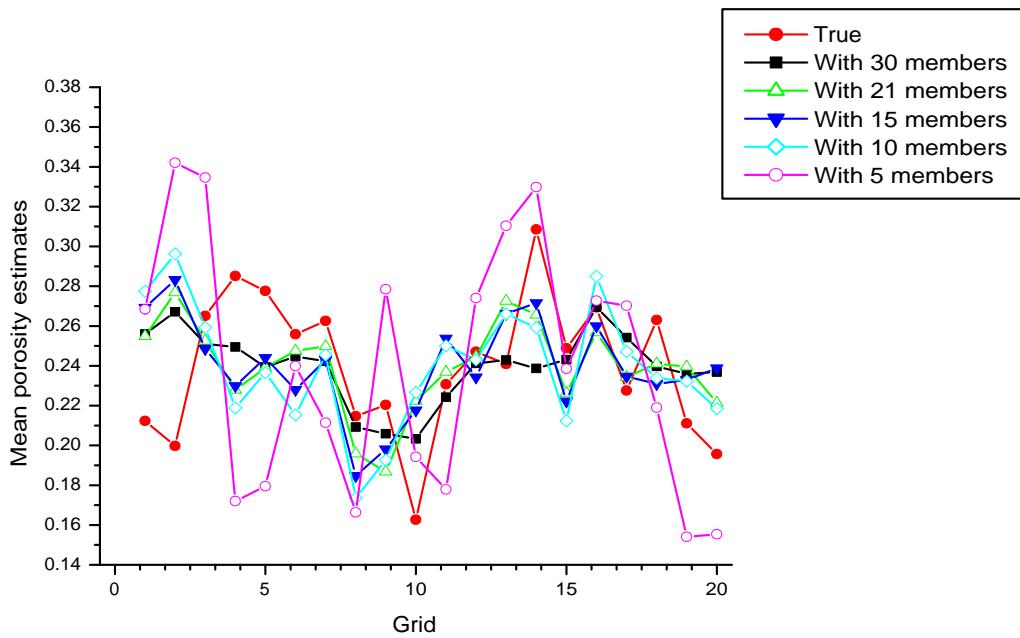


**Figure 5.6:** RMSE and STD of the ensemble porosity estimates with 30 ensemble members from the Conforming EnKF and EnRMLF both with half step length during iterations at all measurement times.

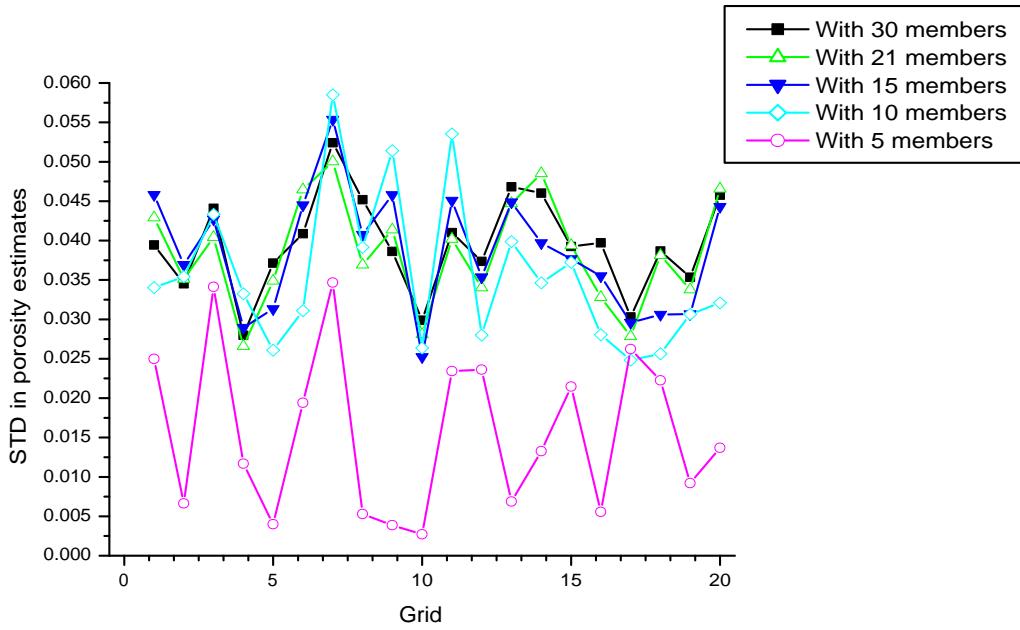
at the fourth measurement time (location=20) from the EnRMLF with different ensemble size. When more ensemble members are used, the mean estimate become smoother and the gridblock STDs are bigger. Fig. 5.9 compares the RMSE and STD of the ensemble porosity estimates with different ensemble size. Generally, with less ensemble members, the STD of the ensemble estimates is smaller. When only 5 members are used, the STD of the estimates drops dramatically and the error of the estimates changes little with data assimilation. When 10 members are used, the error increases after assimilating the 4th data.



**Figure 5.7:** Sensitivity coefficients computed by singular value decomposition (SVD) with different ensemble size at the third measurement time (location= 12).

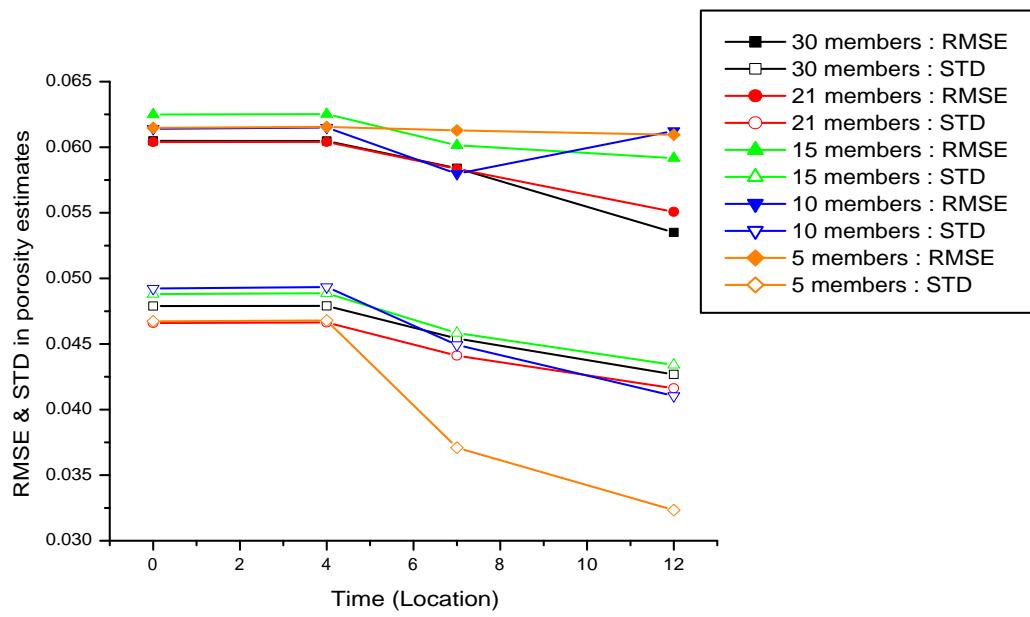


(a) Mean porosity estimates



(b) STD in porosity estimates

**Figure 5.8:** Ensemble mean porosity and gridblock STD of the ensemble porosity estimates at the fourth measurement time (location=20) from the EnRMLF with different ensemble size



**Figure 5.9:** RMSE and STD of the ensemble porosity estimates with different ensemble size from the EnRMLF.

## 5.4 A Non-linear Example

Various methods are tried on the non-linear example: (1) Conforming EnKF, (2) EnRMLF with iteration at every measurement time (EnRMLF1 is used to represent this method), (3) EnRMLF with iteration when needed (EnRMLF2 is used to represent this method), (4) traditional EnKF, and (5) traditional EnKF with truncation of  $S_w$  after Kalman correction. For the iterative methods, the criteria used to stop iterations has been described in Section 5.2.3. In the third method (EnRMLF2), the change in water saturations is used to decide whether or not it is needed to re-solve the flow equations. If the maximum change is smaller than 0.2, the re-computation of the state variables is not necessary and the updated state variables from the EnKF update equation are kept. Otherwise, the flow equations are re-solved from the initial time 0 for consistent state variables. The maximum saturation change criterion is also applied to the Conforming EnKF for the decision of whether or not to add the conforming step after the Kalman correction.

### 5.4.1 Problem description

The example is a water and oil two-phase flow problem in a one-dimensional discretized grid system of a core. There are 32 uniform grids. The physical dimension of each grid is  $60 \times 60 \times 40 \text{ ft}^3$ . The reservoir is initially saturated with oil and connate water. The initial reservoir pressure is 4000 *psia*. The mobility ratio of water and oil is 5. An injector is located at the left end, and a producer is located at the right end. An observation well is located at grid 16. The injector injects water at a constant rate of 150 *STB/Day* and the producer produces at a constant pressure of 1200 *psia*.

Fig. 5.10 shows a schematic setup for the problem.

The procedure to generate the initial rock properties is the same with that shown in Section 3.2.1 and not repeated here. Table 5.5 lists the parameters used to generate the initial realizations. 65 initial members are used for data assimilation.

| 1                       | ... ... ... ... | 16                          | ... ... ... ... | 32                    |
|-------------------------|-----------------|-----------------------------|-----------------|-----------------------|
| injector<br>fixed $q_w$ |                 | obsv. well<br>measure $S_w$ |                 | producer<br>fixed BHP |

**Figure 5.10:** Schematic setup for the non-linear example problem.

|                   | Porosity    | LnK (K in md) |
|-------------------|-------------|---------------|
| prior mean        | 0.2         | 5.5           |
| covariance        | exponential | exponential   |
| STD               | 0.04        | 0.7           |
| range (grid)      | 15          | 15            |
| cross-correlation | 0.6         | 0.6           |

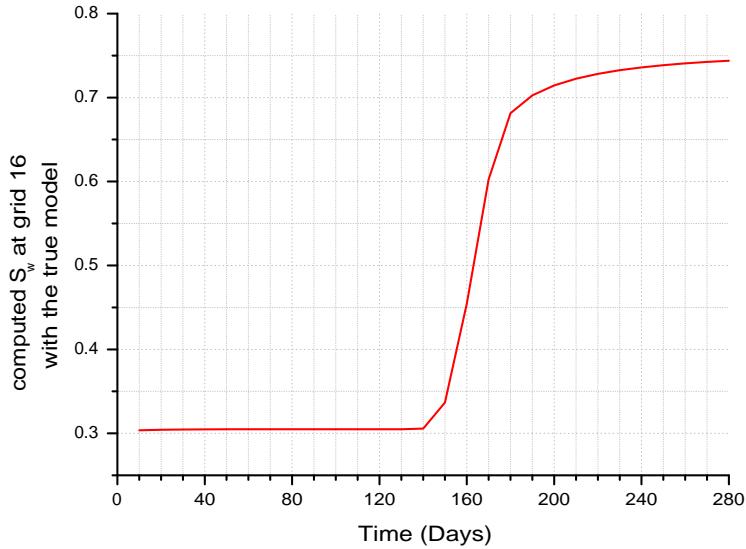
**Table 5.5:** Parameters used to generate initial realizations for the non-linear example.

#### 5.4.2 Observations

We assume that the only observed data available are water saturations measured at the observation well. The constant rate boundary condition at the left end resembles the setup in the linear case (Section 5.3). We have shown in the linear example that porosity distribution determines the speed of water movement for the single-phase fluid flow problem. Similarly for this two-phase problem, water saturation is related to the speed at which water moves, thus to porosity and relative permeabilities of both water and oil. Since the latter is treated without uncertainty in this example, the measured water saturation is sensitive to porosity up to the observation well after the movement of the water is felt on the observation well. It has no sensitivity before water breakthrough at the observation well or after water saturation reaches its highest value.

The true reservoir model is intentionally selected among all the generated models as the one that has the earliest water breakthrough at the observation well. Fig. 5.11 shows the continuous synthetic data profile before adding measurement noise at different times computed directly from the true reservoir model. Water does not break

through to the observation well until some time around day 160. Before the breakthrough time, most of the measured saturations are equal to the connate water saturation and carry no information to infer the rock properties of the core. Only measurements taken after day 160 contain useful knowledge for deducing the properties of the core. The measurement errors used in the water saturations are assumed to be Gaussian with mean 0 and STD 0.007.



**Figure 5.11:** Synthetic data before adding measurement noise at different times computed from the selected true reservoir model.

Two cases are created based on the example described above. In Case I, the first data assimilation is before water breakthrough at the observation well. In Case II, the first data is assimilated after water breakthrough. The results from various methods are shown as follow in different sections.

#### 5.4.3 Case I: first data assimilated before water breaks through the observation well

In this case, the measured water saturations are taken at 8 times: days 40, 90, 140, 190, 210, 230, 250, and 270. The synthetic values for the measurements are listed in Table 5.6. As explained in Section 5.4.2, the first three measurements are not informative to deduce the rock properties of the core. The rock properties can only

be corrected with the data at day 190 and afterwards.

| Time<br>(Days) | Synthetic $S_w$ data<br>w/o measurement noise added |
|----------------|-----------------------------------------------------|
| 40             | 0.3048                                              |
| 90             | 0.3049                                              |
| 140            | 0.3056                                              |
| 190            | 0.7027                                              |
| 210            | 0.7224                                              |
| 230            | 0.7326                                              |
| 250            | 0.7386                                              |
| 270            | 0.7425                                              |

**Table 5.6:** Synthetic data computed from the selected true reservoir model before adding measurement noise at 8 measurement times.

Table 5.7 summarizes the computational cost at the 8 different measurement times for the three iterative methods. One Kalman correction means the calculation of the Kalman gain matrix and updating the state vectors for each of the ensemble models. The cost associated with one Kalman correction is fairly low especially when the number of data at an individual measurement time is small, here only one datum at each measurement time. One “conforming run” means re-computing the state variables for each of the ensemble models at current measurement time by re-initiating the reservoir flow and transport equations at previous measurement time with the most up-to-date model parameters, corrected with the most recent data, and the state variables at the previous measurement time. The computational effort for one “conforming run” equals to running a reservoir simulator for all of the ensemble reservoir model from previous to current measurement time. One Gauss-Newton iteration means updating each of the ensemble state vectors using the Gauss-Newton formula and re-computing the state variables by re-initiating the reservoir flow and transport equation from the initial time 0 with the most up-to-date model parameters. The computation time for one Gauss-Newton iteration is the greatest among the three.

| Time | Conforming EnKF                            | EnRMLF1          | EnRMLF2                                   |
|------|--------------------------------------------|------------------|-------------------------------------------|
| 40   | 1 Kalman correction                        | 1 G-N iteration  | 1 Kalman correction                       |
| 90   | 1 Kalman correction                        | 1 G-N iteration  | 1 Kalman correction                       |
| 140  | 1 Kalman correction                        | 1 G-N iteration  | 1 Kalman correction                       |
| 190  | 6 Kalman correction<br>+ 6 Conforming runs | 6 G-N iterations | 1 Kalman correction<br>+ 3 G-N iterations |
| 210  | 6 Kalman correction<br>+ 6 Conforming runs | 1 G-N iteration  | 1 Kalman correction                       |
| 230  | 6 Kalman correction<br>+ 6 Conforming runs | 1 G-N iteration  | 1 Kalman correction                       |
| 250  | 3 Kalman correction<br>+ 2 Conforming runs | 1 G-N iteration  | 1 Kalman correction                       |
| 270  | 1 Kalman correction                        | 1 G-N iteration  | 1 Kalman correction                       |

**Table 5.7:** Computational cost for the three iterative methods at different measurement times for Case I.

#### 5.4.3.1 Results from the Conforming EnKF

Fig. 5.12 plots the water saturation profiles at day 140 before and after Kalman correction. For all of the profiles shown in this figure and subsequent figures, the red line is the result from the selected true reservoir model, the multiple black lines are the results from the 65 ensemble models, and the vertical straight line denotes the location of the observation well at grid 16. The theoretical data computed from the ensemble models can be seen at the intersection points of the vertical straight line to the saturation profiles. As mentioned above, the data measured before water breaks through the observation well tell no information about the core. Thus the two water saturation profiles are essentially the same before and after the data assimilation.

Fig. 5.13 shows the water saturation profiles at 190 days from the first and sixth iterations. Because the saturation profiles are problematic after the first Kalman correction, see Fig. 5.13(a), a “conforming run” is applied to enforce proper constraints on the state variables. After the porosities and permeabilities are updated using the data at day 190, the state variables at day 190 are re-computed by re-initiating the reservoir flow and transport equations at day 140 with the combination of the newly

updated model parameters and the state variables at day 140. Fig. 5.13(b) is the result from such a “conforming run”. Because of the biased characteristics of the profiles at day 140 (Fig. 5.12(b)), though the model parameters have been corrected with the data and are fairly different from their estimates at day 140, the conformed profiles at day 190 still have similar biased characteristics as that at day 140. Six iterations are applied at this single measurement time in an attempt to match the observed water saturation. Even after this many computational efforts, the observation is not yet honored and the saturation profiles are still severely biased, see Fig. 5.13(d).

Let us also look at how the saturation profiles affect the estimation of the porosity distribution. Fig. 5.14 plots the mean and STD of the ensemble porosity estimates at each gridblock from the 6 different iterations at day 190. After one iteration, there are the substantial improvements in the porosity estimate, see Fig. 5.14(a). The mean porosity values are lowered comparing to the initial mean values in order to match the observed fast speed of the water advancement. The associated STDs of porosity up to the measurement location are also decreased greatly from the initial STDs, while the rest of the STDs do not change much and stay close to their initial STD values due to the data sensitivity. Because of the biased saturation profiles resulted from the bias at day 140, the mean porosity values are further reduced with more iterations trying to move the water faster and catch up with the observed water movement, but the STD changes little after the second iteration.

Fig. 5.15 plots the measures of accuracy for both the estimates of porosity and saturation at day 190 from different iterations. In the  $x$ -axis, 0 iteration means before data assimilation at day 190. The porosity estimates are improved after the first iteration but move further away from the truth with more iterations (black line in Fig. 5.15(a)). The spread in the ensemble porosity estimates changes slightly after the second iteration (red line in Fig. 5.15(a)). Both the RMSE and spread in the saturation estimate oscillate. After a Kalman correction, the error and spread in the

saturation are both reduced. But after a conforming run, they bounce back to a higher values because of the biased saturation profiles. Overall, with more iterations, both the error and spread in the ensemble saturation estimates are slowly decreasing. All the observations made based on this plot are consistent with what has been observed from Fig. 5.13.

In the conforming EnKF, from the fourth measurement time, day 190, to the sixth measurement time, day 230, the number of iterations used at each time equals to the pre-set maximum number of iteration, 6, see column 2 at Table 5.7. All other criteria for solution convergence are not satisfied at these three times. Fig. 5.16 plots the final saturation profiles at different measurement times after day 190. At day 210, the biased characteristics of the conformed profiles still remains as that at days 140 and 190. The bias becomes less severe at day 230. At day 250, after the third Kalman correction, no conforming run is enforced because of the small changes in the saturation profiles and the saturations after Kalman correction are saved for the calculation of state variables at day 270. The saturation profiles after the Kalman correction look fairly decent at day 270, see Fig. 5.16(d). As a whole, the bias seems to diminish with time in this case.

Base on the above explanation, it is not surprising to see that the error in the porosity estimates does not drop until day 190 and becomes larger with time again after day 190 until the last measurement time, day 270, see Fig. 5.17(a). This method also gives the largest error and spread in saturation estimates comparing to the two other iterative methods until day 270, see Fig. 5.17(b).

#### *5.4.3.2 Results from the EnRMLF with iteration at every measurement time*

Fig. 5.18 shows the saturation profiles at day 140 before and after one Gauss-Newton iteration. The two profiles are essentially the same. Therefore there is no need to re-compute the state variables from the initial time 0 when the model parameters

change little, but rather to use the updated results from the Kalman correction. By so doing, the computational cost can be significantly cut down. Section 5.4.3.3 will show the results from applying Gauss-Newton iterations only when needed at some measurement times. However, in this section, Gauss-Newton iterations are used at every measurement time regardless of how small the changes in the model parameters and state variables.

Fig. 5.19 plots the saturation profiles at day 190 from the first and sixth Gauss-Newton iterations. Unlike the Conforming EnKF, after the model parameters are updated at day 190, they are taken to the initial time 0 to compute the state variables. It can be seen from the intersection points of the vertical straight line to the saturation profiles in Fig. 5.19(b) that the datum is honored well after the sixth iteration.

The mean and STD of the ensemble porosity estimates at each gridblock from the six Gauss-Newton iterations at day 190 are plotted in Fig. 5.20. The measures of accuracy in both the estimates of porosity and saturation at day 190 from different iterations are compared in Fig. 5.21. As expected, the errors in both estimates keep decreasing with more iterations.

The third column in Table 5.7 lists the Gauss-Newton iterations applied for this method at the 8 different measurement times. Only at day 190, 6 iterations are used. At all other measurement times, only one iteration is used for the convergence of the solution. Fig. 5.22 plots the final saturation profiles after one Gauss-Newton iteration at each of the four measurement times after day 190. Data are honored fairly well at each time with only one Gauss-Newton iteration.

The error in the estimate of porosity from this method is the smallest among all other methods, see Figs. 5.17(a) and 5.29(a).

#### 5.4.3.3 Results from the EnRMLF with iteration only when needed

As explained with Fig. 5.18 in Section 5.4.3.2, when the changes in the model parameters are small, there is no need to waste the effort to compute the state variables from the initial time 0, but rather to keep the traditional Kalman corrected results. Thus, for this method, at each measurement time, a general Kalman correction to the state vectors is applied first (equivalent to the first Gauss-Newton iteration formula with  $\beta = 1.0$ ). If the maximum water saturation change exceed a pre-set threshold value (0.2 for this case), it is decided that the state variables computed from the Kalman correction is not reliable and they have to be re-computed with the updated model parameters from the initial time 0. Otherwise, trust the results from the Kalman correction for the state variables. The fourth column in Table 5.7 shows that Gauss-Newton iterations is needed only at day 190, and at all other measurement times, only Kalman correction is applied. Comparing to the method with iteration at every measurement time, this method greatly brings down the computational cost. We need to see in the following if it compromises the results.

Figs. 5.23 and 5.24 plot the saturation profiles at day 190 and afterwards. Comparing to their counterparts produced by EnRMLF with iteration at every measurement time in Figs. 5.19 and 5.22, respectively, in terms of the closeness to the true saturation profile, this method gives better estimates of saturation, which are also reflected by the RMSE curve in Fig. 5.17(b). The difference in the porosity estimates from both methods is small, see Fig. 5.17(a). It is clear that the results produced by this method do not compromise at all comparing to these in Section 5.4.3.2 for this case.

#### 5.4.3.4 Results for traditional EnKF with/without sw truncation

For engineering practices, the easiest way to impose constraints on  $S_w$  is to truncate the values that are out of physical bounds after applying Kalman corrections. It hardly adds additional cost for the whole history matching procedure and works

quite well for this example.

Figs. 5.25 and 5.26 plot the saturation profiles after Kalman correction at days 190, 210, and 270 for the traditional EnKF with and without truncation of  $S_w$ , respectively. If the non-physical values of  $S_w$  are not truncated, they are inheritable when propagating forward in time. However, if they are simply truncated to reasonable values, the problematic saturations seem to go away with time.

The error and spread in both of the estimates of porosity and saturation are shown in Fig. 5.27. For the porosity estimate, the error and spread with and without truncation differ slightly. The difference may be caused by the truncation scheme. However, the truncation results in distinction on saturation estimates. Both the error and spread in the saturation estimates are much lower with truncation than without truncation, which can be easily understood by examining the saturation profiles produced by the two methods in Figs. 5.25 and 5.26.

#### 5.4.4 Summary for Case I

Fig. 5.28 summarizes the data match at the 8 different measurement times for the three iterative methods. The solid red dots on each of the plot are the synthetic data computed directly from the selected true reservoir model; the multiple hollow black dots at each measurement time are the final data from the ensemble models. The Conforming EnKF does not match the data at days 190, 210, and 230, although great computing effort are made at those times (6 “conforming runs” at each of the three measurement times). Both EnRMLF methods match all data fairly well.

Fig. 5.29 compares the RMSEs and ensemble STDs of both porosity and saturation estimates from the above five methods. It can be obviously seen from the plot that:

- Conforming EnKF fails to provide valid estimates for porosity. The error in the saturation estimates is the largest until the last measurement time, day 270, among all methods.

- At day 190, the two EnRMLF methods give equivalent estimates of porosity; for the estimates of saturation, iterating when needed obtained better solution than iterating at every measurement time.
- At times after day 190, EnRMLF with iteration at every measurement time achieves slightly better solution for porosity than all other methods but in a somewhat expansive way.
- In terms of the accuracy in the saturation estimate, EnRMLF with iteration when needed outperforms all other methods at all the 8 measurement times.
- The traditional EnKF with and without truncation of  $S_w$  obtain similar answers for porosity estimates, but truncation provides much better saturation profiles.

#### **5.4.5 Case II: first data assimilated after water breaks through the observation well**

For this case, it is assumed that the observed data are taken at 5 times: days 190, 210, 230, 250, and 270. Other setups are exactly the same with those in Case I. Table 5.8 summarizes the computational cost at each of the five measurement times for the three iterative methods. Both EnRMLF methods have the same expense spent at the five times as that in Case I (compare Table 5.8 to Table 5.7). But for the Conforming EnKF, the cost spend on the two cases is quite different. For Case I, 6 conforming runs are applied at each of the three measurement times, days 190, 210, and 230; for case II, only at day 190, two conforming runs are used and no conformation is needed at all other times.

EnRMLF with iteration at every measurement time would perform similarly as in the first case because it always re-solves the reservoir flow and transport equations from the initial time 0. EnRMLF with iteration when needed would change slightly as in the first case. But the Conforming EnKF would perform rather differently than

| Time | Conforming EnKF                       | EnRMLF1          | EnRMLF2                                   |
|------|---------------------------------------|------------------|-------------------------------------------|
| 190  | 3 Kalman correction<br>+ 2 Conforming | 6 G-N iterations | 1 Kalman correction<br>+ 3 G-N iterations |
| 210  | 1 Kalman correction                   | 1 G-N iteration  | 1 Kalman correction                       |
| 230  | 1 Kalman correction                   | 1 G-N iteration  | 1 Kalman correction                       |
| 250  | 1 Kalman correction                   | 1 G-N iteration  | 1 Kalman correction                       |
| 270  | 1 Kalman correction                   | 1 G-N iteration  | 1 Kalman correction                       |

**Table 5.8:** Computational cost for the three iterative methods at different measurement times for Case II.

the first case. Thus we will only examine some of the results from the Conforming EnKF.

Fig. 5.30 plots the saturation profiles at day 190 from the Conforming EnKF. The saturation profile after first Kalman correction would be the same as that in Case I (Fig. 5.13(a)), but the profiles change greatly after the first conforming run, comparing Fig. 5.13(b) to Fig 5.30(a). Because in Case II, the updated model parameters are taken to time 0 since day 190 is the first measurement time, instead of day 140 as in Case I, to re-initiate the state of the reservoir and re-compute the state variables at day 190. This also results in the difference at the subsequent two Kalman corrections at day 190.

Fig. 5.31 plots the error and spread in both estimates of porosity and saturation at day 190. We can see from the black line in Fig. 5.31(a) that the error in the porosity estimates decreases with iteration, unlike that in Case I, see Fig. 5.15(a). The spread in the estimate is also decreasing with iteration. For the saturation estimate, both the error and spread decreases with iteration. At the first iteration, the two points at each of the two curves are from Kalman correction and conforming run. Opposite to what has been observed in Case I, the higher-value point is from Kalman correction and the lower-value one is from conforming. At the second iteration, the distinction between Kalman correction and conforming is hardly seen. Overall, the error and spread in saturation estimate decreases rapidly with iteration.

We would expect that the estimates for the model parameters from the two cases are different at day 190. Fig. 5.32 shows a comparison of the mean porosity estimates from both cases. The mean estimate from Case I has lower values than Case II. It is understandable because the further reduction in porosity is attempting to push the water move faster toward the observation well in Case I due to the biased characteristics.

Fig. 5.33 shows the final saturation profiles at times after day 190. They all are different from their counterparts in Case I, see Fig. 5.16.

#### 5.4.6 Summary for Case II

From the final saturation profiles in Figs. 5.30(d) and 5.33, we can see that the data are honored quite well. Fig. 5.34 confirms the good data match for Case II with the Conforming EnKF.

Fig. 5.35 compares the RMSE and STD in both estimates of porosity and saturation of the 5 methods for Case II. The Conforming EnKF seems to work well for this case. In terms of porosity estimate, it provide reasonable solution as other methods. But the error in the saturation estimates is increasing rapidly with measurement time. EnRMLF with iteration when needed still gives the best saturation estimate.

## 5.5 Discussion

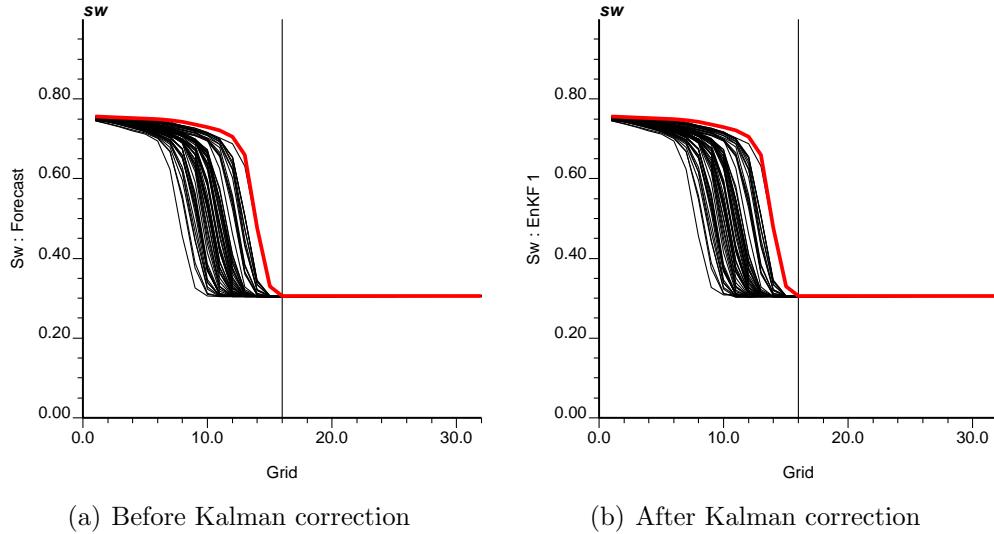
In this chapter, we applied the traditional EnKF and the two iterative methods, Conforming EnKF and EnRMLF, to both linear and non-linear examples.

In the linear example, the EnRMLF achieves equivalent solution as the traditional EnKF, where the validity of the solution is guaranteed when the ensemble is large enough, and matches the observed data well. However the Conforming EnKF could not match the data.

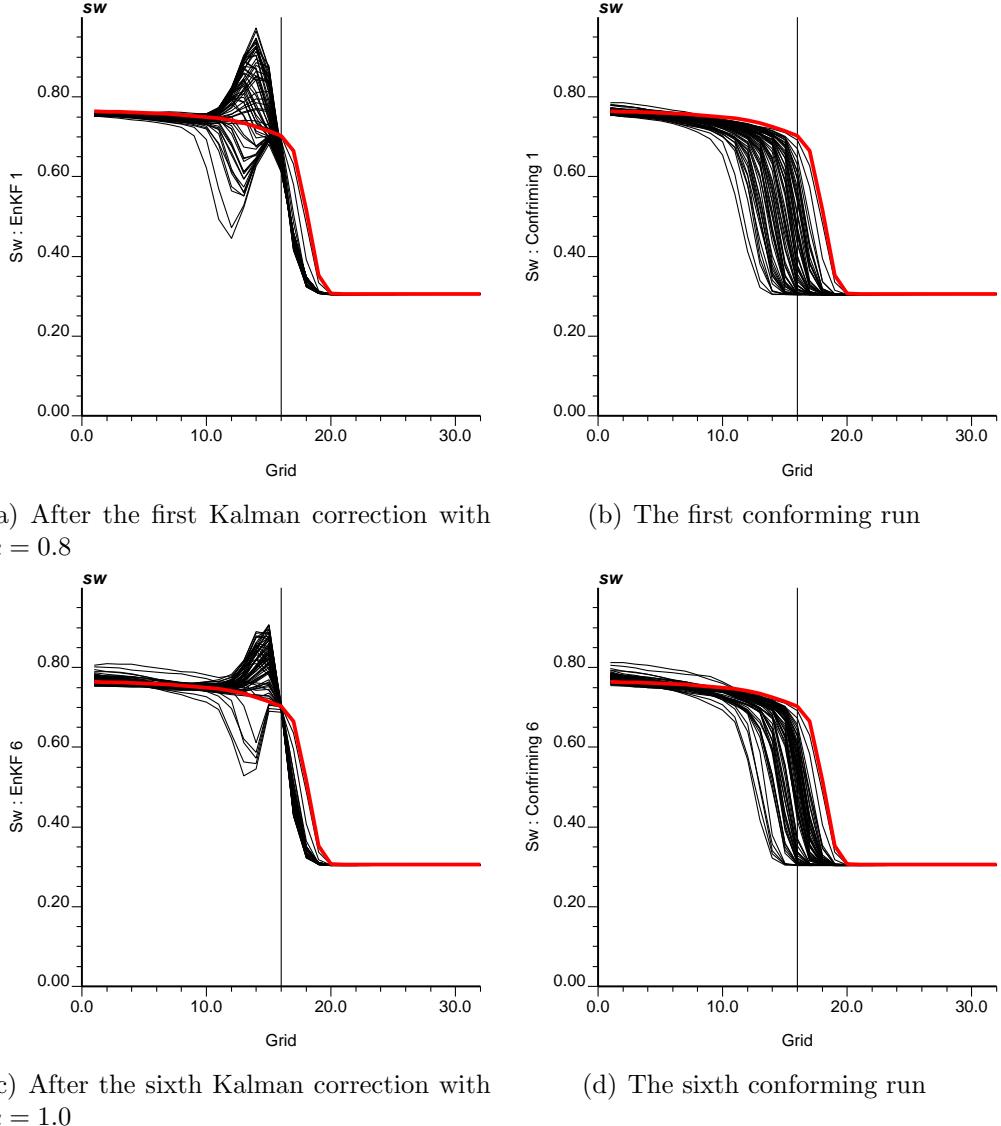
For Case I in the non-linear example, the conforming EnKF failed to provide valid solution for both porosity and saturation estimates because of the apparent

mismatched between model parameters and state variable used to re-initialize the reservoir flow equation at day 190.

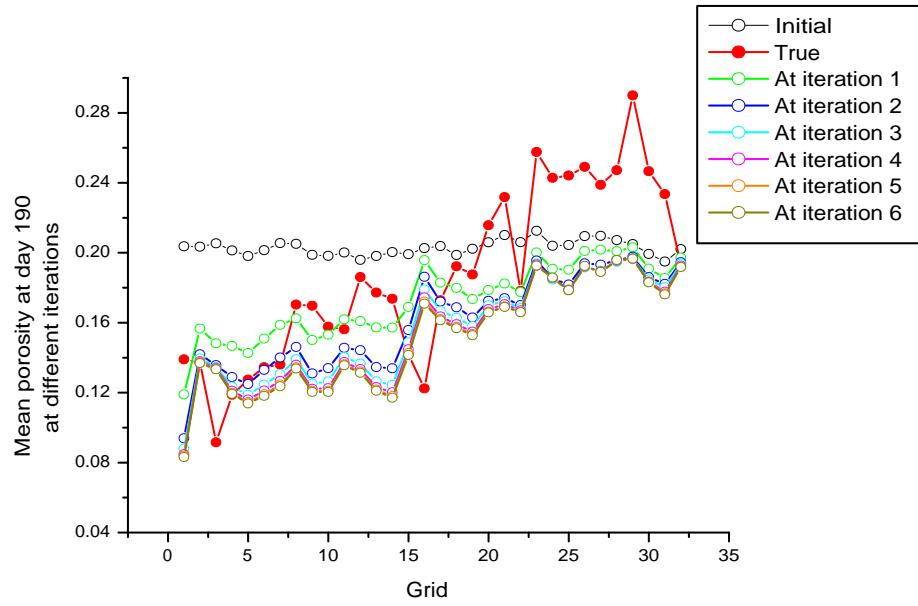
The EnRMLF with iteration at every measurement time is robust and provides best estimates for porosity for both cases in the non-linear example but in a somewhat expensive way. The EnRMLF with iteration only when needed dramatically cuts down the computational cost while not compromising the results.



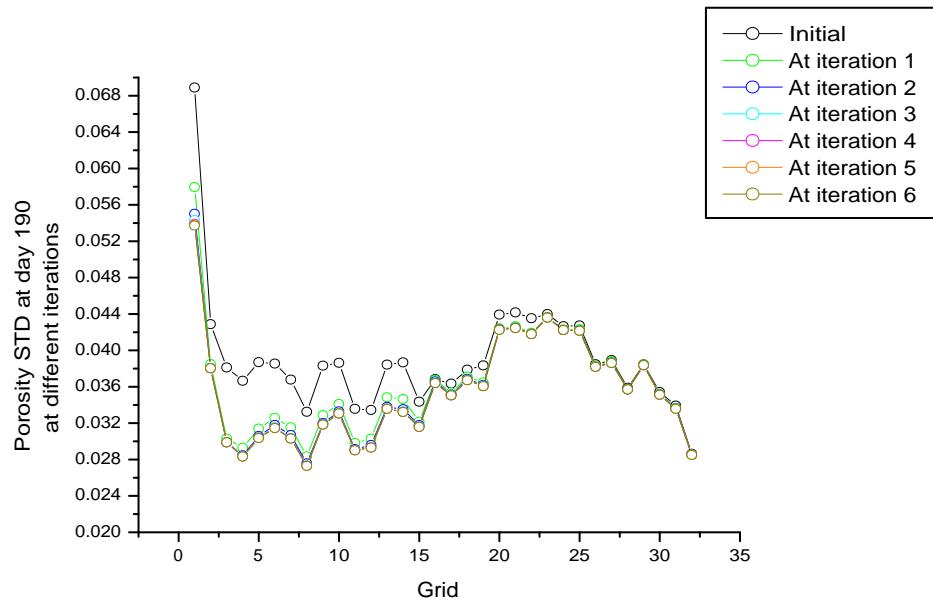
**Figure 5.12:** Conforming EnKF:  $S_w$  profiles at day 140 for Case I.



**Figure 5.13:** Conforming EnKF:  $S_w$  profiles at day 190 for Case I.

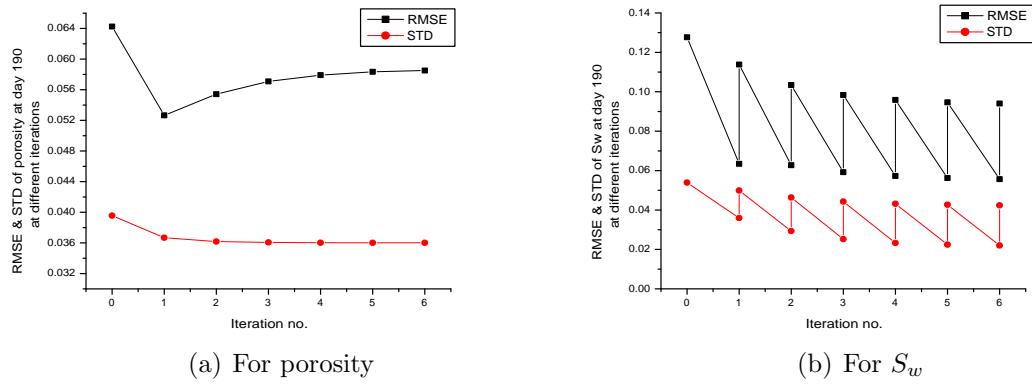


(a) Mean porosity during iterations

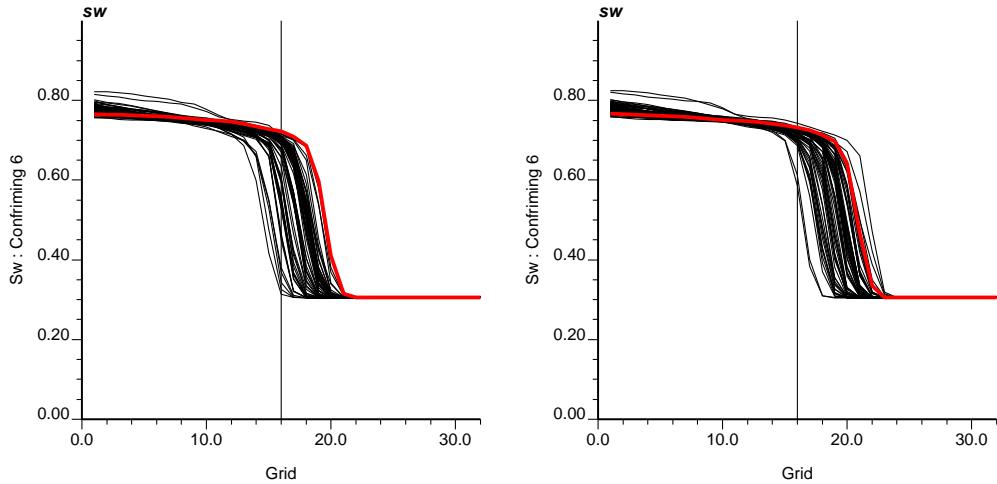


(b) Porosity STD during iterations

**Figure 5.14:** Conforming EnKF: gridblock mean and STD of the ensemble porosity estimates at day 190 from different iterations for Case I.

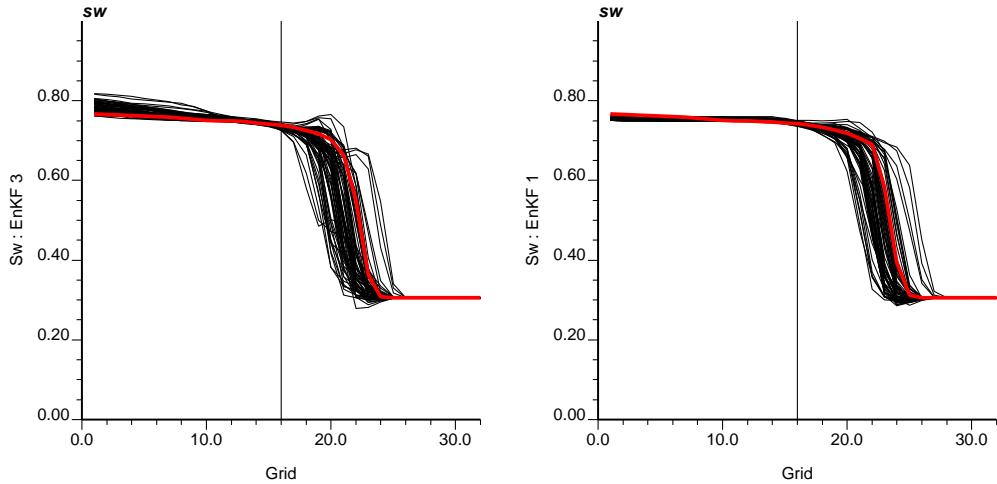


**Figure 5.15:** Conforming EnKF: RMSE and spread of porosity and  $S_w$  estimates at day 190 from different iterations for Case I.



(a) The sixth conforming run at day 210

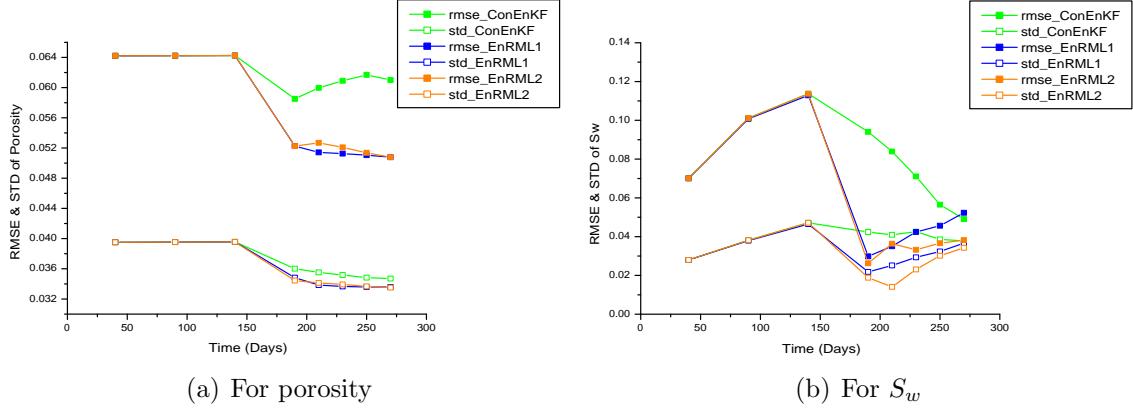
(b) The sixth conforming run at day 230



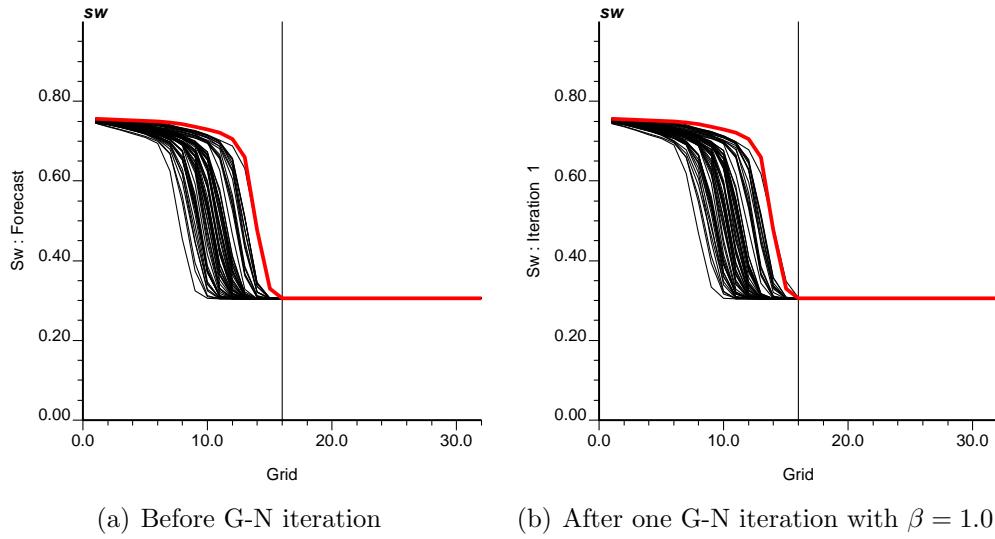
(c) After the third Kalman correction at day 250

(d) After Kalman Correction at day 270

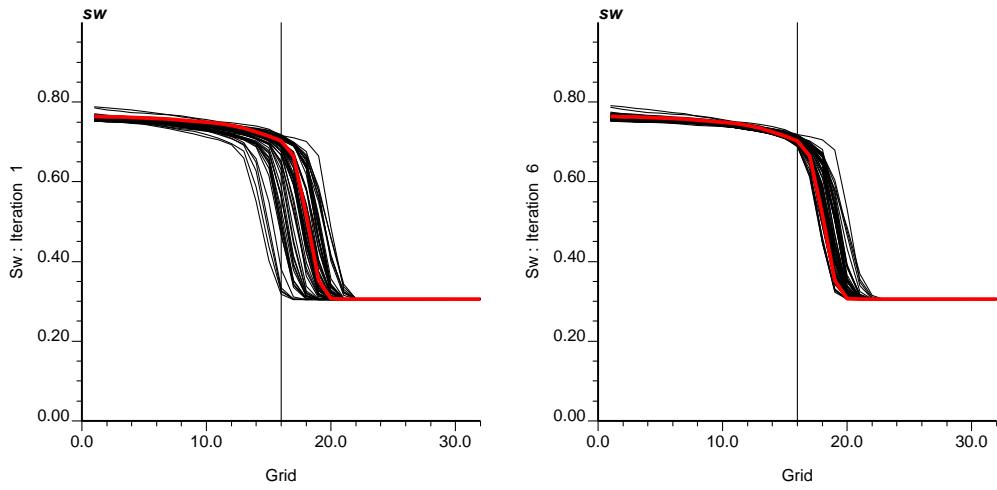
**Figure 5.16:** Conforming EnKF: the final  $S_w$  profiles at days 210, 230, 250, and 270 for Case I.



**Figure 5.17:** RMSE and spread of porosity and  $S_w$  estimates from the three iterative methods for Case I.

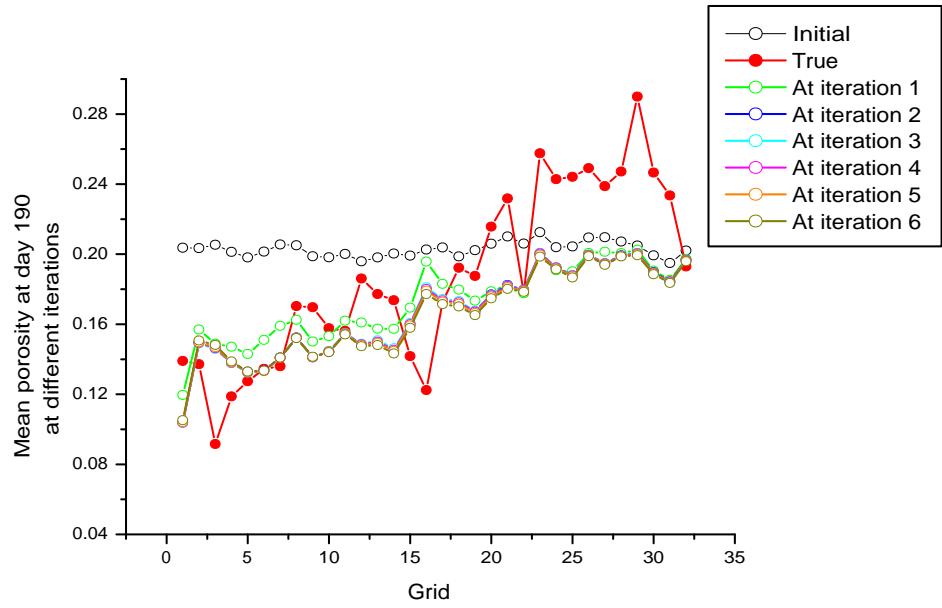


**Figure 5.18:** EnRMLF with iteration at every measurement time:  $S_w$  profiles at day 140 before and after G-N iteration.

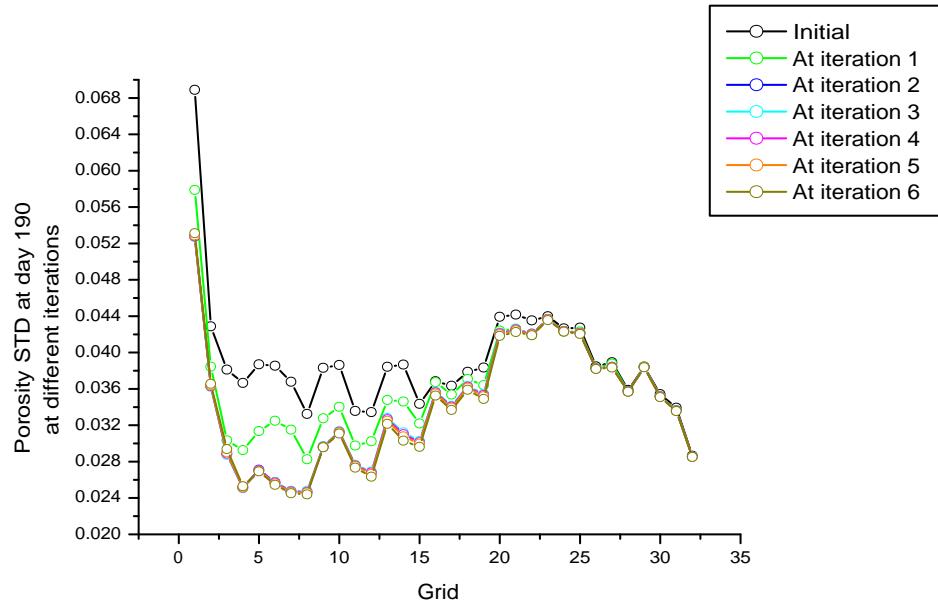


(a) After the first G-N iteration with  $\beta = 0.8$  (b) After the sixth G-N iterations with  $\beta = 1.0$

**Figure 5.19:** EnRMLF with iteration at every measurement time:  $S_w$  profile at day 190.

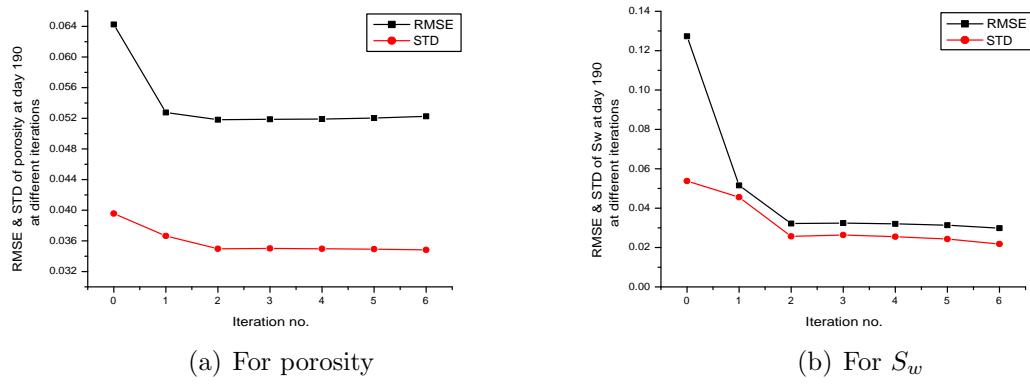


(a) Mean porosity during iterations

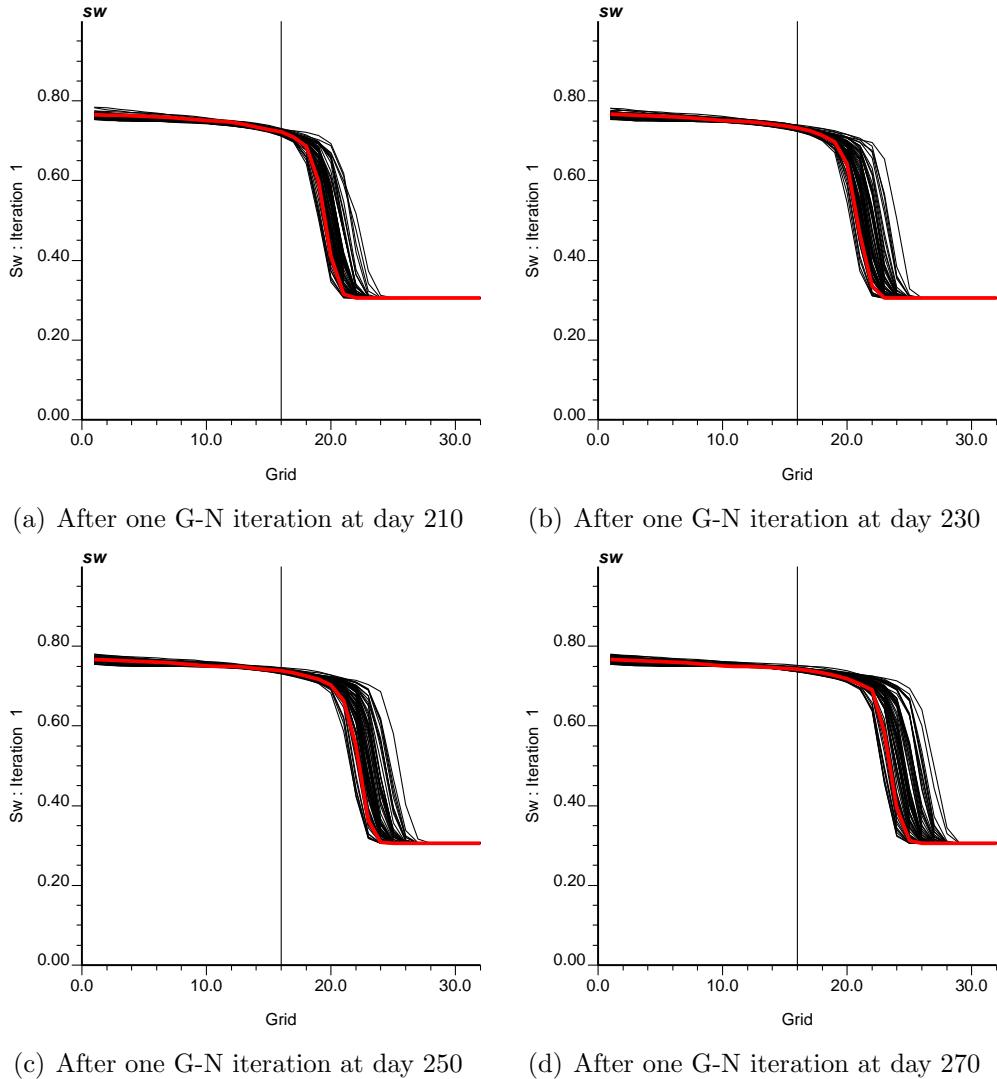


(b) Porosity STD during iterations

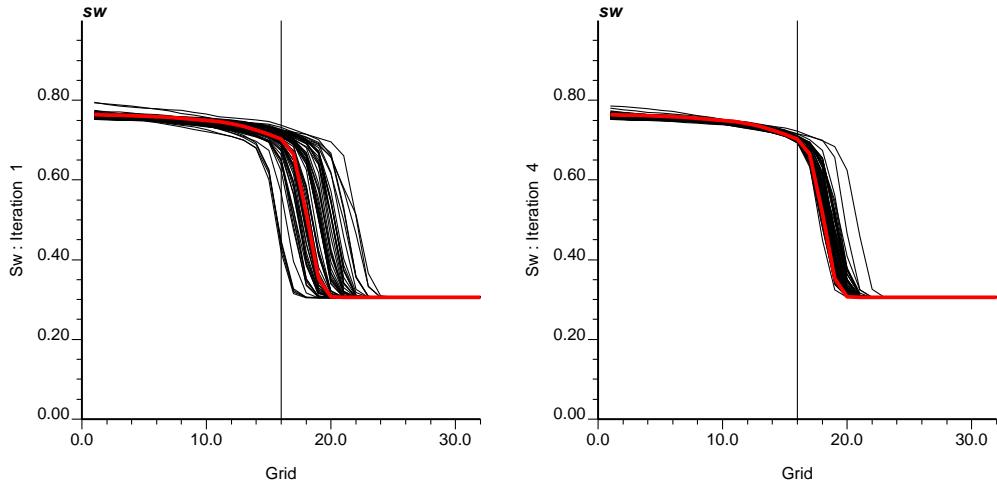
**Figure 5.20:** EnRMLF with iteration at every measurement time: mean and STD of porosity at day 190 from different iterations for Case I.



**Figure 5.21:** EnRMLF with iteration at every measurement time: RMSE and spread of porosity and  $S_w$  estimates at day 190 from different iterations for Case I.



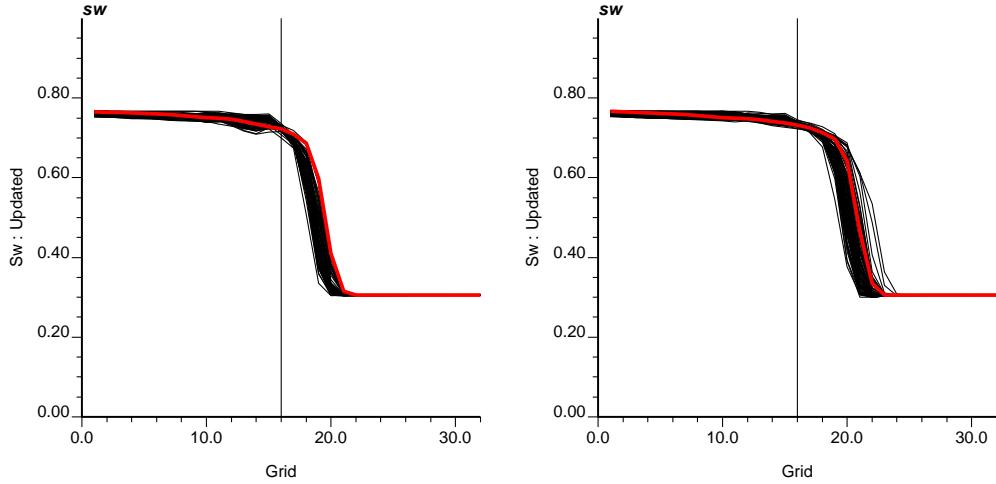
**Figure 5.22:** EnRMLF with iteration at every measurement time: the final  $S_w$  profiles at days 210, 230, 250, and 270 for Case I.



(a) Re-computed from the initial time after the one Kalman correction on the model parameters

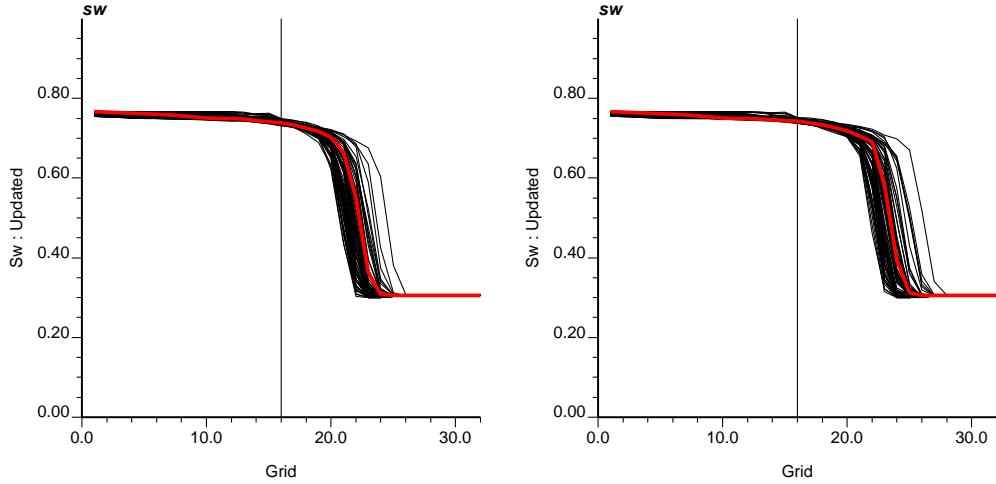
(b) After the third G-N iteration

**Figure 5.23:** EnRMLF iterate when needed:  $S_w$  profiles at day 190 for Case I.



(a) After Kalman correction at day 210

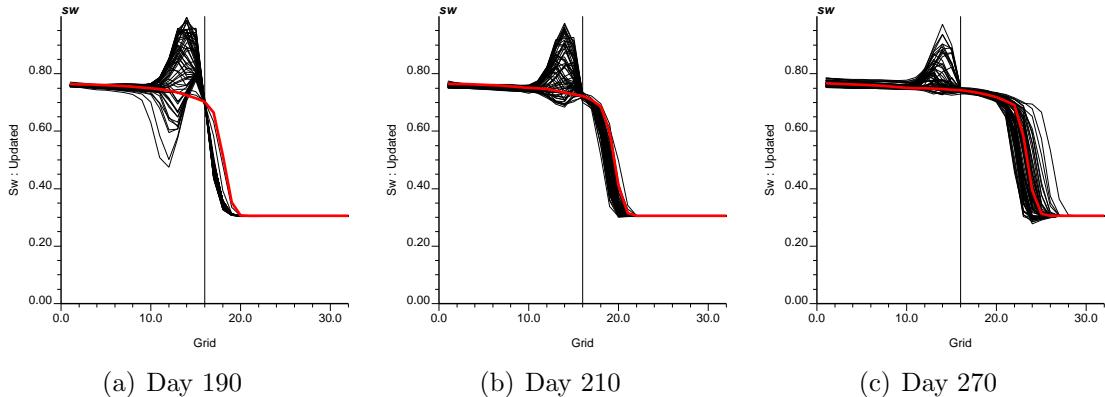
(b) After Kalman correction at day 230



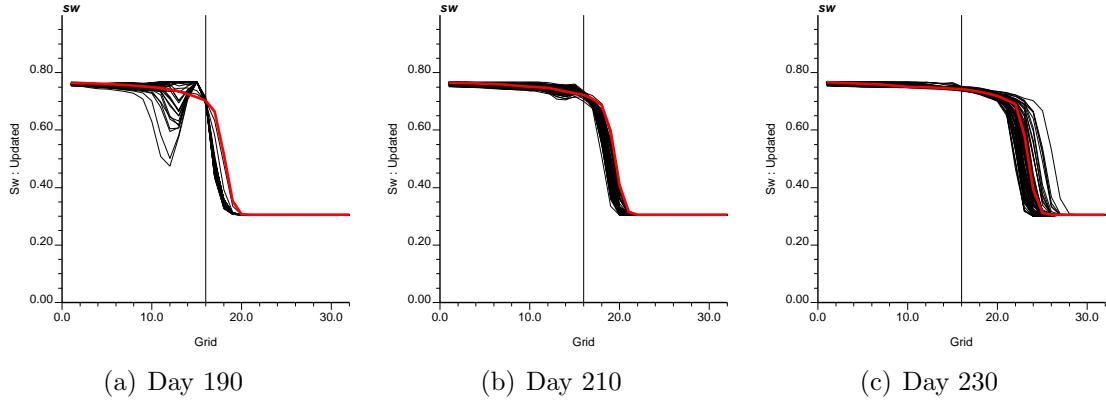
(c) After Kalman correction at day 250

(d) After Kalman correction at day 270

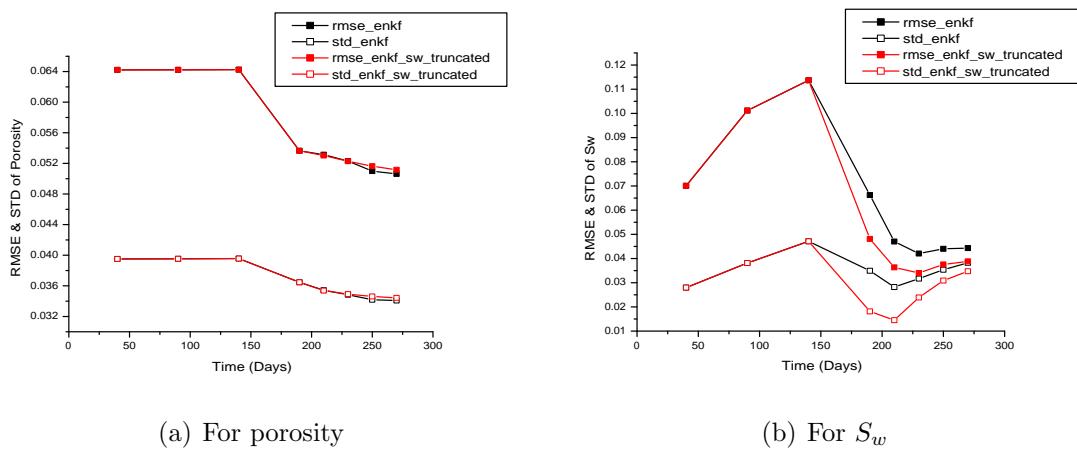
**Figure 5.24:** EnRMLF iterate when needed: the final  $S_w$  profiles at days 210, 230, 250, and 270 for Case I.



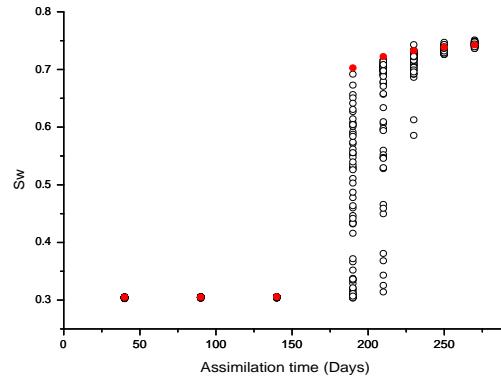
**Figure 5.25:** Traditional EnKF:  $S_w$  profiles at days 190, 210 and 270 for Case I.



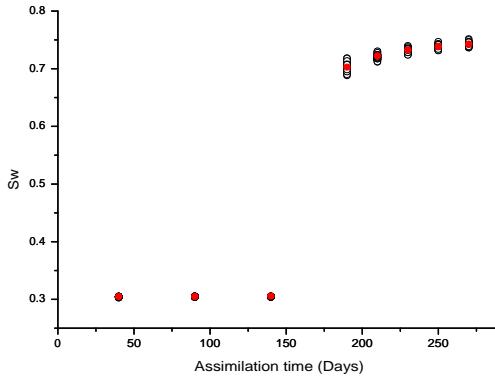
**Figure 5.26:** Traditional EnKF with truncation of  $S_w$  after Kalman correction:  $S_w$  profiles at day 190, 210 and 270 for Case I.



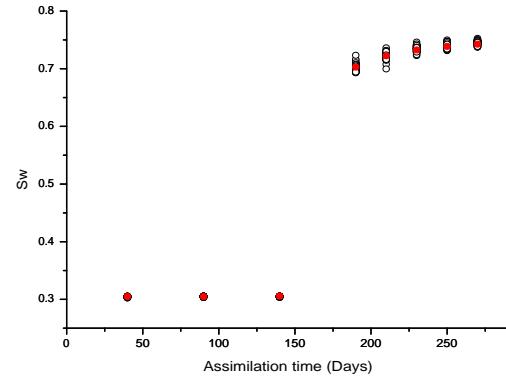
**Figure 5.27:** RMSE and spread for estimates of porosity and  $S_w$  for the traditional EnKF with and without truncation of  $S_w$  after Kalman corrections for Case I.



(a) Conforming EnKF

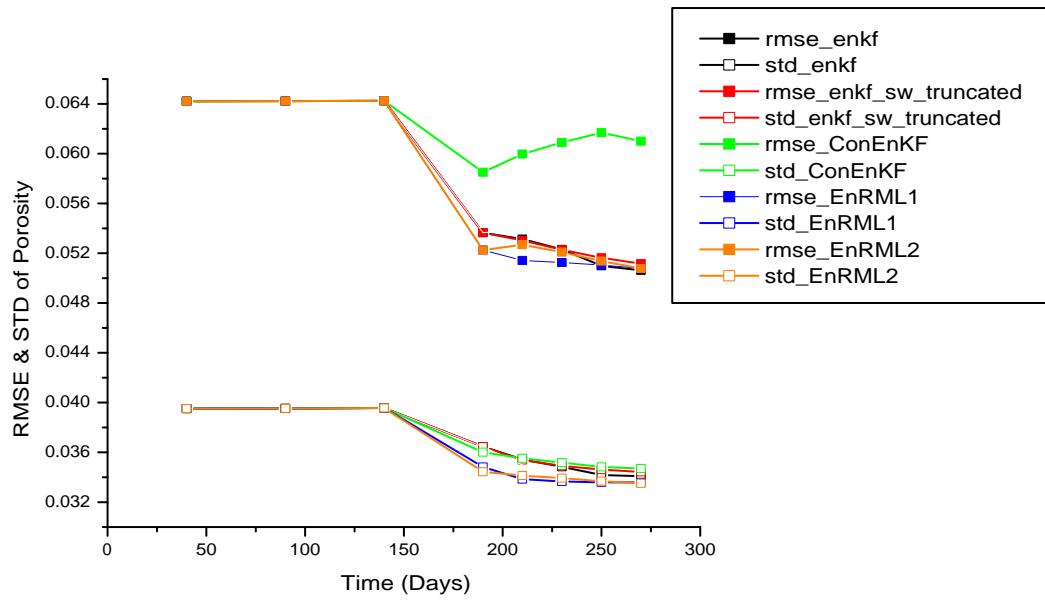


(b) EnRMLF with iteration at every measurement time

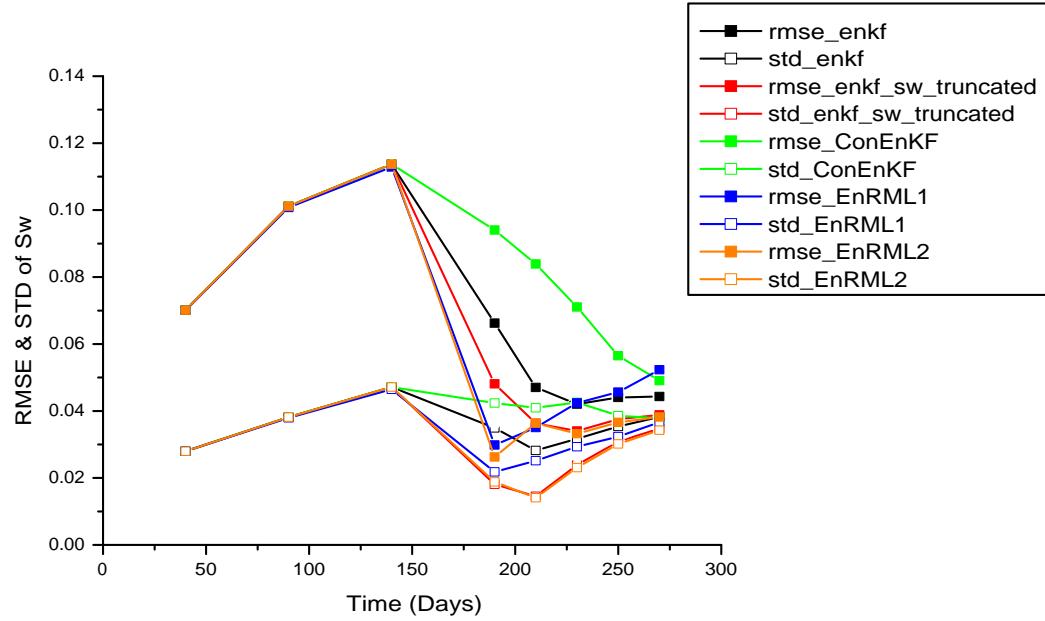


(c) EnRMLF with iteration when needed

**Figure 5.28:** Data match at the 8 different measurement times from the three iterative methods for Case I. The solid red dots on each of the plot are the synthetic data computed directly from the selected true reservoir model. The multiple hollow black dots at each measurement time are the final data from the ensemble models.

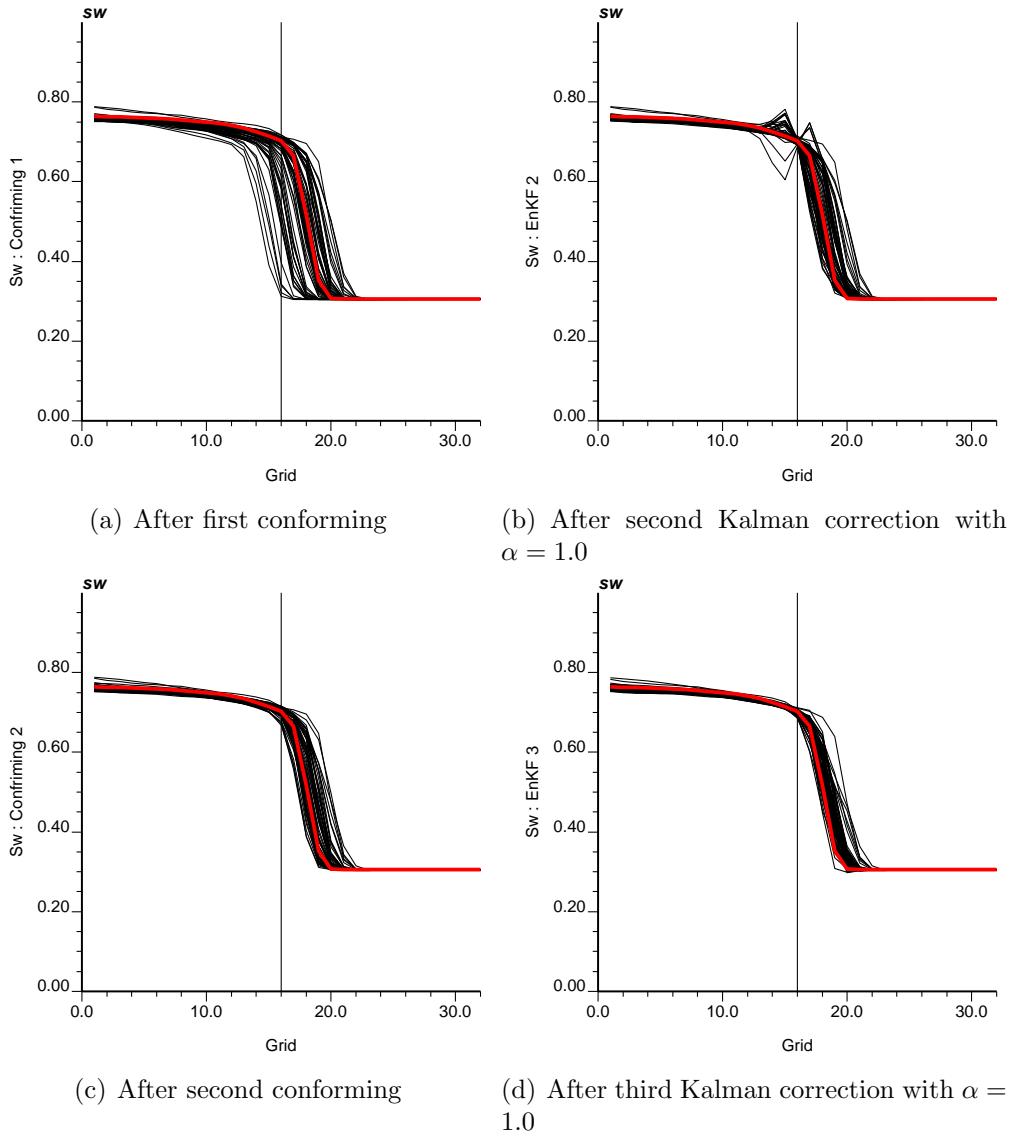


(a) For porosity

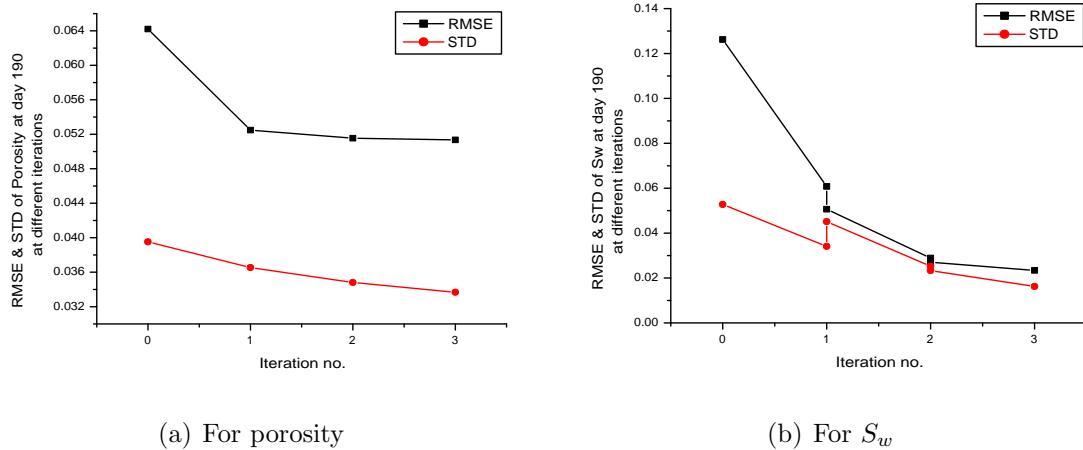


(b) For  $S_w$

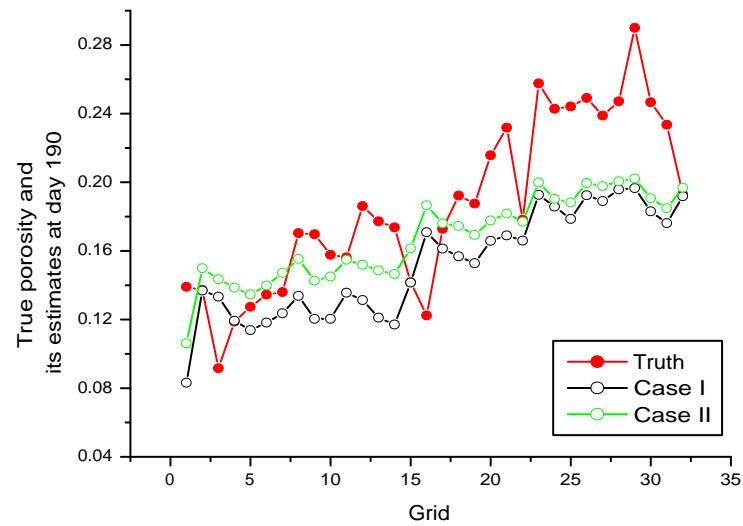
**Figure 5.29:** RMSE and spread of both estimates of porosity and  $S_w$  from various methods for Case I.



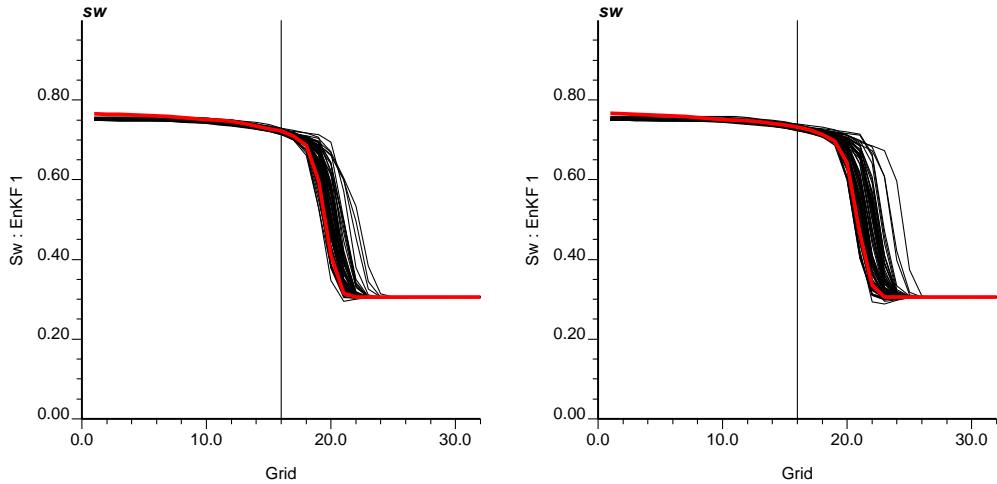
**Figure 5.30:** Conforming EnKF:  $S_w$  profiles at days 190 for Case II.



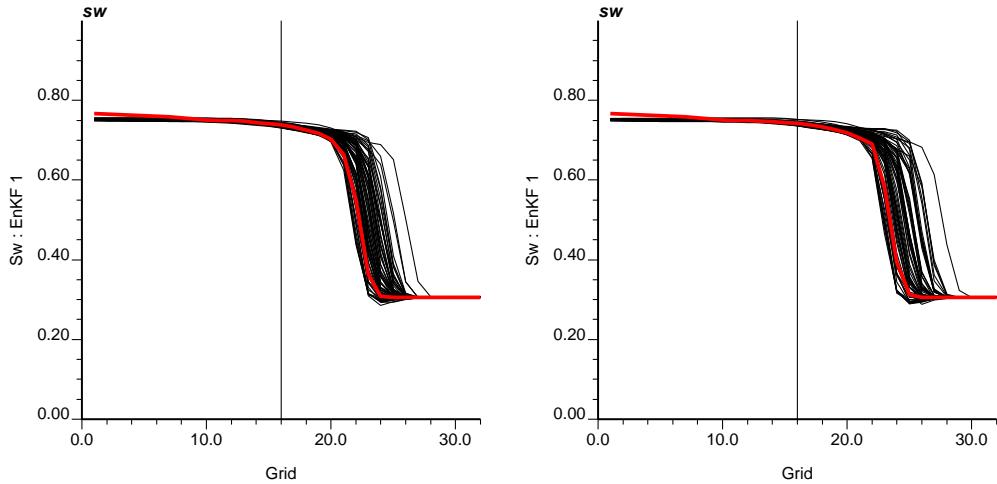
**Figure 5.31:** Conforming EnKF: RMSE and spread of porosity and  $S_w$  estimates at day 190 from different iterations for Case II.



**Figure 5.32:** The comparison of the porosity estimates at day 190 of Case I and II.

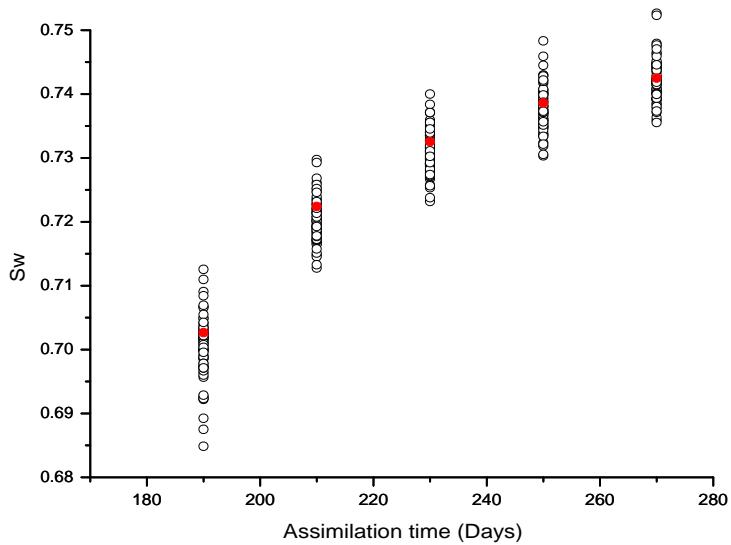


(a) After the Kalman correction at day 210 (b) After the Kalman correction at day 230

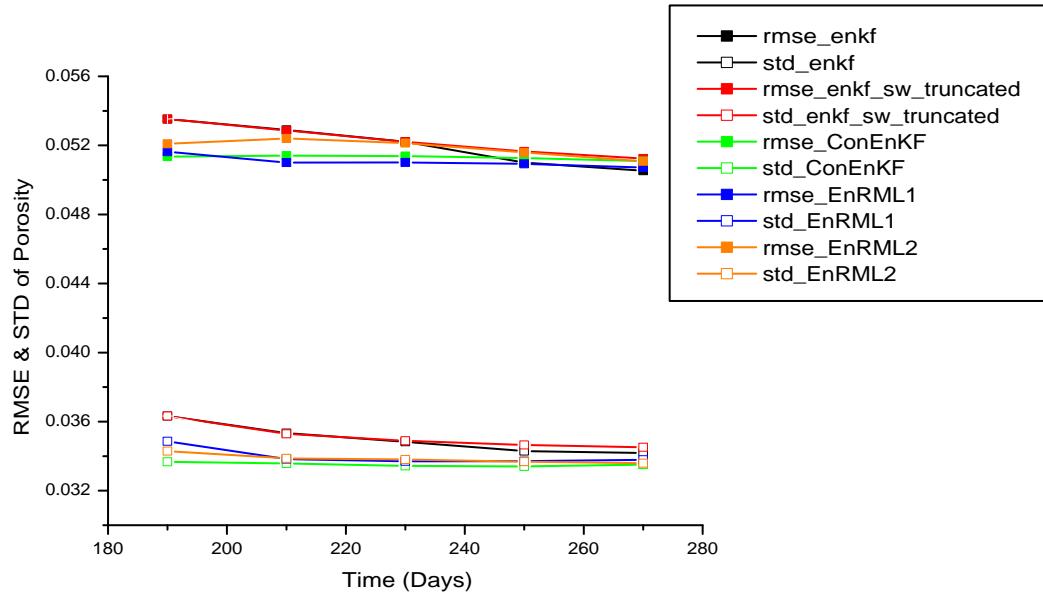


(c) After the Kalman correction at day 250 (d) After the Kalman correction at day 270

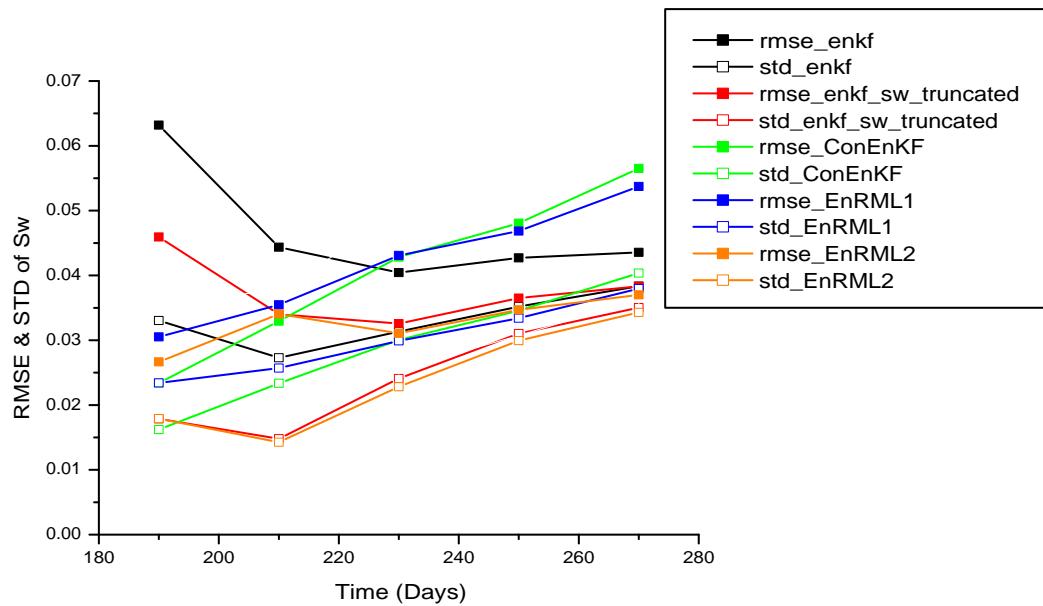
**Figure 5.33:** Conforming EnKF: the final  $S_w$  profiles at days 210, 230, 250, and 270 for Case II.



**Figure 5.34:** Conforming EnKF: data match at different measurement times for Case II. The solid red dots on each of the plot are the synthetic data computed directly from the selected true reservoir model. The multiple hollow black dots at each measurement time are the final data from the ensemble models.



(a) For porosity



(b) For  $S_w$

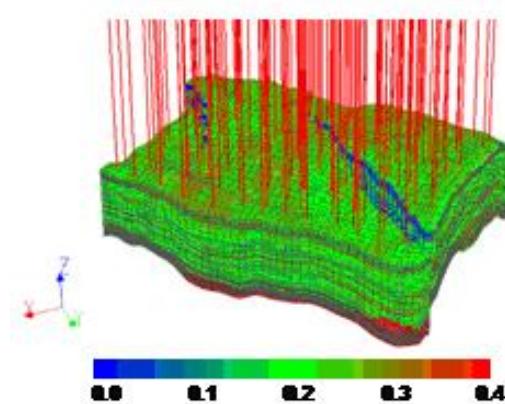
**Figure 5.35:** RMSE and spread of porosity and  $S_w$  estimates from various methods for Case II.

# CHAPTER VI

## A FIELD CASE STUDY WITH THE ENSEMBLE KALMAN FILTER

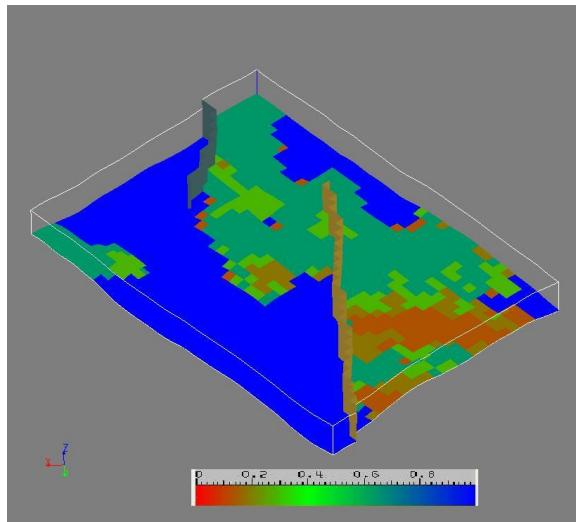
### 6.1 Field Description

The reservoir simulation model under study is a section extracted from a large sandstone reservoir field containing over 1.5 MMMSTB of oil. The sector model, called BBCK, has an average porosity of 0.2 (Fig. 6.1). The simulation model has  $30 \times 46 \times 39$ , totally 53,820 gridblocks. It has two faults, a bottom aquifer, and four different relative permeability zones. The field has been produced for approximately 50 years by primary depletion and phased water-flooding. Recovery to date is about 35% of OOIP (original oil in place) from 150 wells with a field-wide water cut of 93%. However, the sector model used in this study starts at Year 1965 (about 20 years after the first oil) and ends at Year 2001 with only 129 wells. This field was previously used for assisted (Milliken et al., 2000) and generalized travel-time (Cheng et al., 2004) history match studies.



**Figure 6.1:** Porosity distribution for the BBCK sector model.

The sector model has no water injection until Year 1993. So the bottom aquifer is the only water resource until then. Fig. 6.2 plots the initial water saturation at the bottom layer. The blue-colored areas are the aquifer. Most part of the aquifer is originally located at left side of the fault block. The distribution of the initial water will be illustrated later that it plays an important role in achieving successful history matching for most of the wells.



**Figure 6.2:** Initial water saturation at the bottom layer (layer 39) of the BBCK sector model (the blue-colored areas represent the aquifer).

There are two issues need to be mentioned regarding the simulation model.

- **Boundary conditions.** The sector model is a part of a complete field model and should have communications with the other parts of the whole model through its boundaries. However no-flow boundary conditions were assigned to the sector model. This assignment limits the peripheral water influx for wells close to the boundaries. As the only water resource available in the sector model is the bottom aquifer till the open of the injection wells at Year 1993, these wells have much lower computed water cut from the initial models than the recorded history (6D\_89, 7D\_11, 7D\_12, 7E\_41 and etc.), especially at early times.

- **Initial conditions.** The sector model starts simulation 20 years after the first oil. Although the reservoir fluid distribution at that time is probably different from the hydraulic equilibrium, the initial conditions in the sector model were set up assuming a uniform water/oil contact (WOC) depth.

Both the boundary and initial conditions in the sector model may greatly deviate from the unknown reality, thus complicating the history matching study for this case.

## 6.2 Production Data

### 6.2.1 Data pre-processing

The historical data available are oil and water **cumulative** production taken monthly. Based on them, the oil and water production rates within any time duration  $\Delta t$  can be calculated using

$$d_{r,obs}(t) = \frac{d_{c,obs}(t) - d_{c,obs}(t - \Delta t)}{\Delta t}$$

$$t = 1965 + \Delta t, 1965 + 2\Delta t, 1965 + 3\Delta t \dots, \quad (6.1)$$

where  $d_{r,obs}$  denotes the computed rate data and  $d_{c,obs}$  is the measured cumulative data. It can be seen from Eq. (6.1) that using different  $\Delta t$  gives different averaged rate data and results in different overall smoothness.

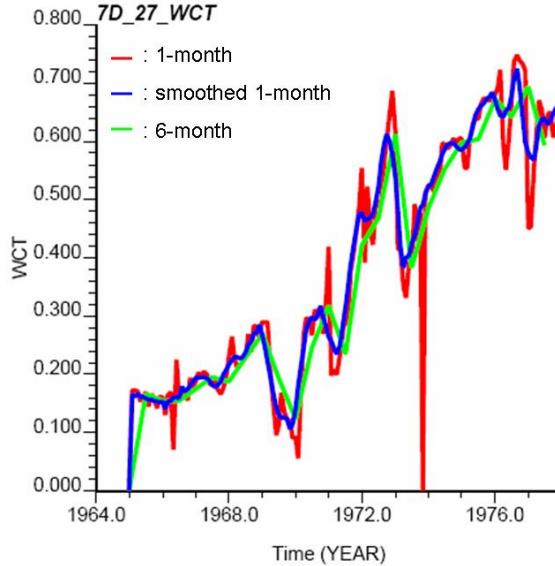
In this study,  $\Delta t = 1$  month and 6 months were used. 1-month data are rougher (Fig. 6.3). Some of the roughness in the data may be due to measurement noise and error and doesn't necessarily carry real information. 6-month data are much smoother than the 1-month data. They capture the major characteristics of the data variation. Using even bigger  $\Delta t$  can give more smoothness, but may mask some useful information in the data.

For field data with unknown noise, smoothing the data may be necessary to mitigate the noise level. Eq. 6.2 is used to smooth the 1-month data.

$$d_{r,obs}(t) = \frac{d_{c,obs}(t + \Delta t) - d_{c,obs}(t - \Delta t)}{2\Delta t}$$

$$t_i = 1965 + i/12, \text{ for } i = 1, 2, \dots, \quad (6.2)$$

where  $\Delta t$  is the half time window used to smooth the 1-month data, here  $\Delta t = 3$  months.



**Figure 6.3:** Water cut for well 7D\_27 computed with different data processing schemes.

Table 6.1 summarizes the above data processing schemes and their advantages and disadvantages.

### 6.2.2 Well constraint and computed data sensitivity from ensemble

In the simulation model, the wells are produced with total liquid voidage constraint. The historical data used are the oil and water production rates (other than the water cut data used in the previous studies. But it is essentially the same with matching the water cut data). The constraint, sometimes, results in small differences of computed rates from the ensemble models, especially at the early production time when not

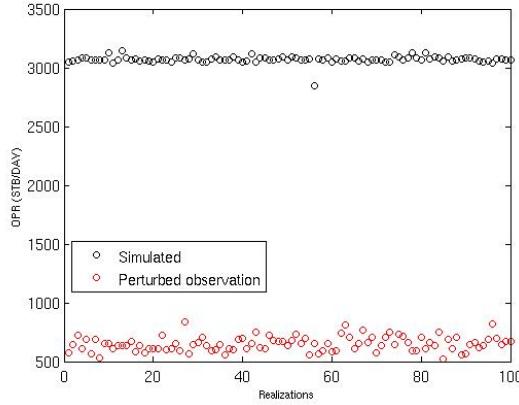
| Processing scheme                                     | Advantages                                      | Disadvantages                            |
|-------------------------------------------------------|-------------------------------------------------|------------------------------------------|
| 1-month interval                                      | capture detailed data changes                   | susceptible to measurement noise & error |
| 1-month interval smoothed w/ 3-month half time window | mitigate the level of data noise & error        |                                          |
| 6-month interval                                      | capture major characteristics of data variation | may mask useful information              |

**Table 6.1:** Data processing schemes used and their advantages and disadvantages.

much water comes out from the sector model. The models have to produce more oil to substitute the water production to satisfy the well constraint. Fig. 6.4 shows the perturbed measurements and computed oil production rates, from 100 models, for well 6D\_89 at Year 1965.5. It can be seen that the computed oil rates from the ensemble models differ slightly and are much more than the observed oil rate because the limited water resource specified in the simulation models can not sustain much water producing out of them at this time. The small difference in the computed data from the ensemble will result in small sensitivity of data to model parameters (the extreme scenario will be that all the ensemble models predict the same data, then no sensitivity can be computed from this). Presumably, bottom-hole pressure data can provide more information for this kind of well constraint. Unfortunately, they are not available.

### 6.2.3 Data selection criterion

As seen in Fig. 6.4, the computed data from the ensemble have small differences and yet are far from the perturbed observations. If we still try to match this piece of datum, we may end up with unrealistic reservoir models and introduce problems to subsequent assimilation times. As sequential data assimilation methods require reasonable adjustment at each measurement time, we need to select carefully what data to match prior to data assimilation at each time. In this application, datum



**Figure 6.4:** perturbed measurements and computed oil production rates, from 100 models, for well 6D\_89 at Year 1965.5.

that is far from its computed values is discarded according to the following criterion

$$h > n(\sigma_{\text{sim}} + \sigma_{\text{obs}}), \quad (6.3)$$

where  $h$  is the distance between a single observed and simulated data;  $\sigma_{\text{sim}}$  is the standard deviation of the simulated data;  $\sigma_{\text{obs}}$  is the standard deviation of the perturbed measurements; and  $n$  is an input threshold parameter,  $n = 4.0$  is used in this application.

By so doing, only datum whose observed and computed values are within reasonable range is used for data assimilation at measurement times. It is possible that data that are abandoned contain useful information and may improve the estimation if used, but screening out data prior to assimilation is a compromise that was necessary in this real case study because of the characteristics of the computed data from the ensemble models illustrated by Fig. 6.4.

### 6.3 Model Parameters

The permeability in the  $x$  direction of each gridblock is altered in the history-matching process to match the historical data. The permeabilities in the  $y$  and  $z$  directions are computed using deterministic ratios to the permeability in the  $x$  direction. The

porosity field is not changed because its spatial variation is relatively minor compared to that of the permeability's.

There is one model originally constructed based on a single upscaled model from the static fine-scale geostatistical model using flow-based upscaling method. This model has been roughly history matched using the assisted history matching method (Milliken et al., 2000). Based on this model, the variogram of the  $x$  permeability is computed. Spatially correlated Gaussian random errors with zero mean, and calculated covariance function are added to the single initial model to generate the initial ensemble reservoir simulation models for the EnKF.

## 6.4 Result Analysis

This study is composed of two stages: history matching phase and prediction phase. The first 13-year history, Year 1965 to 1978, was used to adjust the model parameters. The update models at the final time were used to predict the following 23-year reservoir performances. There are totally 25 wells whose data were used.

Besides the traditional EnKF method, The Conforming EnKF was also implemented and used with the data processed using 6-month interval. Originally for all the investigations, the size of the ensemble is 100. We also tried double-sized ensemble, 200, on the 6-month interval data to see if the bigger ensemble is able to provide better results.

### 6.4.1 History matching results

Among all the 25 wells, 17 wells achieve satisfactory history matching results, 4 wells demonstrate the effects by reservoir boundary and fault, and other 4 wells do not have substantial improvement. Figs. 6.5(a) – 6.5(c) show the well locations of the three categories. For illustration purposes, a few typical wells are picked from each category and shown from Fig. 6.6 to Fig. 6.10. In these figures, the red lines on different plots are the observation calculated with different data processing scheme;

the multiple black lines are results from the ensemble models; the green lines are the mean of the ensemble.

Fig. 6.6 compares the computed water cut from initial models to the history-matched counterparts with different production data processing schemes at well 7D\_27. This well is located close to one of the reservoir boundaries (see Fig. 6.5(a)). At the very beginning of the simulation time, about 18% water cut was immediately observed. As mentioned in Section 6.1, most of the water flux at that time may come from outside of the sector model. The initial models could not produce as much as the measured water cut until 4 years later (Fig. 6.6(a)). However, after data assimilation, it only takes about 2.5 years for the updated models to catch up with the measured water production (See Figs. 6.6(b) - 6.6(f)). Among all the data processing schemes, 6-month data averaging scheme achieves the best match (Figs. 6.6(d) and 6.6(f)). The 100 models are adequate and Conforming does not help at this well. The smoothed 1-month scheme performs fairly differently from that of the 1-month's, but is better in whole (compare Fig. 6.6(c) to 6.6(b)). Marked by the two rectangles on Fig. 6.6(a), the significant draw-downs on the observed water cut (red line) are not reflected by the simulated data from the initial models (black lines). This may indicate that (1) the observations at the two time spots were mistaken, or (2) the re-completion (shut down of the water layers) was not recorded in the simulation deck, or (3) the water layers in the simulation models do not agree with these layers in the reality due to the difference between the simulated and real water movement, so when the water layers are shut down in the real world, there is no reduction on the computed water cut. While 1-month data keep the detailed information, most likely, it is not easy to match them by adjusting the permeability field only. However, for the 6-month and smoothed 1-month data (red lines on Figs. 6.6(d) and 6.6(c)), the two draw-downs are mitigated.

At well 8D\_21, Fig. 6.7, the initial models have the opposite behavior than well

7D\_27: more water is produced from the ensemble models than the recorded history. Similarly as above, 6-month data are matched the best and the Conforming EnKF with iterations does not help at this well either. Smoothed 1-month scheme captures the major trend of the historical data and is better than the non-smoothed 1-month scheme. Similarly with what has been seen at well 7D\_27, the rectangle on Fig. 6.7(a) marks the inconsistency between the substantial water cut draw-down on the observed water cut (red line) and the small drop on the computed water cuts from the initial models. This is also where the 1-month scheme begins to go astray from the historical data.

The observations from the above two wells (7D\_27 and 8D\_21) are applicable to the other wells. From here forward, we will show only the history matched results based on the data processed with the 6-month averaging scheme along with the computed data from the initial models.

Both wells 7D\_11 and 7D\_44 are located on the right side of the longer fault, where initially there is no direct water support from the bottom aquifer at this fault block (see Fig. 6.5(b)). Water on the left side of the fault block has to travel across the fault to get into these wells. Hence fault transmissibility plays an essential role in the water cut at these wells. The observed water cut is much higher than the simulated values from the initial models (Figs. 6.8(a) and 6.9(a)). However, with fixed fault transmissibility in this study, changing  $x$  permeability is not able to bring more water at these two wells (Fig. 6.8(b) and 6.9(b)).

7D\_74 is among one of the 4 wells where history matching does not achieve significant improvement. The initial models produce almost 100% water cut from the very time the well is operated and much more than the historical data (Fig. 6.10(a)). After adjustment, the updated models produce a little less than the initial models but are still much higher than the history (Fig. 6.10(b)).

### 6.4.2 Prediction results

At the end of the history matching at Year 1978, the 100 updated models based on the 6-month interval data predicted the water cut till Year 2001. Figs. 6.11 – 6.14 compare the water cut from the initial models to the predictions from the updated models and the history-matching results from the generalized travel time (Cheng et al., 2004). The legends used here are the same with the above figures. The blue-colored lines are the generalized travel-time (GTT) results.

Only by visual inspection of the results from the 25 wells (whose data are used in the EnKF), there are 8 wells where the GTT matched better than the EnKF predicted, 4 wells EnKF better than GTT, and 13 others comparable from both methods. Fig. 6.12 shows one of the example wells, 7D\_44, where GTT matched better than EnKF predicted. Remember, this well is located close to the longer fault at the right fault block where aquifer is not initially available. Most of the models in the EnKF could not produce enough water, but GTT gave a pretty good match. Interestingly, GTT also matched the two wells located at the boundary of the right fault block (7D\_11 and 7D\_12), while shown in Section 6.4.1, EnKF could not move water across the fault and to the right block (Fig. 6.9(b)).

On the other hand, 7E\_41 (Fig. 6.11), is among one of the wells where EnKF obtained better results. The observed water cut lies in the envelope covered by the predictions from the updated models, while the GTT's match is still further from the observation. Figs. 6.13 and 6.14 show two wells where both had good results and no substantial improvement, respectively.

It is worth noting that the EnKF only matched the first 13-year production history with 25 wells and predicted the following 23-year reservoir performances, while the GTT matched the entire 36-year historical data of 130 wells. Some of the wells used in the EnKF opened late and only a few data at each of these wells were used. In the history matching phase, we observed that some of the data are not matched at

early times but caught up later by the computed data (Fig. 6.6). So it is tempting to assume that when more data are used, some of the wells which do not have substantial improvement in the first 13 years, may be able to behave better.

Fig. 6.15 compares the field wide water cut performances from the initial models to the updated models. After data assimilation, the updated models predict satisfactory reservoir behavior with smaller variation in the predicted values.

#### 6.4.3 Evolution of estimated permeability field

Fig. 6.16 shows the evolution of the mean  $\log k_x$  estimate at layer 34. Some parts of the fault block on the left side at this layer is below the WOC depth. From Year 1968 to 1978, the permeability on both sides of the longer fault and the two lower boundaries is increased, this adjustment is consistent with what has been observed from the well history-match. For example, wells 7D\_27, 7D\_11 and 7D\_44, see Figs. 6.6, 6.8, and 6.9 respectively, need to increase the water cut. By increasing the permeability in these areas, the water from the bottom aquifer flows faster to these wells. The rest of the areas have their permeability decreased, which is also consistent with the changes of the water cut at the wells located in the same region, see Fig. 6.7. The general heterogeneity in the initial model at this layer is still preserved after the adjustment at different times.

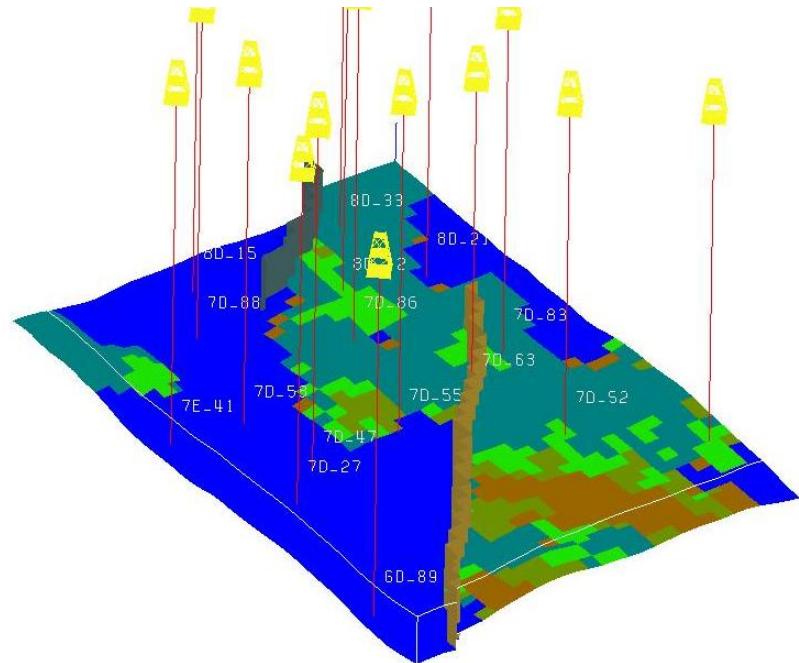
Comparatively, layers 8 and 3 (see Figs. 6.17 and 6.18) do not share the same permeability changing pattern with layer 34. The main reason is layer 34 has direct contact with the bottom aquifer, so the permeability adjustment at that layer is more sensitivity to the data than that of shallow layers.

Fig. 6.19 shows the gridblock STD maps of the ensemble  $\log k_x$  estimates at three different times at layer 34. For all the three times, the lower two boundaries had smaller STD. At both the two upper corners, the STD is relatively higher due to the lack of information at these regions.

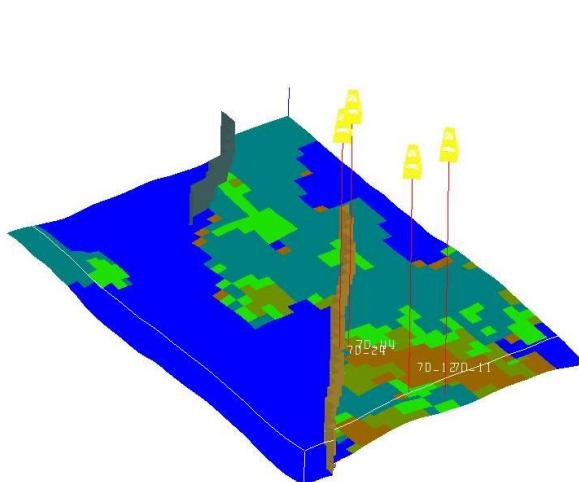
## 6.5 Discussion

The EnKF was applied to a real case history matching study. Some conclusions drawn from the study are:

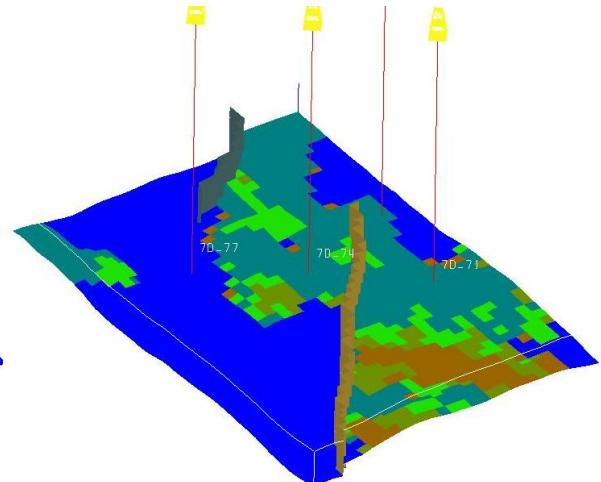
- The history-matching results based on the 6-month interval data outperforms the results based on all other data processing schemes. 100 initial models are sufficient for this study after being compared with the results from 200 models.
- The Conforming EnKF in this case does not improve the history-matching quality.
- Non-smoothed 1-month data are susceptible to measurement noise and error. Thus, most of the wells do not achieve good the history-matching results based on these data. However, the results based on the smoothed data with 1-month interval are much better. This is partly because smoothing mitigates the noise level in the measured data.
- The wells located at the left side of the fault block are generally matched better than the wells at the other side of the block, because they are more easily accessible to water from the bottom aquifer.
- Most of the wells located at the reservoir boundaries observe high water cut immediately after they are opened. Some of the water at that time may come from outside the sector model, thus the water cut data at the earlier times are not matched but the later data are caught up by the match.
- The permeability changes at deep layers are more sensitive to the data than that of the shallow layers, because they are closer to the bottom aquifer. The adjustment for permeability at the deep layers is consistent with what water cut dictates.



(a) Wells achieve good match

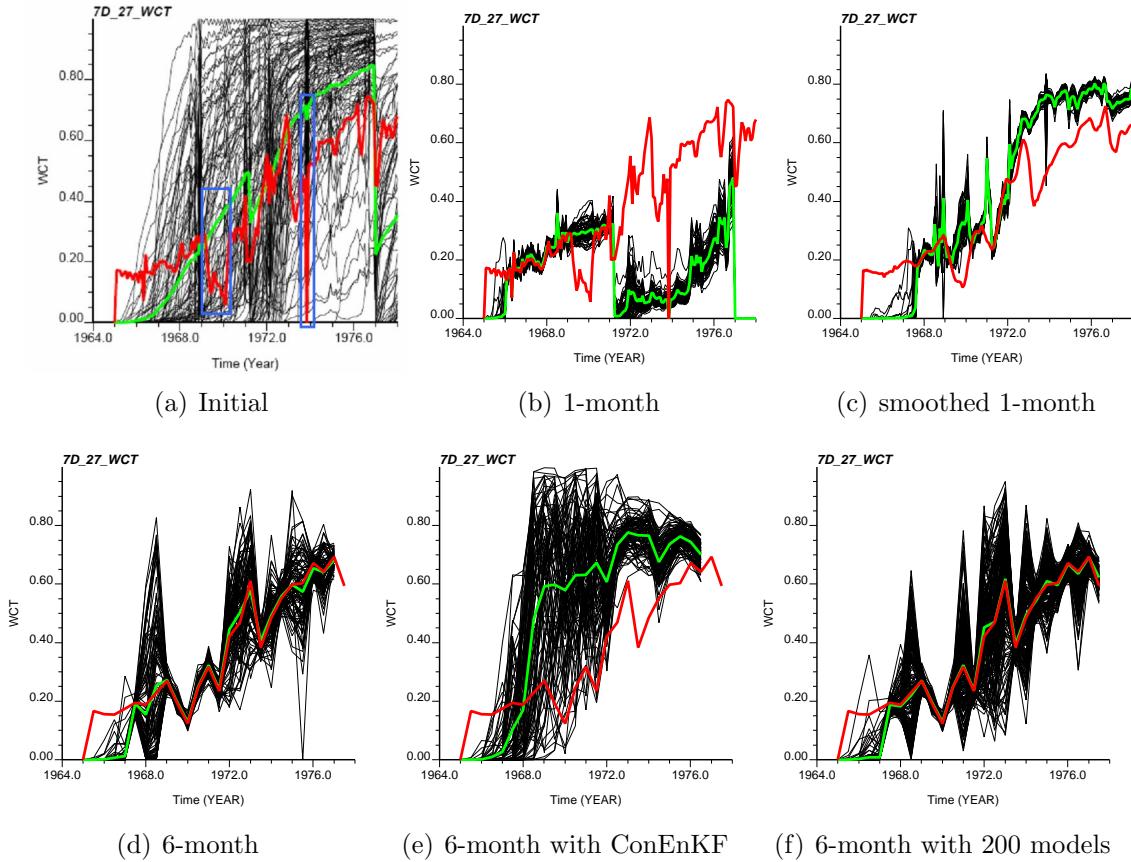


(b) Wells affected by reservoir boundary and fault

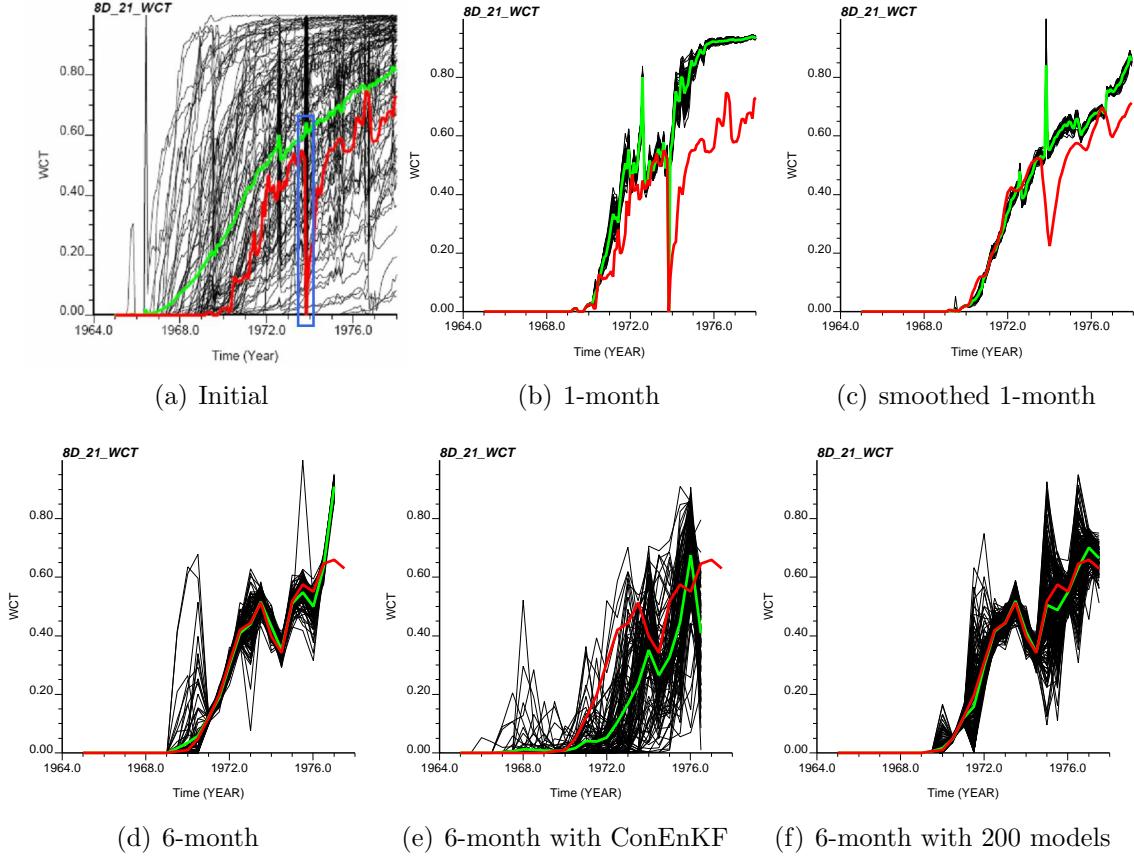


(c) Wells do not have improvement

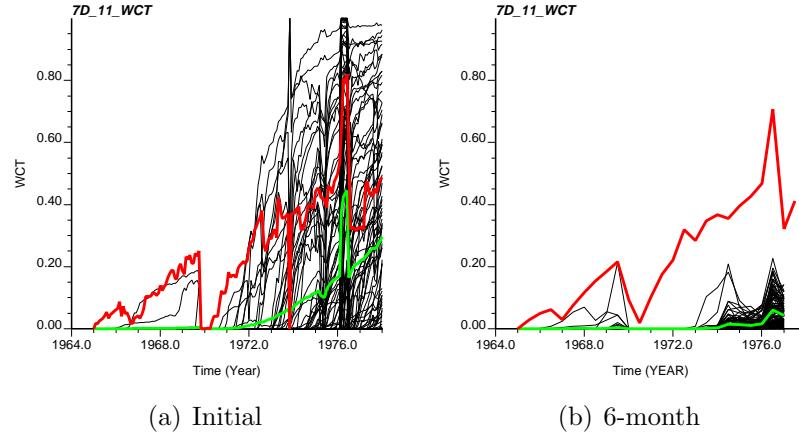
**Figure 6.5:** Wells achieving different history matching results projected on the bottom layer with initial water saturation shown. The blue color represents the aquifer.



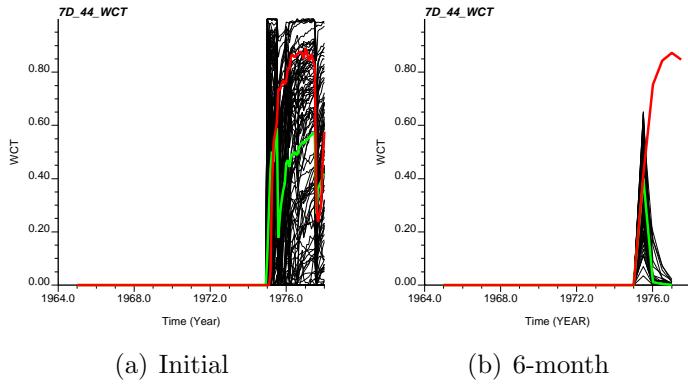
**Figure 6.6:** Water cut at well 7D\_27. Similar behavior wells include: 6D\_89, 7C\_58, 7E\_41, 8D\_24, and 8D\_33. They are all located close to reservoir boundaries and directly above the bottom aquifer which provides resource for water production (see Fig. 6.5(a)).



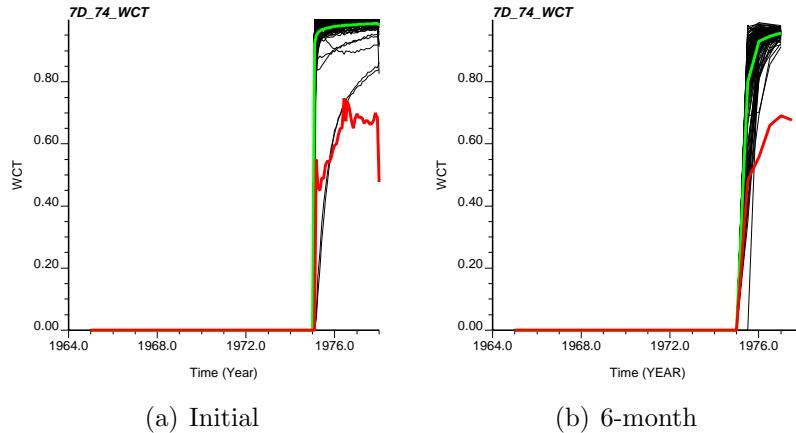
**Figure 6.7:** Water cut at well 8D\_21. Similar behavior wells include: 7D\_47, 7D\_55, 7D\_58, 7D\_83, 7D\_88, 8D\_12, and 8D\_15. They are all located directly above the bottom aquifer and water resource is guaranteed (see Fig. 6.5(a)).



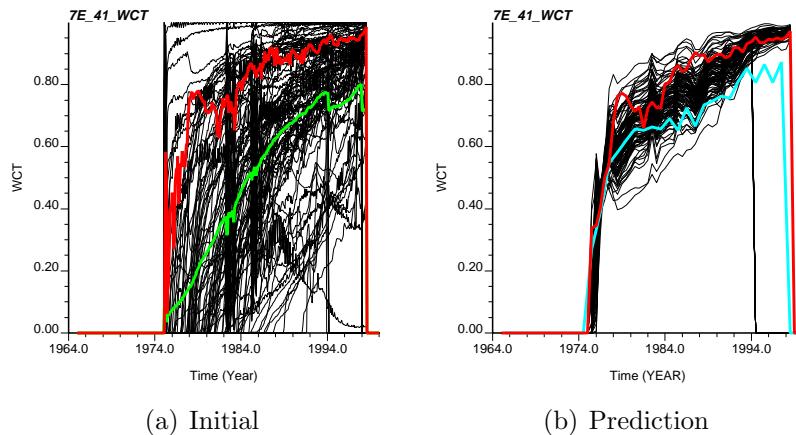
**Figure 6.8:** Water cut at well 7D\_11. Similar behavior well include: 7D\_12. They are all located close to one of the reservoir boundaries and not much direct aquifer support for this fault block at early times (see Fig. 6.5(b)).



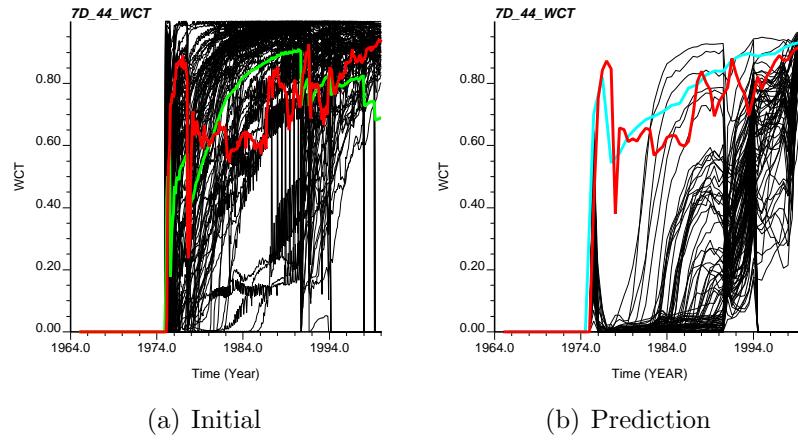
**Figure 6.9:** Water cut at well 7D\_44. Similar behavior well include: 7D\_24. They are all located close to right side the longer fault and not much direct aquifer support for this fault block at early times (see Fig. 6.5(b)).



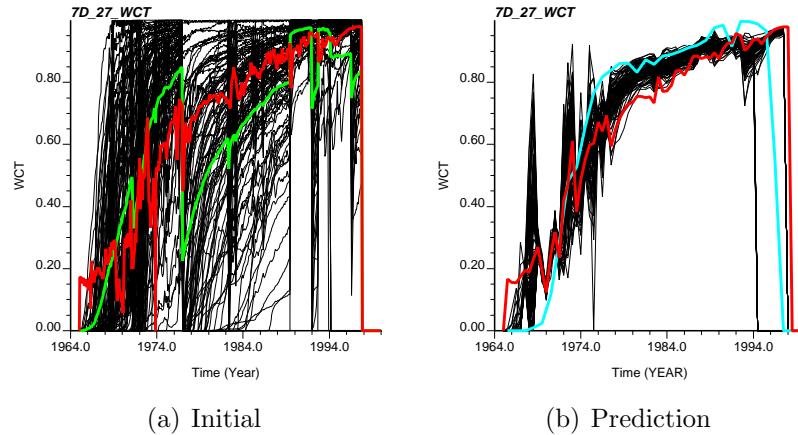
**Figure 6.10:** Water cut at well 7D\_74. Similar behavior wells include: 7D\_71, 7D\_77, and 8C\_18 (see Fig. 6.5(c)).



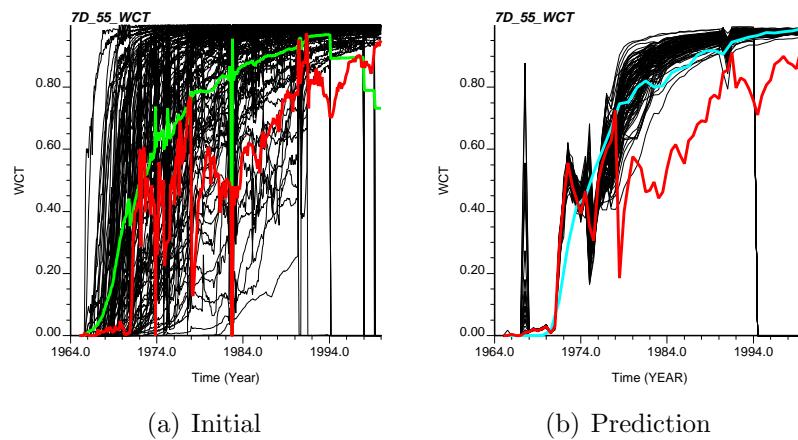
**Figure 6.11:** Initial and predicted water cut at well 7E\_41. The blue-colored line on the second plot is the history-matching result obtained by generalized travel-time.



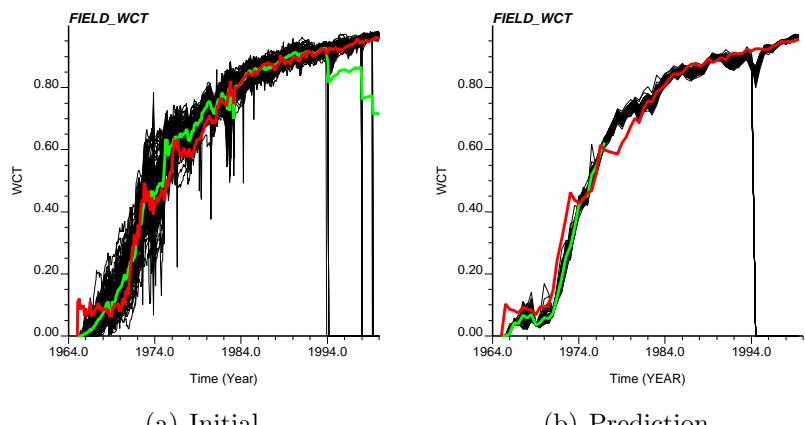
**Figure 6.12:** Initial and predicted water cut at well 7D\_44. Same legends are used as Fig. 6.11.



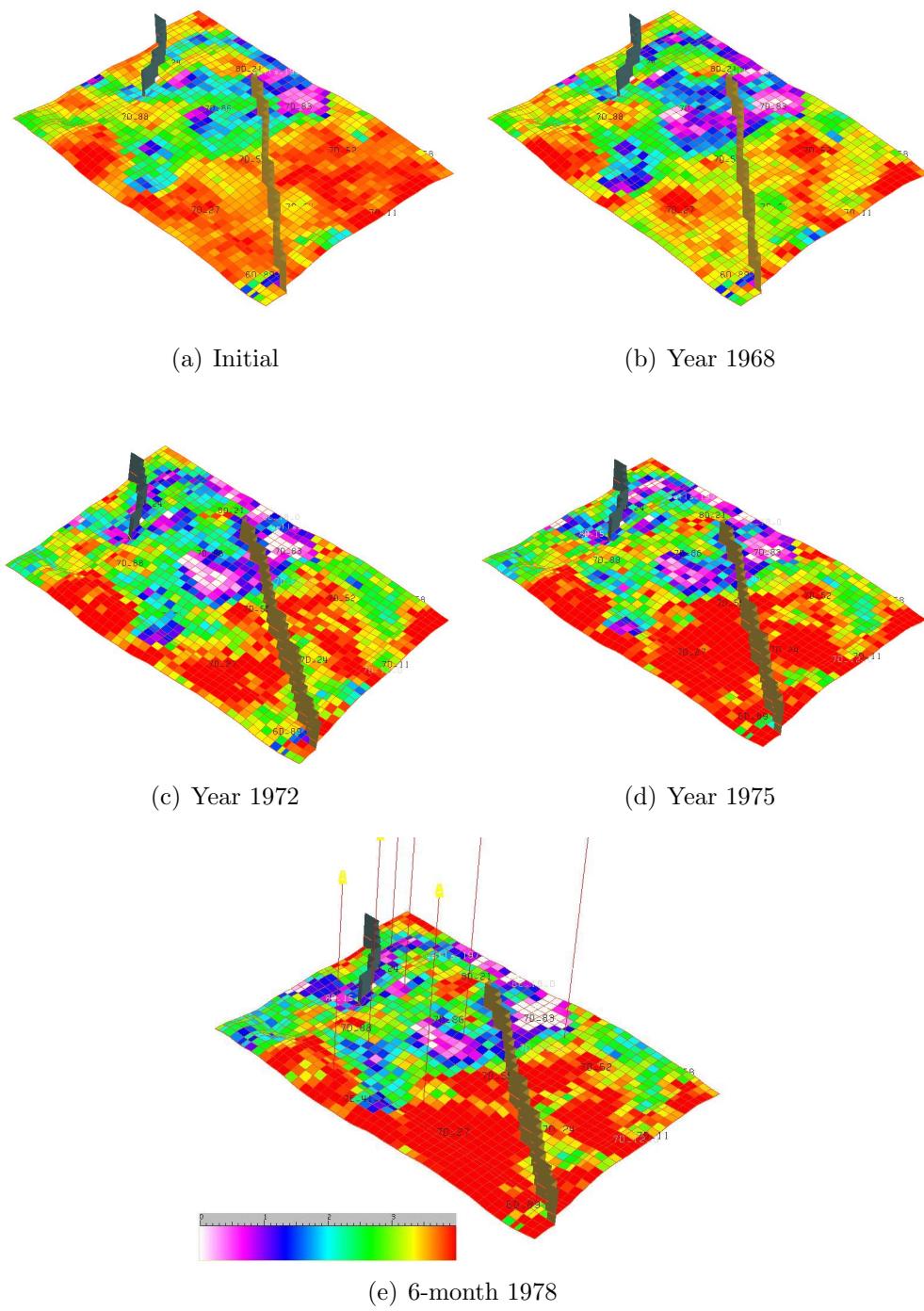
**Figure 6.13:** Initial and predicted water cut at well 7D\_27. Same legends are used as Fig. 6.11.



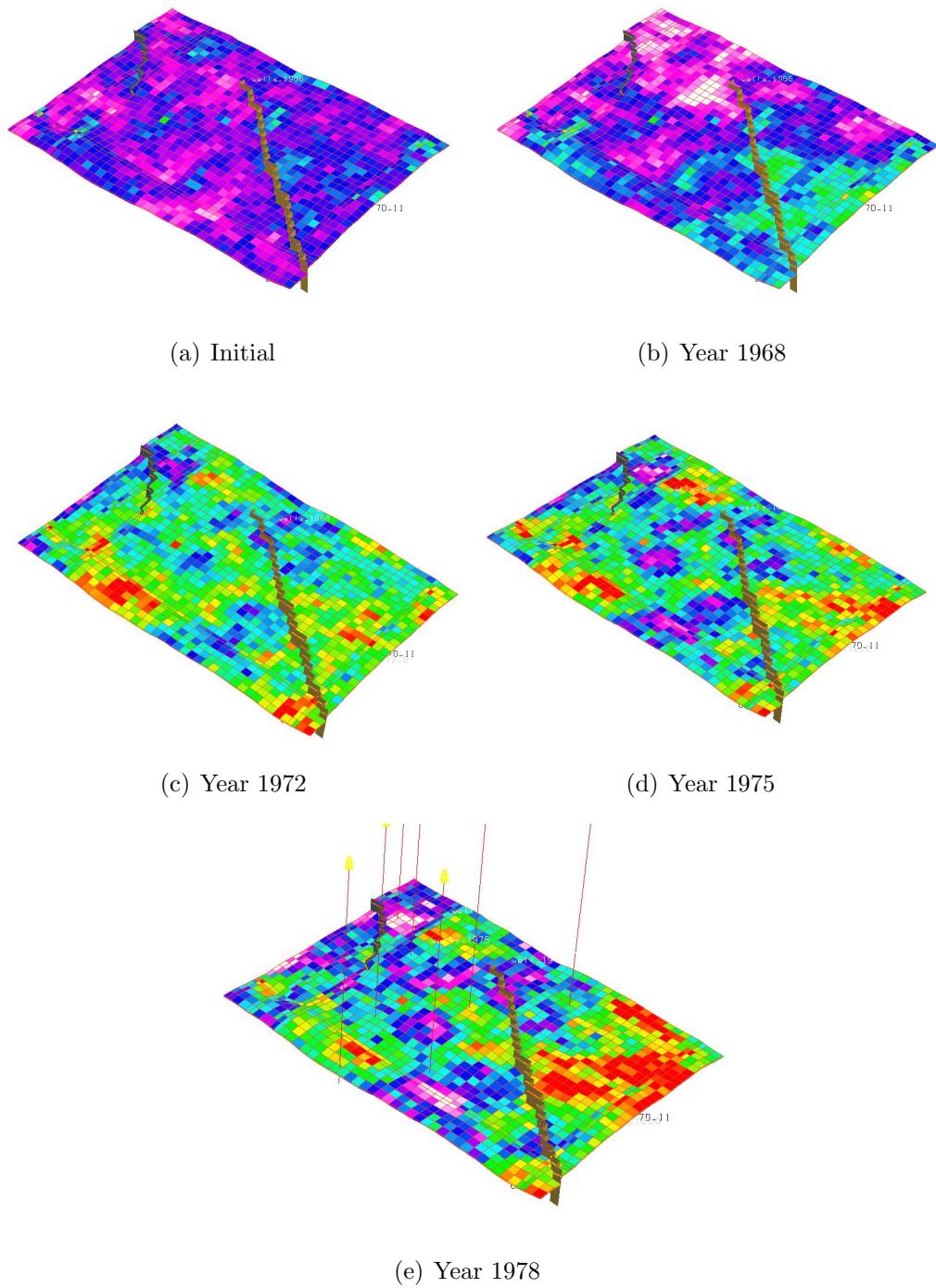
**Figure 6.14:** Initial and predicted water cut at well 7D\_55. Same legends are used as Fig. 6.11.



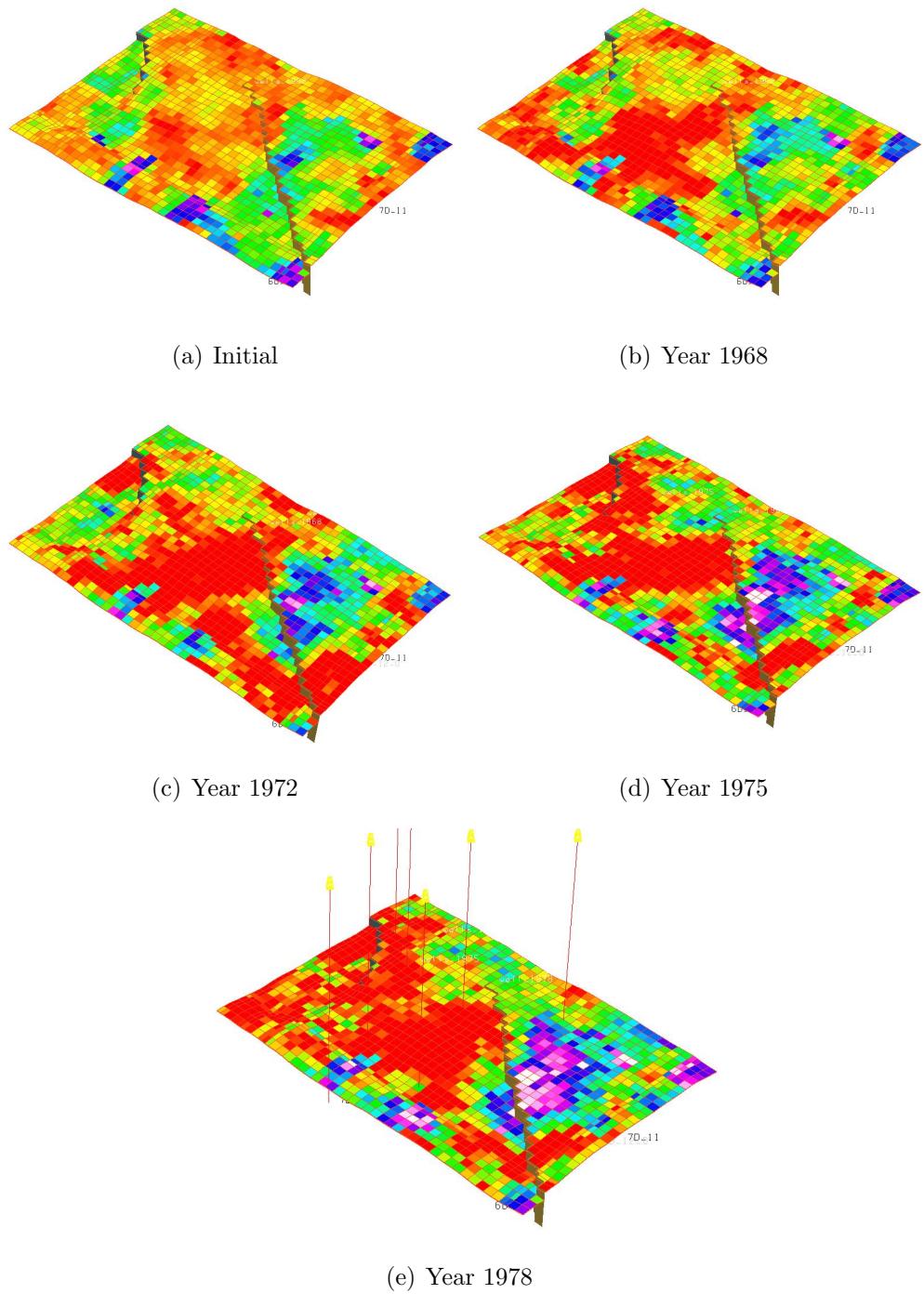
**Figure 6.15:** Initial and predicted field wide water cut.



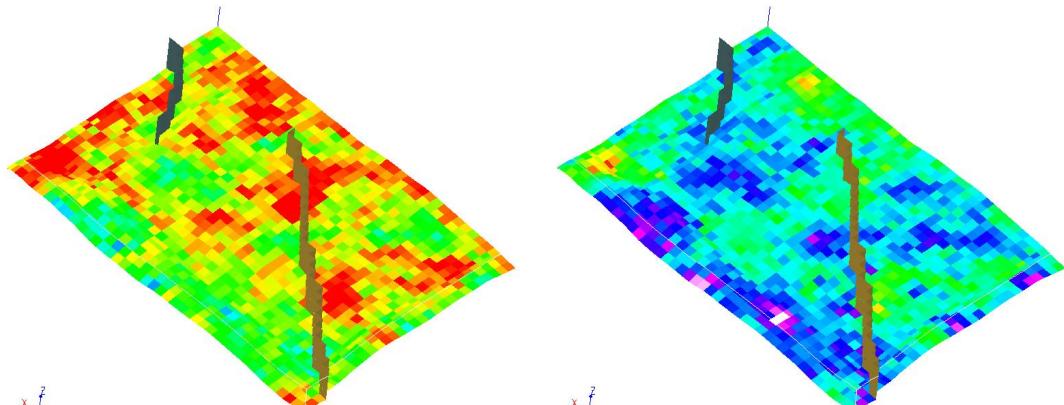
**Figure 6.16:** Evolution of the mean  $\log k_x$  estimate at layer 34 based on the 6-month interval data.



**Figure 6.17:** Evolution of the mean  $\log k_x$  estimate at layer 8 based on the 6-month interval data.

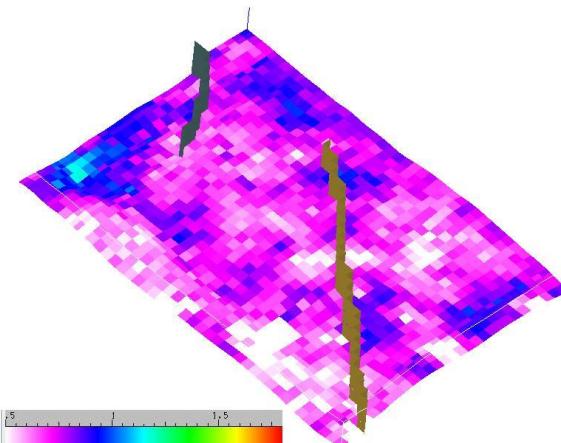


**Figure 6.18:** Evolution of the mean  $\log k_x$  estimate at layer 3 based on the 6-month interval data.



(a) Year 1972

(b) Year 1975



(c) Year 1978

**Figure 6.19:** Gridlock STD for the  $\log k_x$  estimates at layer 34.

# CHAPTER VII

## CONCLUSIONS

In this dissertation, the ensemble Kalman filter (EnKF) was introduced as an alternative method to traditional history matching approaches and applied to several reservoir applications. The EnKF takes in data sequentially whenever they become available and its framework is compatible with real-time reservoir monitoring with data from permeant down-hole gauges. As a Monte Carlo type of method, the correlations between model variables and theoretical data can be estimated from the ensemble models directly. By so doing, the EnKF avoids the complex calculation of the adjoint system for the forward problem, which may be required by efficient gradient based optimization methods. Moreover, the EnKF itself can be coded as a generic library, which enables transfer among different simulators in a very efficient way. After the observation history being matched, the EnKF outputs a collection of updated simulation models and they can be used for uncertainty analysis.

Through a few synthetic case studies, we found that the EnKF is suitable for data from time series when the changes made to the model parameters and state variables are both small at every measurement time. However, when the changes in the variables are large, the EnKF may provide invalid solutions.

Wen and Chen (2005a,b) suggested an intuitive remedy for the EnKF by adding a “conforming step” after the update step at each measurement time. The conforming step applies constraints on the state variables by re-initializing the system governing equations at the previous measurement time with the newly updated model parameters, conditional to data up to the current measurement time, and the state variables, conditional to data up to the previous measurement time. We showed through both

linear and non-linear examples that the results from the conforming method are incorrect. We proposed a new iterative scheme, called ensemble randomized maximum likelihood filter (EnRMLF) to handle the strong non-linearity at some measurement times. The new method was proven to be robust and work better than the EnKF.

The EnKF and Conforming EnKF of Wen and Chen was applied to a real data set from an oil field in east Asia. The first data available for assimilation is 20 years after the field was first produced. Due to the limited assessment of the reservoir conditions in the middle of production, there is a large uncertainty associated with the assumed reservoir initial conditions. However, without taking the uncertainty into consideration, we found that the history-matching results were affected by the simplified initial conditions.

## Bibliography

- Anderson, B., *Optimal Filtering*, Pergamon-Hall, Inc., England Cliffs, N.J., 1979.
- Anderson, J. L. and S. L. Anderson, A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, *Monthly Weather Review*, **127**(12), 2741–2758, 1999.
- Aziz, K. and A. Settari, *Petroleum Reservoir Simulation*, Elsevier Applied Science Publishers, London, 1979.
- Barker, J. W., M. Cuypers, and L. Holden, Quantifying uncertainty in production forecasts: Another look at the PUNQ-S3 problem, *SPE Journal*, **6**(4), 433–441, 2001.
- Bertino, L., G. Evensen, and H. Wackernagel, Sequential data assimilation techniques in oceanography, *International Statistical Review*, **71**(2), 223–241, 2003.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, Analysis scheme in the ensemble kalman filter, *Monthly Weather Review*, **126**(6), 1719–1724, 1998.
- Chen, Y. and D. Zhang, Data assimilation for transient flow in geological formations vis ensemble Kalman filter, *Advances in Water Resources*, **29**, 1107–1112, 2006.
- Cheng, H., X.-H. Wen, W. J. Milliken, and A. Datta-Gupta, Field experiences with assisted and automatic history matching using streamline models, SPE 89857, in *Proceedings of the 2004 SPE Annual Technical Conference and Exhibition*, 2004.
- Corser, G. P., J. E. Harmse, B. A. Corser, M. W. Weiss, and G. L. Whitflow, Field test results for a real-time intelligent drilling monitor, SPE-59227, in *Proceedings of the 2000 IADC/SPE Drilling Conference*, 2000.
- Deutsch, C. V. and A. G. Journel, *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 1998.
- Dong, Y., Y. Gu, and D. S. Oliver, Sequential assimilation of 4d seismic data for reservoir description using the ensemble Kalman filter, *Journal of Petroleum Science and Engineering*, **53**(1–2), 83–99, 2006.
- Eisenmann, P., M.-T. Gounot, B. Juchereau, and S. J. Whittaker, Improved rxo measurements through semi-active focusing, SPE-28437, in *Proceedings of the SPE 69th Annual Technical Conference and Exhibition*, 1994.
- Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics, *Journal of Geophysical Research*, **99**(C5), 10,143–10,162, 1994.

- Evensen, G., Advanced data assimilation for strongly nonlinear dynamics, *Monthly Weather Review*, **125**(6), 1342–1354, 1997.
- Evensen, G., The ensemble kalman filter: Theoretical formulation and practical implementation, *Ocean Dynamics*, **53**(4), 343 – 367, 2003.
- Evensen, G., Sampling strategies and square root analysis schemes for the EnKF, *Ocean Dynamics*, **54**(6), 539–560, 2004.
- Evensen, G. and P. J. van Leeuwen, Assimilation of geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model, *Monthly Weather Review*, **124**(1), 85–96, 1996.
- Floris, F. J. T., M. D. Bush, M. Cuypers, F. Roggero, and A.-R. Syversveen, Methods for quantifying the uncertainty of production forecasts: A comparative study, *Petroleum Geoscience*, **7**(SUPP), 87–96, 2001.
- Gao, G., M. Zafari, and A. C. Reynolds, Quantifying uncertainty for the PUNQ-S3 problem in a Bayesian setting with RML and EnKF, SPE 93324, in *Proceedings of the 2005 SPE Reservoir Simulation Symposium*, 2005.
- Gelb, A., *Applied Optimal Estimation*, The M.I.T Press, Cambridge, Massachusetts, and London, England, 1979.
- Goovaerts, P., *Geostatistics for Natural Resources Evaluation*, Osford Univ. Press, New York, 1997.
- Gu, Y. and D. S. Oliver, History matching of the PUNQ-S3 reservoir model using the ensemble Kalman filter, *SPE Journal*, **10**(2), 217–224, 2005.
- Gu, Y. and D. S. Oliver, The ensemble Kalman filter for continuous updating of reservoir simulation models, *Journal of Energy Resources Technology*, **128**(1), 79–87, 2006.
- Hamill, T. M., C. Snyder, D. P. Baumhefner, Z. Toth, and S. L. Mullen, Ensemble forecasting in the short to medium range: Report from a workshop, *Bull. Amer. Meteor. Soc.*, p. (accepted), 2000.
- Houtekamer, P. L. and H. L. Mitchell, Data assimilation using an ensemble Kalman filter technique, *Monthly Weather Review*, **126**(3), 796–811, 1998.
- Houtekamer, P. L. and H. L. Mitchell, A sequential ensemble Kalman filter for atmospheric data assimilation, *Monthly Weather Review*, **129**(1), 123–137, 2001.
- Kalman, R. E., A new approach to linear filtering and prediction problems, *Transactions of the ASME-Journal of Basic Engineering*, **82**(Series D), 35–45, 1960.
- Kalman, R. E. and R. S. Bucy, New results in linear filtering and prediction theory, *Transactions of the ASME-Journal of Basic Engineering*, **83**, 97–107, 1961.

- Kepert, J. D., On ensemble representation of the observation-error covariance in the ensemble Kalman filter, *Ocean Dynamics*, **54**(6), 561–569, 2004.
- Kitanidis, P. K., *Introduction to Geostatistics – Applications in Hydrogeology*, Cambridge University Press, Cambridge, UK, 1997.
- Liu, N. and D. S. Oliver, Critical evaluation of the ensemble Kalman filter on history matching of geologic facies, *SPE Reservoir Evaluation and Engineering*, **8**(6), 470–477, 2005.
- Lorentzen, R., G. Nvdal, B. Valls, A. Berg, and A.-A. Grimstad, Analysis of the ensemble Kalman filter for estimation of permeability and porosity in reservoir models , SPE-96375, in *Proceedings of the 2005 SPE Annual Technical Conference and Exhibition*, 2005.
- Maybeck, P. S., *Stochastic Models, Estimation, and Control (Volume 1)*, Academic Press, New York, 1979.
- Milliken, W. J., A. Emanuel, and A. Chakravarty, Application of 3d streamline simulation to assist history matching, SPE 63155, in *Proceedings of the 2000 SPE Annual Technical Conference and Exhibition*, 2000.
- Nævdal, G., L.M.Johnsen, S.I.Aanonsen, and E.H.Vefring, Reservoir monitoring and continuous model updating using ensemble Kalman filter, *SPE Journal*, **10**(1), 66–74, 2005.
- Nævdal, G., T. mannseth, and E.H.Vefring, Near-well reservoir monitoring through ensemble Kalman filter, SPE-75235, in *Proceedings of the 2002 SPE/DOE Improved Oil Recovery Symposium*, 2002.
- Pacheco, P. S., *Parallel programming with MPI*, Morgan Kaufmann Publishers, Inc., San Francisco, 1997.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in Fortran*, Cambridge University Press, Cambridge, N.Y., Australia, 1992.
- Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, Hydrologic data assimilation with the ensemble Kalman filter, *Monthly Weather Review*, **130**(1), 103–114, 2002.
- Skjervheim, J.-A., G. Evensen, S. Aanonsenand, and T. Johansen, Incorporating 4d seismic data in reservoir simulation models using ensemble Kalman filter, SPE-95789, in *Proceedings of the 2005 SPE Annual Technical Conference and Exhibition*, 2005.
- Tarantola, A., *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, Amsterdam, The Netherlands, 1987.
- Wen, X.-H. and W. H. Chen, Real-time reservoir model updating using ensemble Kalman filter, SPE-92991, in *Proceedings of the 2005 SPE Reservoir Simulation Symposium*, 2005a.

Wen, X.-H. and W. H. Chen, Some practical issues on real-time reservoir model updating using ensemble Kalman filter, IPTC-11024, in *Proceedings of the 2005 International Petroleum Technology Conference*, 2005b.

Zafari, M. and A. Reynolds, Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter, SPE-95750, in *Proceedings of the 2005 SPE Annual Technical Conference and Exhibition*, 2005.

Zang, X. and P. Malanotte-Rizzoli, A comparison of assimilation results from the ensemble Kalman filter and a reduced-rank extended Kalman filter, *Nonlinear Processes in Geophysics*, **10**(6), 477–491, 2003.

Zhang, F. and A. C. Reynolds, E48: Optimization algorithms for automatic history matching of production data, in *Proceedings of the 8th European Conference on the Mathematics of Oil Recovery*, 2002.