A HYBRID ENSEMBLE KALMAN FILTER

FOR NONLINEAR DYNAMICS

A Thesis

by

SHINGO WATANABE

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

MASTER OF SCIENCE

December 2009

Major Subject: Petroleum Engineering

A HYBRID ENSEMBLE KALMAN FILTER

FOR NONLINEAR DYNAMICS

A Thesis

by

SHINGO WATANABE

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

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Akhil Datta-Gupta
Committee Members,	A. Daniel Hill
	Yalchin Efendiev
Head of Department,	Stephen A. Holditch

December 2009

Major Subject: Petroleum Engineering

ABSTRACT

A Hybrid Ensemble Kalman Filter for Nonlinear Dynamics.

(December 2009)

Shingo Watanabe, B.E., Waseda University

Chair of Advisory Committee: Dr. Akhil Datta-Gupta

In this thesis, we propose two novel approaches for hybrid Ensemble Kalman Filter (EnKF) to overcome limitations of the traditional EnKF. The first approach is to swap the ensemble mean for the ensemble mode estimation to improve the covariance calculation in EnKF. The second approach is a coarse scale permeability constraint while updating in EnKF. Both hybrid EnKF approaches are coupled with the streamline based Generalized Travel Time Inversion (GTTI) algorithm for periodic updating of the mean of the ensemble and to sequentially update the ensemble in a hybrid fashion.

Through the development of the hybrid EnKF algorithm, the characteristics of the EnKF are also investigated. We found that the limits of the updated values constrain the assimilation results significantly and it is important to assess the measurement error variance to have a proper balance between preserving the prior information and the observation data misfit. Overshooting problems can be mitigated with the streamline based covariance localizations and normal score transformation of the parameters to support the Gaussian error statistics.

The swapping mean and mode estimation approach can give us a better matching of the data as long as the mode solution of the inversion process is satisfactory in terms of matching the observation trajectory.

The coarse scale permeability constrained hybrid approach gives us better parameter estimation in terms of capturing the main trend of the permeability field and each ensemble member is driven to the posterior mode solution from the inversion process. However the WWCT responses and pressure responses need to be captured through the inversion process to generate physically plausible coarse scale permeability data to constrain hybrid EnKF updating.

Uncertainty quantification methods for EnKF were developed to verify the performance of the proposed hybrid EnKF compared to the traditional EnKF. The results show better assimilation quality through a sequence of updating and a stable solution is demonstrated.

The potential of the proposed hybrid approaches are promising through the synthetic examples and a field scale application.

DEDICATION

To our almighty God for encouraging my faith and guiding me through my life. To my family for supporting my decisions and trusting in me. To my fiancée, Kecia, for her affection, patience, and respect for me. To people around me to help and share their lives with me.

ACKNOWLEDGEMENTS

First of all, I would like to acknowledge my academic adviser, Dr. Akhil Datta-Gupta, for his financial support, academic guidance and encouragement throughout my master degree program.

I would like to acknowledge my committee members, Dr. Daniel Hill and Dr. Yalchin Efendiev, for their advice that shaped my master's thesis.

I would like to acknowledge Dr. Norio Arihara at Waseda University for his recommendation for me to pursue studying at Texas A&M University.

I would like to give special thanks to my senior students in MCERI research group, Ichiro Osako, Deepak Devegowda, Ajitabh Kumar, Akella Santha, Eduardo Jimenez, Adedayo Oyerinde, Ahmed Al-Huthali, Jong-Uk Kim, Prannay Parihar, Sarwesh Kumar, Matthew Talbert and Jiang Xie, for their contributions and achievements in our research area. Also thanks to the current students, Alvaro Jose Rey Amaya, Jichao Yin, Eric Bhark, Qing Tao, Song Du, Han-Young Park, Baljit Sehbi, Yip Yeung Han, Satyajit Taware and Suksang Kang, for their friendship, great inspiration and informative discussions.

NOMENCLATURE

β	Scalar weighting on regularization terms
$\mathbf{C_D}$	Data covariance matrix
$\mathbf{C}_{M^s,d}$	Cross-covariance matrix between data and model parameters
$\mathbf{C}_{\Psi}$	Model covariance matrix
$\mathbf{D}_{obs,k}$	Ensemble of the observation data
$\mathbf{d}_{obs}$	Observation data vector
$\mathbf{d}_{cal}$	Calculated or theoretical observation vector
ε	Noise in the data
$g(\circ)$	Forward model operator
$\mathbf{H}$	Measurement matrix
K	Absolute permeability
k_r	Relative permeability
$\mathbf{m}^s$	Vector of static model variables
$\mathbf{m}^d$	Vector of dynamic model variables
N_d	Number of observation data
N_e	Number of ensemble members
ϕ	Porosity
p_o, p_w, p_g	Phase pressure
ρ	Density

$\boldsymbol{\rho}$	Covariance localizing function
$\mathbf{R}$	Precision of the coarse-scale data
$\mathbf{S}$	Sensitivity matrix
S_o, S_w, S_g	Phase saturation
t	Time
$\mathbf{U}$	Upscaling operator
τ	Time of flight
$\mathbf{y}$	Model state vector
μ	Viscosity
σ	Standard deviation
$\boldsymbol{\Psi}$	Ensemble of model state vector

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION v

ACKNOWLEDGEMENTS vi

NOMENCLATURE vii

TABLE OF CONTENTS ix

LIST OF FIGURES xii

CHAPTER

I INTRODUCTION 1

1.1 Background and Literature Review 3
1.2 Objectives of Study 6
1.3 Thesis Outline 8

II MATHEMATCAL FORMULATION FOR ENSEMBLE KALMAN FILTER 10

2.1 Time-independent Case 10
2.1.1 State Estimation Problem 11
2.1.2 Bayesian Formulation of the Kalman Equations 14
2.1.3 Extension to Multivariate Case 17
2.2 Time-dependent Sequential Model Updating Case 20
2.2.1 Kalman Filter for a Scalar Case with Linear Dynamics 21
2.2.2 Kalman Filter for a Vector Case with Linear Dynamics 24
2.2.3 Extended Kalman Filter for Nonlinear Dynamics 25
2.3 Ensemble Kalman Filter 28
2.3.1 Sampling Representation of Error Statistics 28
2.3.2 Analysis Scheme 29
2.4 EnKF for History Matching in Reservoir Problem 32
2.4.1 History Matching Terminology for EnKF 32
2.4.2 Ensemble of the State Vector 35

CHAPTER Page

2.4.3 Forecast Step and Update Step in EnKF 36
2.5 Hybrid Ensemble Kalman Filter Formulations 40
2.5.1 Swapping of Mean and Mode Estimation 42
2.5.2 Coarse Scale Permeability Constraint 44

III CHARACTERISTICS OF ENSEMBLE KALMAN FILTER 47

3.1 Application to a Synthetic Case 47
3.2 Streamline Based Covariance Localization 51
3.3 Normal Score Transformation of the Parameter 58
3.4 Goldsmith Field Application 62
3.4.1 Sensitivity with Updating Limit Constraint 65
3.4.2 Sensitivity with the Measurement Error Variance 70
3.4.3 Comparison with Localization Methods 72
3.4.4 Normal Score Transformation with Localization Methods ... 74
3.5 Summary of Chapter III 77

IV A HYBRID ENSEMBLE KALMAN FILTER APPLICATION 78

4.1 Swapping Mean and Mode Estimation Hybrid Approach 79
4.1.1 Sensitivity with Initial Model Biasness 83
4.2 Hybrid EnKF with Coarse Scale Permeability Constraint 86
4.2.1 Effect of the Upscaling Factor 88
4.2.2 Effect of the Coarse Scale Data Error Variance 91
4.3 Another Synthetic Case 92
4.3.1 Test of the Validity of the Approach 94
4.3.2 Inversion Coupling Hybrid EnKF 97
4.4 Another Nine Spot Synthetic Case 100
4.4.1 Comparison of Plain EnKF and Hybrid EnKF 103
4.4.2 Hybrid EnKF with Streamline Trajectory Localization 107
4.4.3 WWCT and WBHP Simultaneous Inversion 109
4.5 Uncertainty Quantification Methods for Hybrid EnKF 113
4.5.1 Water Front Movement 114
4.5.2 Multi-dimensional Scaling of the Swept Volume Changes .. 116
4.5.3 Quantitative Comparison and Validation of Hybrid EnKF 118
4.6 Gold Smith Field Application 123
4.7 Summary of Chapter IV 126

V CONCLUDING REMARKS AND RECOMMENDATIONS 128

5.1 Conclusions 128

Page

5.2 Recommendations 132

REFERENCES 135

APPENDIX A 138

APPENDIX B 140

VITA 142

LIST OF FIGURES

Page

Figure 3.1 Reference model log permeability and its spatial histogram 38

Figure 3.2 Field cumulative oil production responses from the ensemble models; the reference model is in red line; (a) is the initial ensemble models, (b) is the EnKF updated models. 40

Figure 3.3 Well water cut responses from the ensemble models; the reference model is in red line; first row is the initial ensemble models, the second is the EnKF updated models. 40

Figure 3.4 Mean of the ensemble permeability fields; reference model in the left, the initial mean of the ensemble models in the center, and the mean of the final updated ensemble models. 40

Figure 3.5 Spatial histograms of the log permeability values comparison initial models and updated models; in the first row and from the left, reference model, the ensemble No.1 model, and the ensemble No.50, the mean of the ensemble models in the second row from the left, the ensemble No.1 model, and the ensemble No. 50, the mean of the ensemble models..... 41

Figure 3.6 An example for the streamline trajectory localization; the localizing function is plotted for each producer, from left P1, P2 and P3. The red color is 1's and blue color is 0's. 43

Figure 3.7 An example for the streamline sensitivity localization; the localizing function is plotted for each producer, from left P1, P2 and P3. 44

Figure 3.8 Field cumulative oil production responses from the ensemble models; the reference model is in red line; (a) is the initial ensemble models, (b) is the EnKF updated models with streamline trajectory localization. 45

Figure 3.9 Well water cut responses from the ensemble models; the reference model is in red line; first row is the initial ensemble models, the second is the EnKF updated models with streamline trajectory localization. 45

Page

Figure 3.10 Mean of the ensemble permeability fields; reference model in the left, the mean of the EnKF final updated ensemble models in the center, and the mean of the EnKF updated ensemble models with streamline trajectory localization..... 46

Figure 3.11 (a)Well configuration of the synthetic problem from an initial model and (b)streamline trajectory of the model..... 46

Figure 3.12 Spatial histograms of the log permeability values comparison EnKF updated models and EnKF updated models with streamline trajectory localization; in the first row and from the left, reference model, the ensemble No.1 model, and the ensemble No. 50, the mean of the ensemble models in the second row from the left, the ensemble No.1 model, and the ensemble No. 50, the mean of the ensemble models. 47

Figure 3.13 Spatial histograms of the log permeability values comparison initial ensemble models and normal scored ensemble models; in the first row and from the left, initial permeability spatial histogram of ensemble No. 1, No. 25 and No. 50 models, in the second row and from left, the normal scored permeability spatial histogram of ensemble No. 1, No. 25, and No. 50 models..... 48

Figure 3.14 Field cumulative oil production responses from the ensemble models; the reference model is in red line; (a) is the initial ensemble models, (b) is the EnKF updated models, (c) is EnKF updated models with the normal score transformation of the permeability..... 49

Figure 3.15 Well water cut responses from the ensemble models; the reference model is in red line; first row is the initial ensemble models, the second row is the EnKF updated models, and the third row is the EnKF updated models with normal score transformation. 50

Figure 3.16 Mean of the ensemble permeability fields; reference model in the left, the mean of the EnKF final updated ensemble models in the center, and the mean of the EnKF final updated ensemble models with normal score transformation..... 50

Page

Figure 3.17 Spatial histograms of the log permeability values comparison; in the first row and from the left, initial permeability spatial histograms of ensemble No. 1,No. 25 and No. 50 models, in the second row and from left, EnKF updated permeability spatial histograms of ensemble No. 1, No. 25, and No. 50 models and in the third row and from left, Normal EnKF updated permeability spatial histogram of ensemble No. 1, No. 25, and No. 50 models. 51

Figure 3.18 Gold Smith well configuration map 52

Figure 3.19 Gold Smith simulation model..... 52

Figure 3.20 Initial 50 ensemble model WWCT responses; the observation data in blue points and the initial ensemble model responses are in grey lines, in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5, and P6, and in the third row, from left, P7, P8 and P9. . 53

Figure 3.21 Initial ensemble permeability fields; from the left, ensemble No. 1, No. 25 and No. 50 at the depth of 1080 ft, 2042 ft and 3004 ft from the top..... 54

Figure 3.22 EnKF final 50 updated ensemble model WWCT responses without limit value constraints; the observation data in blue points and the ensemble model responses are in light blue lines; in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5,and P6, and in the third row, from left, P7, P8 and P9..... 55

Figure 3.23 The ensemble No. 1 permeability field and spatial distribution comparison between (a) the mean of the initial ensemble models and the mean of the updated ensemble models. 56

Figure 3.24 EnKF final 50 updated ensemble model WWCT responses with limit value constraints; the observation data in blue points and the ensemble model responses are in light blue lines; in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5,and P6, and in the third row, from left, P7, P8 and P9. Vertical line shows the last assimilation time step..... 57

Figure 3.25 Updated ensemble permeability fields; from the left ensemble No. 1, No. 25 and No. 50 at the depth of 1080 ft, 2042 ft and 3004 ft from the top..... 58

Page

Figure 3.26 Spatial histograms of the log permeability values comparison between initial ensemble models and EnKF updated models; in the first row and from the left, the initial ensemble No. 1, No. 25 and No. 50 models, in the second row from the left, the updated ensemble No. 1, No. 25 and No. 50 models. 58

Figure 3.27 EnKF final 50 updated ensemble model WWCT responses with limit value constraints for measurement error variance =5 %; the observation data in blue points and the ensemble model responses are in light blue lines, in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5, and P6, and in the third row, from left, P7, P8 and P9.Vertical line shows the last assimilation time step. 59

Figure 3.28 The ensemble No. 1 permeability field and spatial distribution..... 60

Figure 3.29 EnKF final 50 updated ensemble model WWCT responses comparisons between (a) ENKF with streamline trajectory localization and (b) EnKF with streamline sensitivity localization; the observation data in blue points and the ensemble model responses are in light blue lines, in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5, and P6, and in the third row, from left, P7, P8 and P9.Vertical line shows the last assimilation time step. 61

Figure 3.30 The ensemble No. 1 permeability field and spatial histogram comparison between the initial ensemble model, the updated ensemble model from EnKF-ST and the updated ensemble model from EnKF-SS. 61

Figure 3.31 EnKF final 50 updated ensemble model WWCT responses comparisons between (a) normal scored ENKF with streamline trajectory localization and (b) normal scored EnKF with streamline sensitivity localization; the observation data in blue points and the ensemble model responses are in light blue lines, in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5, and P6, and in the third row, from left, P7, P8 and P9. Vertical line show the last assimilation time step..... 62

Figure 3.32 The ensemble No. 1 permeability field and spatial histogram and changes from the initial model No. 1; from the left the initial ensemble model No1, the updated ensemble model No. 1 from NST-EnKF and the change that we make from NST-EnKF updating from the initial model No. 1. 63

Page

Figure 3.33 The ensemble No. 1 permeability field and spatial histogram and changes from the initial model No. 1; from the left the initial ensemble model No1, the updated ensemble model No. 1 from NST-EnKF and the change that we make from NST-EnKF updating from the initial model No. 1.. 63

Figure 4.1 Reference model log permeability and its spatial histogram 66

Figure 4.2 Field cumulative oil production responses from the ensemble models; the reference model is in red line, Plain EnKF ensemble model responses in light blue, and Hybrid-SMM ones in light green; (a) is the initial ensemble models, (b) is Plain EnKF updated models, and (c) is Hybrid-SMM-EnKF updated models and (d) is the comparison between the mean of Plain EnKF updated ensemble models and the mean of Hybrid-SMM- EnKF updated ensemble models. 67

Figure 4.3 Well water cut responses from the ensemble models; the reference model is in red line; the first row is the initial ensemble models, the second row is Plain EnKF updated ensemble models, and the third row is Hybrid-SMM-EnKF-updated ensemble models, the forth row is the comparison between the mean of the updated ensemble models from Plain EnKF and Hybrid-SMM-EnKF..... 68

Figure 4.4 Mean of the ensemble permeability fields comparison; (a) reference model, (b) the initial mean of the ensemble models, and (c) the mean of the final updated ensemble models from Hybrid-SMM EnKF, (c) the mean of the final updated ensemble models from Plain EnKF. 69

Figure 4.5 Reference model log permeability..... 70

Figure 4.6 Field cumulative oil production responses from the ensemble models; the reference model is in red line, Plain EnKF ensemble model responses in light blue, and Hybrid-SMM ones in light green; (a) is the initial ensemble models, (b) is Plain EnKF updated models, and (c) is Hybrid-SMM EnKF and (d) is the comparison between the mean of Plain EnKF updated ensemble models and the mean of Hybrid EnKF updated ensemble models. 70

Page

Figure 4.7 Well water cut responses from the ensemble models; the reference model is in red line; the first row is the initial ensemble models, the second row is Plain EnKF updated ensemble models, and the third row is Hybrid-SMM EnKF-updated ensemble models, the forth row is the comparison between the mean of the updated ensemble models from Plain EnKF and Hybrid-SMM-EnKF..... 71

Figure 4.8 Mean of the ensemble permeability fields comparison; (a) reference model, (b) the initial mean of the ensemble models in the center, and (c) the mean of the final updated ensemble models from Hybrid-SMM EnKF, (d) the mean of the final updated ensemble models from Plain EnKF..... 72

Figure 4.9 Reference model log permeability field..... 74

Figure 4.10 Well water cut responses comparison from the ensemble models; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line, (a) Initial ensemble models, (b) Plain EnKF updated models, in the top row from left, P1 and P2 and the bottom row from left, P3 and P4. 74

Figure 4.11 Well water cut responses comparison from the ensemble models; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line; (a) Plain EnKF updated models, (b) Hybrid EnKF updated models with upscaling factor=2, (b) Hybrid EnKF updated models with upscaling factor=5, and (c) Hybrid EnKF updated models with upscaling factor=10, in the top row from left, P1 and P2 and the bottom row from left, P3 and P4. 75

Figure 4.12 Well water cut responses comparison from the upscaled permeability models; the reference model is in red line, the upscaled model with upscaling factor=2 in blue, the upscaled model with upscaling factor=5 in light blue, the upscaled model with upscaling factor=10 in light green, in the top row from left, P1 and P2 and the bottom row from left, P3 and P4. 76

Page

Figure 4.13 Well water cut responses comparison from the ensemble models; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line; (a) Updated models with error std lnK=1, (b) Updated models with error std lnK=4. 77

Figure 4.14 Initial 50 ensemble model WWCT and WBHP responses; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line, in the top row from the left, P1, P2, and P3 in the bottom row. 79

Figure 4.15 Well water cut responses comparison from the ensemble models; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line; (a) Plain EnKF updated models, (b) Hybrid EnKF updated models with reference coarse scale permeability data, in the top row from the left, P1, P2, and P3.in the bottom row. 80

Figure 4.16 Well bottom-hole pressure responses comparison from the ensemble models; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line; (a) Plain EnKF updated models, (b) Hybrid EnKF updated models with reference coarse scale permeability data, in the top row from the left, I1 and P1, and P2 and P3 in the bottom row.. 80

Figure 4.17 The ensemble permeability fields comparison; in the first row from left, the reference model, initial model No. 1and No. 25, the initial mean model, and in the second row from left, Plain EnKF updated model No. 1 and No. 25, the updated mean model, and in the third row from left, Hybrid EnKF updated model No. 1and No. 25, updated mean model. 81

Figure 4.18 The ensemble spatial permeability histogram comparison; in the first row from left, the reference model, initial model No. 1and No. 25, the initial mean model, and in the second row from left, Plain EnKF updated model No. 1and No. 25, the updated mean model, and in the third row from left, Hybrid EnKF updated model No. 1and No. 25, the updated mean model. 81

Page

Figure 4.19 Well water cut responses comparison from the ensemble models; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line; (a) Plain EnKF updated models, (b) Hybrid EnKF updated models, in the top row from the left, P1, P2, and P3 in the bottom row. 83

Figure 4.20 Well bottom-hole pressure responses comparison from the ensemble updated models; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line; (a) Plain EnKF updated models, (b) Hybrid EnKF updated models, in the top row from the left, I1, P1, and P2, P3 in the bottom row. 83

Figure 4.21 Well water cut inversion matching results at 2400 days; the reference model response is in blue point, initial model response in light green triangles and the final updated model response in pink square, (a) P1, (b) P2, and (c) P3 respectively. 84

Figure 4.22 Relative objective function of the GTTI inversion process; the GTTI misfit is in blue line, and the amplitude misfit is in pink line. 84

Figure 4.23 Permeability fields; from left (a) reference model, (b) initial model No. 1 and (c) initial mean model 86

Figure 4.24 Initial 50 ensemble model WWCT and WBHP responses; the reference model is in red line, ensemble model responses in light green lines and the response from the initial mean ensemble model in blue line, (a) WWCT responses in the first row from the left, P1, P2, and P3, in the second row P4, P5, and P6, and in the third row from left P7 and P8, (b) WBHP responses in the first row from the left, I1, P1, and P2, in the second row P3, P4, and P5, and in the third row from left P6, P7, and P8. 87

Figure 4.25 Ensemble updated model WBHP matching results comparison between (a) Plain EnKF (b) Hybrid EnKF; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line, in the first row from the left, I1, and P1, P2, in the second row P3, P4, and P5, and in the third row from left P6, P7 and P8. 88

Page

Figure 4.26 Ensemble updated model WWCT matching results comparison between (a) Plain EnKF and (b) Hybrid EnKF; the reference model is in red line, ensemble model responses in light green lines and the response from the mean updated ensemble model in blue line, in the first row from the left, P1, P2, and P3, in the second row P4, P5, and P6, and in the third row from left P7, P8. 88

Figure 4.27 Permeability fields comparisons between the initial models, Plain EnKF updated models and Hybrid EnKF updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left Inversion mode model, Hybrid EnKF updated model No. 1, No. 25, and the updated mean ensemble model. 89

Figure 4.28 Permeability spatial histogram comparisons between the initial models, Plain EnKF updated models and Hybrid EnKF updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left inversion solution model, Hybrid EnKF updated model No. 1, and No. 25, the updated mean ensemble model. 90

Figure 4.29 Ensemble updated model WBHP matching results comparison between (a) Plain EnKF with streamline trajectory localization and (b) Hybrid EnKF with streamline trajectory localization; the Reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line, in the first row from the left, I1, and P1, P2, in the second row P3, P4, and P5, and in the third row from left P6, P7 and P8. 91

Figure 4.30 Ensemble updated model WWCT matching results comparison between (a) Plain EnKF with streamline trajectory localization and (b) Hybrid EnKF; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line, in the first row from the left, P1, P2, and P3, in the second row P4, P5, and P6, and in the third row from left P7 and P8. 91

Page

Figure 4.31 Permeability fields comparisons between the initial models, Plain EnKF-ST updated models and Hybrid EnKF-ST updated models; in the first row from left reference model, initial model No. 1, and No. 25,the initial mean model, in the second row, Plain EnKF-ST updated model No. 1, No. 25, and the updated mean ensemble model, in the third row,from left inversion solution model, Hybrid EnKF-ST updated model No. 1, and No. 25, the updated mean ensemble model. 92

Figure 4.32 Permeability spatial histogram comparisons between the initial models, Plain EnKF-ST updated models and Hybrid EnKF-ST updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF-ST updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left inversion solution model, Hybrid EnKF-ST updated model No. 1, No. 25, and the updated mean ensemble model..... 92

Figure 4.33 Ensemble updated model WWCT matching results comparison between (a) Plain EnKF and (b) Hybrid EnKF; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line, in the first row from the left, I1, and P1, P2, in the second row P3, P4, and P5, and in the third row from left P6, P7 and P8..... 94

Figure 4.34 Ensemble updated model WBHP matching results comparison between (a) Plain EnKF and (b) Hybrid EnKF; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line, in the first row from the left, P1, P2, and P3, in the second row P4, P5, and P6, and in the third row from left P7, P8. 94

Figure 4.35 Permeability fields comparisons between the initial models, Plain EnKF updated models and Hybrid EnKF updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left inversion solution model, Hybrid EnKF updated model No. 1, and No. 25, the updated mean ensemble model. 95

Page

Figure 4.36 Permeability spatial histogram comparisons between the initial models, Plain EnKF updated models and Hybrid EnKF updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left inversion solution model, Hybrid EnKF updated model No. 1, and No. 25, the updated mean ensemble model. 96

Figure 4.37 Water saturation map at 2500 days comparisons between the initial models, Plain EnKF updated models and Hybrid EnKF updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left inversion solution model, Hybrid EnKF updated model No. 1, and No. 25, the updated mean ensemble model. 97

Figure 4.38 Water saturation map at 4000 days comparisons between the initial models, Plain EnKF updated models and Hybrid EnKF updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left inversion solution model, Hybrid EnKF updated model No. 1, and No. 25, the updated mean ensemble model. 98

Figure 4.39 Multi-dimensional scaling of the initial ensemble models and updated ensemble models from Plain EnKF and Hybrid-Corse EnKF; (a) 2 dimensions plot, (b) 3 dimensions plot..... 100

Figure 4.40 RMSE comparisons; Plain EnKF is Plain EnKF without localization, Plain EnKF SL is the plain EnKF with localization and Hybrid EnKFis Hybrid EnKF without localization. 102

Figure 4.41 Error covariance estimation evolutions; Plain EnKF SL is the plain EnKF with localization and Hybrid EnKF is Hybrid EnKF without localization. 103

Figure 4.42 χ^2 validation test comparison; the conventional EnKF without streamline trajectory localization, the conventional EnKF with streamline trajectory localization, and the hybrid EnKF 104

Page

Figure 4.43 Innovation distribution comparison; (a) Plain EnKF with streamline trajectory localization, (b) Hybrid EnKF without localization..... 105

Figure 4.44 Initial 50 ensemble model WWCT responses; the observation data in red points and the initial ensemble model responses are in grey lines; in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5, and P6, and in the third row, from left, P7, P8 and P9..... 106

Figure 4.45 EnKF final 50 updated ensemble model WWCT responses comparisons between (a) Plain ENKF and (b)Hybrid EnKF; the observation data in red points and the ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line,in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5,and P6, and in the third row, from left, P7,P8 and P9. Vertical line shows the last assimilation time step..... 107

Figure 4.46 Updated ensemble permeability fields comparison of (a) Plain EnKF , from the left, ensemble No. 1, No. 25 and (b) Hybrid EnKF, from the left, ensemble No. 1, No. 25 at the depth of 1080 ft, 2042 ft and 3004 ft from the top. 108

Figure A.1 Fine scale cells and coarse scale cell 118

CHAPTER I

INTRODUCTION

History matching is reconciling reservoir simulation models to the dynamic response of the reservoir history data for building reliable reservoir performance models and assessing the underlying uncertainties in the geological models. Integration of dynamic data generally leads to an inverse problem and requires solution of the flow equations several times using iterative procedures. Traditionally, history matching is performed manually by applying local and regional changes to reservoir properties. Such trial-and error involves considerable subjective judgment and time-consuming work, and very often creates artificial discontinuities with loss of geologic realism in the updated models.

Over the past few years, automatic history matching or production data integration methods were developed by utilizing inverse theory to minimize an appropriately defined misfit function to obtain a history match. If the misfit function is solely defined from the observation data at the wells, the solutions will be non-unique and potentially unstable. Formally, this class of inverse problem is known as ill-posed, and must be regularized by constraining the solution to the independent prior information.

This thesis follows the style and format of the *SPE Journal*.

However, the increase in the development of permanent sensors for monitoring pressure, temperature, resistivity, or flow rate has added impetus to the continuous model updating method development. Because of the high data frequency, simultaneously matching all available data to update 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 updated sequentially. Moreover, there has been a paradigm shift from history matching a single reservoir model to generating a suite of realizations. This provides for a measure of uncertainty in production forecasts and assists better management decisions for reservoir development. Both the heavy computational burden and the high data sampling frequency require a new kind of history-matching method.

Ensemble Kalman Filter (EnKF) is one such promising technique for generating a suite of plausible reservoir models (Evensen 1994). It is a Monte Carlo approach, in which a suite of reservoir models or an ensemble of models is used, observation data are sequentially assimilated or matched as they become available, and the ensemble of models are continuously updated to honor the current data without re-matching previous history data.

The EnKF utilizes the correlation between reservoir responses (e.g. rates, bottom-hole pressures, gas-oil ratios, and water cuts) and reservoir variables (e.g. porosity and permeability) estimated directly from the sample covariance matrix between model parameters and model responses. It should be noted that the EnKF is optimal only in linear dynamics and Gaussian statistics systems, and is sub-optimal for nonlinear

dynamics and non-Gaussian statistics systems. As the sample size grows larger, however, it has been shown to asymptotically provide the best linear estimates of the states and parameters.

1.1 Background and Literature Review

The EnKF has been applied in field-scale problems for reservoir characterization (Naevdal et al. 2003; Gu and Oliver 2004; Zafari and Reynolds 2005; Skjervheim et al. 2005); however, there are several difficulties that arise during theses applications. The set of parameters to be estimated is typically orders of magnitude larger than the ensemble size for reservoir history matching problems. A small ensemble however, may lead to noisy or inaccurate covariance and cross-covariance estimations and degrade the performance of the EnKF over a sequence of updates (Anderson and Anderson 1999; Devegowda et al. 2007). It is crucial to accurately estimate the covariance matrices for large-scale problems from a finite ensemble size.

One potential problem in the EnKF is that updating both the model parameters and state variables simultaneously may violate their nonlinear relationship (Gu and Oliver 2006). The model parameter and state variables are related to each other by a nonlinear governing equation. The linear updating of the model parameter affects the state updating and that results in the material balance error or destroys the physical realism of the problem. 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. For a nonlinear problem, however, when both state and parameters are updated, it may be impossible to update the state variables to be consistent with the updated model parameters without re-solving the nonlinear forward equations when the changes are large.

Wen and Chen (2005) proposed an intuitive remedy for this problem. They suggested introducing a so-called "conforming step" at each measurement time. This is simply rerunning the forward simulation with updated parameters to achieve the consistent state variables at the current measurement time. It can be easily inferred that their approach doubles the computing time compared to that of the EnKF since there are two simulation runs for each simulation process in the forecast and conforming steps. They also suggested iterating this conforming step is necessary only when nonlinearity of the problem is strong.

Another way to overcome the nonlinearity in the problem is Ensemble Randomized Maximum Likelihood Filter (EnRMLF) or iterative Ensemble Kalman Filter proposed by (Gu and Oliver 2006; Li and Reynolds 2007). 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. Their approach adopts the iterative Gauss-Newton formula to update the model parameters. After the new model parameters are obtained, the system governing equations are re-initialized from time zero to compute the state variables for consistency at the current state. The drawback of this method is the intensive computational cost from rerunning the forward simulation iteratively.

However, they claimed 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, they found that it outperforms the traditional EnKF.

In the atmospheric science literature, there is a different way of approaching the non linearity of the state estimation problem. Evolution of the ensemble model responses can deviate from the true model trajectory of the problem by a sequence of updating. It is because of the linear updating and mean and covariance estimation discrepancy in non-Gaussian system. Fuqing Zhang and Meng Zhang (2009) proposed a hybrid Ensemble Kalman filter approach where coupling deterministic four-dimensional variational assimilation (4DVAR) with an ensemble Kalman filter (EnKF) produced a superior data assimilation approach. The coupled assimilation scheme (E4DVAR) benefits from using the state-dependent uncertainty provided by EnKF while taking advantage of 4DVAR in preventing filter divergence.

It is also widely recognized that integration of data at various scales could further reduce uncertainty in our estimates of the reservoir variables (Efendiev et al. 2005; Lee et al. 2002). Akella et al. (2008) proposed a novel approach to integrate data at different scales using the ensemble Kalman filter (EnKF), such that the finest scale data is sequentially estimated, subject to the available data at the coarse scale(s), as additional constraints. Their results show that higher precision in the coarse-scale data yielded improved estimates. By constraining the high resolution variations in model properties to coarse-scale information, the estimates of the unknowns are better resolved by reducing

the dimensionality of the underlying problem. Devegowda et al. (2009) proposed coupling EnKF with a deterministic inversion algorithm to impose coarse-scale spatial saturation distribution of the reservoir to infer saturation distributions using tracer responses and showed better estimate of the saturation profiles compared to the traditional EnKF.

1.2 Objectives of Study

The primary objective of this study is to develop a hybrid Ensemble Kalman filter algorithm to overcome traditional EnKF issues related to nonlinear dynamics, parameter overshooting and loss of geologic information in the updated models.

Our goal is to improve EnKF performance without computationally demanding iterative rerunning of forward simulations to achieve the consistency in model parameter and state variable updating. Instead, by coupling the inversion algorithm on the mean of the ensemble model, we intend to make the responses of the mean follow the observation trajectory and attain better mode estimation from the inversion process. For that purpose, streamline-derived generalized travel time inversion (GTTI) technique is coupled with EnKF to obtain a better estimate of permeability distribution.

Two approaches for a hybrid Ensemble Kalman filter are proposed: swapping of mean and mode estimation during EnKF and a coarse scale permeability constraint during EnKF updating. For the swapping of mean and mode estimation approach, we simply change the way to estimate the covariance matrix by substituting the ensemble

mean with the inversion mode model. For the coarse scale permeability constrained updating approach, a flow-based upscaling method is utilized to relate the fine and coarse-scale permeability data, and the coarse-scale data constrains a sequential second stage EnKF updating. The coarse-scale data is obtained by an iterative inversion of the production data starting with the ensemble mean. By imposing upscaled coarse-scale permeability data on EnKF updating, each ensemble model updating is constrained towards the mode model and we obtain a better assimilation and prevent filter divergence simultaneously. In addition, by integrating the different scale of the data, our estimates of the permeability captures the large structure of the geological features, and consequently it reduces the uncertainty of the main features of the spatial continuity of the permeability field.

Through the hybrid EnKF development, the EnKF characteristics in history matching is investigated under various conditions such as non-Gaussian parameter distribution and the localization effects of the covariance estimation. Sensitivity of EnKF updating with respect to the observation error values is also investigated, and overshooting problems in the updated model are examined.

We also analyze the uncertainty quantification on the updated models by using a multi-dimensional scaling method to visualize the model separation in a Euclidean space. Moreover, to quantify the performance of the proposed hybrid EnKF compared to the plain EnKF, four validation methods employed by Zupanski (2004) are used to verify the hybrid EnKF results.

The advantages of the proposed hybrid EnKF approach over the general EnKF are demonstrated through synthetic problems and a field application problem.

1.3 Thesis Outline

Chapter II contains the mathematical formulations for Ensemble Kalman Filter from the state estimation problem to the combined parameter and state estimation problem. It discusses the derivation of the Kalman equations thoroughly, and the evolution from Kalman filter to Extended Kalman filter and Ensemble Kalman filter is explained. Finally, it summarizes the EnKF application in reservoir problems and the proposed Hybrid EnKF equations are expressed.

Chapter III covers the characteristics of EnKF in reservoir problems. Performance of the EnKF is investigated in synthetic examples and the Goldsmith San Andres Unit field application. Sensitivity of the updating results with EnKF parameters is discussed in detail. Moreover, the covariance localization methods, streamline trajectory-based localization and streamline sensitivity-based localization, are compared in terms of both quality of the history matching and the parameter distributions. In addition to that, normal score transformation of the permeability variables circumvents the difficulty for non-Gaussian or bi-modal prior model distribution.

Chapter IV contains the proposed hybrid EnKF research work. Both swapping of mean and mode estimation approach and the coarse scale permeability constraint approach are applied to the same synthetic case as in Chapter III and their performances

are investigated in detail. The validation tests of the coarse scale permeability constraint hybrid EnKF are conducted. Finally the same filed case as in Chapter III is tested by the hybrid EnKF.

Chapter V summarizes all conclusions drawn from this study and categorized in each perspective of the study. Also, the recommendations and the future study areas are suggested.

Appendix A explains the flow based upscaling algorithm used in the coarse scale permeability constraint approach.

Appendix B explains the algorithm of Multi-dimensional scaling method (MDS) in the uncertainty analysis.

CHAPTER II

MATHEMATCAL FORMULATION FOR ENSEMBLE KALMAN FILTER

In this chapter, the derivation of the Kalman filter (Evensen 2006) is shown from the state estimation problem and extended to both time and space domains from the time-independent case to the time-dependent case, and from a scalar case to a multivariate case. Furthermore the transition from the linear dynamics to nonlinear dynamics for Kaman filter is discussed from the linearized approximation of the evolution of the error covariance in the Extended Kalman Filter (EKF) to the Monte Carlo approach of the estimation of the covariance calculations in the Ensemble Kalman Filter (EnKF). Finally the ensemble Kalman filter formulation is related to the combined parameter and state estimation problem in the scope of the reservoir characterization problem and a novel approach of a Hybrid EnKF formulation is introduced in two different ways: swapping of the mean estimate to a mode solution from the deterministic inversion algorithm and a coarse scale permeability constrained EnKF updating.

2.1 Time-independent Case

This section discusses the problem of how to improve a model prediction of a state variable at a given time by integrating available measurements at the particular time.

We start by deriving the optimal linear and unbiased estimator for the scalar state estimation problem.

2.1.1 State Estimation Problem

Suppose the true state y_{true} (e.g. a pressure at a particular location and time) is estimated as follows

$$y^f = y_{true} + p^f \tag{2.1}$$

$$d_{obs} = y_{true} + \varepsilon \tag{2.2}$$

where y^f is a model forecast state and d_{obs} is a measurement of y_{true}. The term p^f denotes the forecast error, which is unknown and ε is the unknown measurement error. In order to improve the estimate of y_{true}, the following assumptions are made:

$$\begin{aligned} &\overline{p^f} = 0, && \overline{\left(p^f\right)^2} = C_{yy}^f \\ &\overline{\varepsilon} = 0, && \overline{(\varepsilon)^2} = C_{\varepsilon\varepsilon} \\ &\overline{p^f \varepsilon} = 0. \end{aligned} \tag{2.3}$$

Here the over bar denotes ensemble averaging or expected value. Now, we seek a linear estimator for the analyzed state estimate y^a as

$$y^a = y_{true} + p^a = \alpha_1 y^f + \alpha_2 d_{obs}, \tag{2.4}$$

where we define

$$\overline{p^a} = 0, \qquad \overline{\left(p^a\right)^2} = C^a_{yy}. \tag{2.5}$$

The definition of the Eq.(2.5) means that we assume that the error p^a, in the analyzed estimate is unbiased. In the other words, the analyzed estimate becomes an unbiased estimate of the true state y_{true}, i.e. $\overline{y^a} = y_{true}$.

Inserting the two estimates from Eq. (2.1) and Eq. (2.2) in the Eq. (2.4), we get

$$y_{true} + p^a = \alpha_1 (y_{true} + p^f) + \alpha_2 (y_{true} + \varepsilon) \tag{2.6}$$

Take the expectation of the above equation, we obtain

$$y_{true} = \alpha_1 y_{true} + \alpha_2 y_{true} = (\alpha_1 + \alpha_2) y_{true}. \tag{2.7}$$

Thus, the requirement for α_1 and α_2 is

$$\alpha_1 + \alpha_2 = 1, \text{ or } \alpha_1 = 1 - \alpha_2. \tag{2.8}$$

And the linear unbiased estimator for y_{true}, Eq.(2.4) is rewritten as

$$y^a = (1 - \alpha_2) y^f + \alpha_2 d_{obs} = y^f + \alpha_2 (d_{obs} - y^f). \tag{2.9}$$

Substituting Eq.(2.1),(2.2) and (2.4) in the above equation gives the analysis error as follows,

$$p^a = p^f + \alpha_2(\varepsilon - p^f)\,. \tag{2.10}$$

From Eq. (2.5), the error variance becomes

$$\begin{aligned}\overline{(p^a)^2} = C^a_{yy} &= \overline{(p^f + \alpha_2(\varepsilon - p^f))^2} \\ &= \overline{(p^f)^2} + 2\alpha_2\,\overline{p^f(\varepsilon - p^f)} + \alpha_2^2\,\overline{(\varepsilon^2 - 2\varepsilon p^f + (p^f)^2)} \\ &= C^f_{yy} - 2\alpha_2 C^f_{yy} + \alpha_2^2(C_{\varepsilon\varepsilon} + C^f_{yy})\,.\end{aligned} \tag{2.11}$$

The minimum variance is drawn by taking the derivative of Eq. (2.11) as

$$\frac{dC^a_{yy}}{d\alpha_2} = -2C^f_{yy} + 2\alpha_2(C_{\varepsilon\varepsilon} + C^f_{yy}) = 0\,. \tag{2.12}$$

Solving for α_2 gives

$$\alpha_2 = \frac{C^f_{yy}}{C_{\varepsilon\varepsilon} + C^f_{yy}}\,. \tag{2.13}$$

Thus, the analyzed estimate, Eq. (2.9), is rewritten as

$$y^a = y^f + \frac{C_{yy}^f}{C_{\varepsilon\varepsilon} + C_{yy}^f}(d_{obs} - y^f)$$
$$= y^f + K(d_{obs} - y^f) \tag{2.14}$$

where $K(=\alpha_2)$ is the Kalman Gain. Further, the error variance of the analyzed estimate is calculated from Eq. (2.11) and Eq.(2.13) by plugging in the Kalman Gain

$$C_{\mathbf{yy}}^a = C_{yy}^f - 2\frac{C_{yy}^f}{C_{\varepsilon\varepsilon} + C_{yy}^f}C_{yy}^f + \left(\frac{C_{yy}^f}{C_{\varepsilon\varepsilon} + C_{yy}^f}\right)^2 (C_{\varepsilon\varepsilon} + C_{yy}^f)$$
$$= C_{yy}^f - \frac{(C_{yy}^f)^2}{C_{\varepsilon\varepsilon} + C_{yy}^f} = C_{yy}^f\left(1 - \frac{C_{yy}^f}{C_{\varepsilon\varepsilon} + C_{yy}^f}\right)$$
$$= C_{yy}^f(1-K) \tag{2.15}$$

2.1.2 Bayesian Formulation of the Kalman Equations

Given a probability density function $f(y)$ for the forecast state estimate y^f, and a likelihood function $f(d_{obs}|y)$ for the measurement d_{obs}, the Bayes' theorem expresses the following,

$$f(y|d_{obs}) \propto f(d_{obs}|y)f(y). \tag{2.16}$$

Therefore, the posterior density function for the estimate of y given the measurement d_{obs}, is proportional to the product of the prior density function $f(y)$ times the likelihood function for the measurement d_{obs}.

Similarly in sec.2.1.1, consider the two estimates Eq. (2.1) and Eq. (2.2) of the true state y_{true}. For the case with Gaussian statistics we can define both the prior and the likelihood functions as

$$f(y) \propto \exp\left(-\frac{1}{2}(y-y^f)(C_{yy}^f)^{-1}(y-y^f)\right) \tag{2.17}$$

and

$$f(d_{obs}|y) \propto \exp\left(-\frac{1}{2}(y-d_{obs})(C_{\varepsilon\varepsilon})^{-1}(y-d_{obs})\right) \tag{2.18}$$

Thus, the posterior density function can be written as

$$f(y|d_{obs}) \propto \exp\left(-\frac{1}{2}J[y]\right), \tag{2.19}$$

where

$$J[y] = (y-y^f)(C_{yy}^f)^{-1}(y-y^f) + (y-d_{obs})(C_{\varepsilon\varepsilon})^{-1}(y-d_{obs}) \tag{2.20}$$

The least squares solution, y^a, of the above equation, which is a minimum for J, is equivalent to a maximum of the posterior density function $f(y|d_{obs})$. In other words, it is

the maximum a posterior estimate in this case. This will always hold as long as all the error terms can be described by the Gaussian distribution.

The minimum value of J is found from

$$\frac{dJ[y]}{dy} = 2\left(y - y^f\right)\left(C_{yy}^f\right)^{-1} + 2\left(y - d_{obs}\right)\left(C_{\varepsilon\varepsilon}\right)^{-1} = 0\,. \tag{2.21}$$

Solving for y^a gives

$$\begin{aligned}
&\left(\left(C_{yy}^f\right)^{-1} + \left(C_{\varepsilon\varepsilon}\right)^{-1}\right)y^a = y^f\left(C_{yy}^f\right)^{-1} + \left(C_{\varepsilon\varepsilon}\right)^{-1} d_{obs} \\
&\Leftrightarrow \left(\left(C_{\varepsilon\varepsilon}\right) + \left(C_{yy}^f\right)\right)y^a = y^f\left(C_{\varepsilon\varepsilon}\right) + y^f\left(C_{yy}^f\right) - y^f\left(C_{yy}^f\right) + \left(C_{yy}^f\right)d_{obs} \\
&\Leftrightarrow y^a = y^f + \frac{C_{yy}^f}{C_{\varepsilon\varepsilon} + C_{yy}^f}(d_{obs} - y^f)
\end{aligned} \tag{2.22}$$

This is the same result as in Eq. (2.14) derived from the minimum variance estimate for the analyzed state. Thus the minimum variance estimate is also the maximum a posteriori estimate in the case of Gaussian priors.

In summary, the solution for the Kalman filter is based on maximizing the posterior probability density function (PDF) of the state estimation in Bayesian formulation. And it is equivalent to minimizing the error variances of the posterior state estimate with the assumption that the following variables are Gaussian (Maybeck (1979), Anderson (1979)).

- Model errors,
- Measurement errors, and

• The initial state estimate error.

In addition to the Gaussian assumptions, the model and measurement errors are also assumed to be unbiased and uncorrelated in time. And it is worthwhile to note that for a linear dynamic model and the estimate of the initial state is Gaussian, both the prior and posterior PDFs would be Gaussian because product of prior and likelihood Gaussian distribution results in a Gaussian distribution. Both the nonlinear dynamics and violation of the Gaussianity of the prior state values can not preserve the Gaussian statistics in the posterior PDF. In other words, the mean and covariance are sufficient to describe a Gaussian PDF, but for Non-Gaussian PDF, however, the mean and covariance are incomplete description of PDF. In such cases, although all the calculations of the Kalman filter can still be applied, the analyzed estimate would be sub-optimal.

2.1.3 Extension to Multivariate Case

Now, we extend the scalar state variable y to the spatial dimensional state vector $\mathbf{y}(\mathbf{x})$ with $\mathbf{x} = (x, y, z)$ for a three dimensional space.

Assume a multidimensional variable (e.g. a pressure field), and a vector of measurements $\mathbf{d}_{obs}$, which is related to the true state through the measurement matrix $\mathbf{H}$:

$$\mathbf{y}^f = \mathbf{y}^t + \mathbf{p}^f \quad , \tag{2.23}$$

$$\mathbf{d}_{obs} = \mathbf{H}\mathbf{y}^t + \boldsymbol{\varepsilon} \, , \tag{2.24}$$

The actual values of the errors $\mathbf{p}^f$ and $\boldsymbol{\varepsilon}$ are not known. Thus, we make a statistical hypothesis with the following assumption.

$$\begin{aligned} &\overline{\mathbf{p}^f} = \mathbf{0}, && \overline{\mathbf{p}^f (\mathbf{p}^f)^{\mathbf{T}}} = \mathbf{C}_{\mathbf{yy}}^f \\ &\bar{\boldsymbol{\varepsilon}} = \mathbf{0}, && \overline{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathbf{T}}} = \mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} \\ &\overline{\mathbf{p}^{\mathbf{f}} \boldsymbol{\varepsilon}^{\mathbf{T}}} = \mathbf{0} \end{aligned} \qquad (2.25)$$

Similar to the derivation of the scalar case, we seek a linear unbiased estimator for $\mathbf{y}_k^t(\mathbf{x})$ as

$$\mathbf{y}_k^a = \mathbf{y}_k^f + \mathbf{K}_k(\mathbf{d}_{obs} - \mathbf{H}\mathbf{y}_k^f) \qquad (2.26)$$

and the error in the analysis is from Eq. (2.23) and Eq.(2.24) given as

$$\mathbf{p}_k^a = \mathbf{p}_k^f + \mathbf{K}_k(\boldsymbol{\varepsilon} - \mathbf{H}\mathbf{p}_k^f) \qquad (2.27)$$

where the matrix $\mathbf{K}_k$ is often called the Kalman gain. This Kalman Gain matrix can be derived by minimizing the following error variance,

$$\begin{aligned} &\mathbf{C}_{\mathbf{yy}}^a(t_k) = \overline{\mathbf{p}_k^a (\mathbf{p}_k^a)^{\mathbf{T}}} \\ &= \overline{\left(\mathbf{p}_k^f + \mathbf{K}_k(\boldsymbol{\varepsilon} - \mathbf{H}\mathbf{p}_k^f)\right)\left(\mathbf{p}_k^f + \mathbf{K}_k(\boldsymbol{\varepsilon} - \mathbf{H}\mathbf{p}_k^f)\right)^{\mathbf{T}}} \\ &= \overline{\left(\mathbf{p}_k^f\right)\left(\mathbf{p}_k^f\right)^{\mathbf{T}}} - \overline{\mathbf{p}_k^f \left(\mathbf{K}_k \mathbf{H}\mathbf{p}_k^f\right)^{\mathbf{T}}} - \overline{\mathbf{K}_k \mathbf{H}\mathbf{p}_k^f \left(\mathbf{p}_k^f\right)^{\mathbf{T}}} + \overline{\mathbf{K}_k(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathbf{T}})\mathbf{K}_k^{\mathbf{T}}} + \overline{\mathbf{K}_k \mathbf{H}\mathbf{p}_k^f \left(\mathbf{p}_k^f\right)^{\mathbf{T}} \mathbf{H}^{\mathbf{T}} \mathbf{K}_k^{\mathbf{T}}} \\ &= \mathbf{C}_{\mathbf{yy}}^f(t_k) - \mathbf{C}_{\mathbf{yy}}^f(t_k)\mathbf{H}^{\mathbf{T}}\mathbf{K}_k^{\mathbf{T}} - \mathbf{K}_k \mathbf{H}\mathbf{C}_{\mathbf{yy}}^f(t_k) + \mathbf{K}_k\left(\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} + \mathbf{H}\mathbf{C}_{\mathbf{yy}}^f \mathbf{H}^{\mathbf{T}}\right)\mathbf{K}_k^{\mathbf{T}} \\ &= \mathbf{C}_{\mathbf{yy}}^f(t_k) - 2\mathbf{K}_k \mathbf{H}\mathbf{C}_{\mathbf{yy}}^f(t_k) + \mathbf{K}_k\left(\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} + \mathbf{H}\mathbf{C}_{\mathbf{yy}}^f \mathbf{H}^{\mathbf{T}}\right)\mathbf{K}_k^{\mathbf{T}} \end{aligned} \qquad (2.28)$$

We have used the matrix multiplication property $(\mathbf{AB})^{\mathbf{T}} = \mathbf{B}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}$. And $\mathbf{C}_{\mathbf{yy}}^{f}(t_k)$ is a symmetric matrix, $\left(\mathbf{C}_{\mathbf{yy}}^{f}(t_k)\right)^{\mathbf{T}} = \mathbf{C}_{\mathbf{yy}}^{f}(t_k)$.

Since we want to find an optimal $\mathbf{K}_k$, we minimize the individual terms along the major diagonal of $\mathbf{C}_{\mathbf{yy}}^{a}(t_k)$, because these terms represent the estimation error variance for the elements of the analysis state vector. We apply the matrix properties as shown below

$$\begin{aligned} \frac{d[trace(\mathbf{AB})]}{d\mathbf{A}} &= \mathbf{B}^{\mathbf{T}} \\ \frac{d[trace(\mathbf{ACA}^{\mathbf{T}})]}{d\mathbf{A}} &= 2\mathbf{AC}^{\mathbf{T}} \end{aligned} \tag{2.29}$$

From the Eq. (2.28) and (2.29), we get

$$\frac{d[trace\ \mathbf{C}_{\mathbf{yy}}^{a}(t_k)]}{d\mathbf{K}_k} = -2\left(\mathbf{HC}_{\mathbf{yy}}^{f}(t_k)\right)^{\mathbf{T}} + 2\mathbf{K}_k(\mathbf{HC}_{\mathbf{yy}}^{f}(t_k)\mathbf{H}^{t} + \mathbf{C}_{\boldsymbol{\varepsilon\varepsilon}}) \tag{2.30}$$

Now we set the derivative equation (2.30) equal to zero and solve for the optimal Kalman gain $\mathbf{K}_k$ as

$$\mathbf{K}_k = \mathbf{C}_{\mathbf{yy}}^{f}(t_k)\mathbf{H}^{\mathbf{T}}(\mathbf{HC}_{\mathbf{yy}}^{f}(t_k)\mathbf{H}^{t} + \mathbf{C}_{\boldsymbol{\varepsilon\varepsilon}})^{-1}. \tag{2.31}$$

Substituting Eq.(2.31) to Eq.(2.28), the posterior error estimate becomes

$$\begin{aligned}\mathbf{C}_{\mathbf{yy}}^{a}(t_k) &= \mathbf{C}_{\mathbf{yy}}^{f}(t_k) - 2\mathbf{K}_k\mathbf{H}\mathbf{C}_{\mathbf{yy}}^{f}(t_k) + \mathbf{C}_{\mathbf{yy}}^{f}(t_k)\mathbf{H}^{\mathbf{T}}\mathbf{K}_k^{\mathbf{T}} \\ &= \mathbf{C}_{\mathbf{yy}}^{f}(t_k) - 2\mathbf{K}_k\mathbf{H}\mathbf{C}_{\mathbf{yy}}^{f}(t_k) + \left(\mathbf{H}\mathbf{C}_{\mathbf{yy}}^{f}(t_k)\right)^{\mathbf{T}}\mathbf{K}_k^{\mathbf{T}} \\ &= \mathbf{C}_{\mathbf{yy}}^{f}(t_k) - \mathbf{K}_k\mathbf{H}\mathbf{C}_{\mathbf{yy}}^{f}(t_k) \\ &= (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{C}_{\mathbf{yy}}^{f}(t_k) \end{aligned} \tag{2.32}$$

To summarize, the multivariate Kalman Filter equations are as follows,

The unbiased linear estimate equation,

$$\mathbf{y}_k^a = \mathbf{y}_k^f + \mathbf{K}_k(\mathbf{d}_{obs} - \mathbf{H}\mathbf{y}_k^f),$$

with error estimate covariance matrix

$$\mathbf{C}_{\mathbf{yy}}^{a}(t_k) = (1 - \mathbf{K}_k\mathbf{H})\mathbf{C}_{yy}^{f}(t_k),$$

where the optimal Kalman Gain Matrix

$$\mathbf{K}_k = \mathbf{C}_{yy}^{f}(t_k)\mathbf{H}^{\mathbf{T}}(\mathbf{H}\mathbf{C}_{yy}^{f}(t_k)\mathbf{H}^{\mathbf{T}} + \mathbf{C}_{\varepsilon\varepsilon})^{-1}.$$

2.2 Time-dependent Sequential Model Updating Case

This section deals with time dependant problems where sequential data assimilation methods use the analysis scheme to sequentially update the model state. Such methods have proven useful for many applications in meteorology and oceanography where new observations are sequentially assimilated into the model when

they become available. We start with the linear dynamics case where the state variables can be predicted in time by linear governing equations. We then extend to the nonlinear dynamics by introducing the Extended Kalman Filter (EKF) approach.

2.2.1 Kalman Filter for a Scalar Case with Linear Dynamics

For linear dynamics, the optimal sequential assimilation method is the Kalman filter. In the Kalman filter, the first and the second-order statistical moment (mean and covariance) are evolved forward in time to predict error statistics for the model forecast. The error statistics are applied to calculate a variance minimizing estimate as measurements becomes available.

Assume now that a discrete linear dynamical model for the true state of a scalar y_{true} can be expressed as

$$y_{true}(t_k) = G y_{true}(t_{k-1}) + q(t_{k-1}) \tag{2.33}$$

$$y_{true}(t_0) = y_{init} + a \tag{2.34}$$

$$d_{obs} = y_{true} + \varepsilon \tag{2.35}$$

where G is a linear model operator, q is the model error over one time step and $y_{true}(t_k)$ is the true state at time t_k, $\mathbf{y}_{init}$ is an initial state condition with error a. The model error is normally not known and a numerical model will therefore evolve according to

$$y^f(t_k) = G y^a(t_{k-1}) \tag{2.36}$$

$$y^a(t_0) = y_{init}\,. \tag{2.37}$$

This means, given a best estimate y^a, for y at time t_{k-1}, a forecast y^f is calculated at time t_k by using the approximate Eq. (2.36). Now subtract Eq. (2.36) from (2.33) to get

$$y_k^t - y_k^f = G(y_{k-1}^t - y_{k-1}^a) + q_{k-1} \tag{2.38}$$

where we have defined $y_k^t = y_{true}(t_k)$ and $y_k^f = y^f(t_k)$ and $q_{k-1} = q(t_{k-1})$. The error covariance matrix for the forecast at time t_k is

$$\begin{aligned} C_{yy}^f(t_k) &= \overline{(y_k^t - y_k^f)^2} \\ &= G^2\overline{(y_{k-1}^t - y_{k-1}^a)^2} + \overline{q_{k-1}^2} + 2G\overline{(y_{k-1}^t - y_{k-1}^a)q_{k-1}} \\ &= G^2 C_{yy}^a(t_{k-1}) + C_{qq}(t_{k-1})\,. \end{aligned} \tag{2.39}$$

We have defined the error covariance for the model state as

$$C_{yy}^a(t_k) = \overline{(y_k^t - y_k^a)^2}\,, \tag{2.40}$$

the model error covariance as

$$C_{qq}(t_{k-1}) = \overline{q_{k-1}^2}\,, \tag{2.41}$$

and the initial error covariance as

$$C_{yy}(t_0) = C_{aa} = \overline{a^2}\,. \tag{2.42}$$

It is also assumed that there are no correlations between the error in the state, $y_{k-1}^t - y_{k-1}^a$,the model error q_{k-1}, and the initial error a.Thus, we have a consistent set of dynamical equations for the model evolution Eq. (2.36) and the initial condition Eq.(2.37), and the error covariance evolution Eq. (2.39), the model error covariance Eq.(2.41) and the initial error covariance Eq.(2.42). At times when there are measurements available, an analyzed state estimate can be calculated using Eq. (2.14) and Eq. (2.15) as

$$y_k^a = y_k^f + \frac{C_{yy}^f(t_k)}{C_{\varepsilon\varepsilon} + C_{yy}^f(t_k)}(d_{obs} - y^f(t_k))$$
$$C_{\mathbf{yy}}^a(t_k) = C_{yy}^f(t_k)\left(1 - \frac{C_{yy}^f(t_k)}{C_{\varepsilon\varepsilon} + C_{yy}^f(t_k)}\right).$$

When there are no measurements available, we just set $y_k^a = y_k^f$ and $C_{yy}^a(t_k) = C_{yy}^f(t_k)$, and proceed to the next updating. Theses equations define the Kalman filter for a linear scalar model, and it is the optimal sequential data assimilation method for the linear model given that the priors are all Gaussian and unbiased.

2.2.2 Kalman Filter for a Vector Case with Linear Dynamics

Similar to sec 2.1.3, the extension to multiple dimensions of the true state $y_{true}(\mathbf{x})$ can be represented by the state vector $\mathbf{y}^t$. It is assumed that the true state advances according to a dynamical model.

$$\mathbf{y}_k^t(\mathbf{x}) = \mathbf{G}\mathbf{y}_{k-1}^t(\mathbf{x}) + \mathbf{q}_{k-1}, \tag{2.43}$$

where **G** is a linear model operator (matrix) and $\mathbf{q}$ is the unknown model error over one time step. And a vector of measurement $\mathbf{d}_{obs}$ is defined by

$$\mathbf{d}_{obs} = \mathbf{H}\mathbf{y}_k^t(\mathbf{x}) + \boldsymbol{\varepsilon} \tag{2.44}$$

where **H** is a measurement matrix to relate the true state vector to the measurement values and $\boldsymbol{\varepsilon}$ is a vector of measurement error. In this case, a numerical model will forecast the state vector in time by

$$\mathbf{y}_k^f(\mathbf{x}) = \mathbf{G}\mathbf{y}_{k-1}^a(\mathbf{x}). \tag{2.45}$$

This means, given the best possible estimate for $\mathbf{y}_{k-1}^f$ at time t_{k-1}, a forecast is calculated at time t_k by using the approximate Eq. (2.45).The forecast error covariance equation is derived using a similar procedure as was used for Eq.(2.39) and becomes

$$\mathbf{C}_{\mathbf{yy}}^{f}(t_k) = \overline{(\mathbf{y}_k^t - \mathbf{y}_k^f)^2}$$
$$= \mathbf{G}\mathbf{C}_{\mathbf{yy}}^{a}(t_{k-1})\mathbf{G}^{\mathbf{T}} + \mathbf{C}_{\mathbf{qq}}(t_{k-1}) \quad . \tag{2.46}$$

Thus, the standard Kalman filter consists of the dynamical state forecast Eq. (4.45) and forecast error Eq.(2.46) together with the analysis equations which are derived in Eq.(2.26), (2.31), and (2.32) in the sec.2.1.3.

2.2.3 Extended Kalman Filter for Nonlinear Dynamics

For nonlinear dynamics, the extended Kalman filter (EKF) may be applied, in which an approximate linearized equation is used for the prediction of error statistics. In the scalar case, assume now that we have a nonlinear dynamical scalar model

$$y_k^t = g(y_{k-1}^t) + q_{k-1} \tag{2.47}$$

where $g(y)$ is a nonlinear model operator and q is again the unknown model error over one step. This numerical model will evolve according to the approximate equation as

$$y_k^f = g(y_{k-1}^a) . \tag{2.48}$$

Subtracting Eq. (2.48) from Eq. (2.47) gives

$$y_k^t - y_k^f = g(y_{k-1}^t) - g(y_{k-1}^a) + q_{k-1} . \tag{2.49}$$

Now use the Taylor expansion in the above equation as

$$g(y_{k-1}^t) = g(y_{k-1}^a) + g'(y_{k-1}^a)(y_{k-1}^t - y_{k-1}^a) + \frac{1}{2} g''(y_{k-1}^a)(y_{k-1}^t - y_{k-1}^a)^2 + \cdots \qquad (2.50)$$

Substituting Eq. (2.50) in Eq. (2.49), we get

$$y_k^t - y_k^f = g'(y_{k-1}^a)(y_{k-1}^t - y_{k-1}^a) + \frac{1}{2} g''(y_{k-1}^a)(y_{k-1}^t - y_{k-1}^a)^2 + \cdots + q_{k-1}. \qquad (2.51)$$

By squaring and taking the expected value, the equation for the evolution of the forecast error covariance $\mathbf{C}_{yy}^f(t_k)$ becomes

$$\begin{aligned} \mathbf{C}_{yy}^f(t_k) &= \overline{(y_k^t - y_k^f)^2} \\ &= \overline{(y_{k-1}^t - y_{k-1}^a)^2}\left(g'(y_{k-1}^a)\right)^2 \\ &+ \overline{(y_{k-1}^t - y_{k-1}^a)^3} g'(y_{k-1}^a) g''(y_{k-1}^a) \\ &+ \frac{1}{4}\overline{(y_{k-1}^t - y_{k-1}^a)^4}\left(g''(y_{k-1}^a)\right)^2 + \cdots + C_{qq}(t_{k-1}). \end{aligned} \qquad (2.52)$$

This equation can be closed by discarding moments of third and higher order, which result in an approximate equation for the forecast error covariance as

$$\mathbf{C}_{yy}^f(t_k) \approx \mathbf{C}_{yy}^a(t_{k-1})\left(g'(y_k^a)\right)^2 + C_{qq}(t_{k-1}). \qquad (2.53)$$

Together with the equations for the analyzed state estimate Eq. (2.14) and the analyzed error covariance Eq. (2.15) and the dynamical forecast Eq. (2.48) and forecast error Eq. (2.53) constitute the extended Kalman filter (EKF) in the case of a scalar state variable for nonlinear dynamical model.

In Matrix form, similar to the scalar case, assume a nonlinear model, but now the true state vector at time t_k is calculated from

$$\mathbf{y}_k^t = \mathbf{g}(\mathbf{y}_{k-1}^t) + \mathbf{q}_{k-1} \tag{2.54}$$

and a forecast is calculated from the approximate equation

$$\mathbf{y}_k^f = \mathbf{g}(\mathbf{y}_{k-1}^a) \tag{2.55}$$

where the model is now dependant on both time and space. The error statistics are then described by the error covariance matrix $\mathbf{C}_{\mathbf{yy}}^f(t_k)$ which evolves according to the equation with the assumption that the contributions from all the higher order statistical moments and higher order derivative of the nonlinear model operator are negligible as

$$\mathbf{C}_{\mathbf{yy}}^f(t_k) \approx \mathbf{G}'_{k-1}\mathbf{C}_{\mathbf{yy}}^a(t_{k-1})\mathbf{G}'^T_{k-1} + \mathbf{C}_{\mathbf{qq}}(t_{k-1}) \tag{2.56}$$

where $\mathbf{C}_{\mathbf{qq}}(t_{k-1})$ is the model error covariance matrix, $\mathbf{G}'_{k-1}$ is the Jacobi matrix or tangent linear operator,

$$\mathbf{G}'_{k-1} = \left.\frac{\partial \mathbf{g}(\mathbf{y})}{\partial \mathbf{y}}\right|_{k-1} . \tag{2.57}$$

2.3 Ensemble Kalman Filter

Another sequential data assimilation method is the Ensemble Kalman Filter (EnKF). The method was originally proposed as a stochastic or Monte Carlo alternative to the deterministic Extended Kalman Filter (EKF) by Evensen (1994a). The EnKF was designed to resolve the two major problems which arise from the EKF application for nonlinear dynamics in large state spaces: the use of an approximation of the error covariance evolution and the huge computational requirements associated with the memory storage and the calculation of the sensitivity matrix for the forecast error covariance matrix.

The EnKF has gained popularity because of its simple conceptual formulation and its relatively easy implementation. It does not require derivation of the sensitivity matrix or adjoint equations and integrations backwards in time. In addition to that, especially for reservoir problems, it is easily connected to existing commercial reservoir simulators because of the independent process of EnKF.

2.3.1 Sampling Representation of Error Statistics

The error covariance matrices for the predicted and the analyzed estimate, $\mathbf{C}_{\mathbf{yy}}^{f}$ and $\mathbf{C}_{\mathbf{yy}}^{a}$, are defined in terms of the true state as

$$\mathbf{C}_{\mathbf{yy}}^{f} = \overline{\left(\mathbf{y}^{f} - \mathbf{y}^{t}\right)\left(\mathbf{y}^{f} - \mathbf{y}^{t}\right)^{\mathbf{T}}}, \tag{2.58}$$

$$\mathbf{C}_{\mathbf{yy}}^{a} = \overline{\left(\mathbf{y}^{a} - \mathbf{y}^{t}\right)\left(\mathbf{y}^{a} - \mathbf{y}^{t}\right)^{\mathbf{T}}}, \tag{2.59}$$

where the over line denotes the ensemble averaging. However, the true state is not known, and we therefore define the ensemble covariance matrices around the ensemble mean $\overline{\mathbf{y}}$,

$$\left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^{f} = \overline{\left(\mathbf{y}^{f} - \overline{\mathbf{y}^{f}}\right)\left(\mathbf{y}^{f} - \overline{\mathbf{y}^{f}}\right)^{\mathbf{T}}}, \tag{2.60}$$

$$\left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^{a} = \overline{\left(\mathbf{y}^{a} - \overline{\mathbf{y}^{a}}\right)\left(\mathbf{y}^{a} - \overline{\mathbf{y}^{a}}\right)^{\mathbf{T}}}, \tag{2.61}$$

where the over line denotes an average over the ensemble. Thus, we can infer from Eq. (2.60) and Eq.(2.61) that ensemble mean is the best estimate and the spread of the ensemble around the mean defines the error in the ensemble mean.

2.3.2 Analysis Scheme

The Kalman filter analysis scheme utilizes the definition of $\mathbf{C}_{\mathbf{yy}}^{f}$ and $\mathbf{C}_{\mathbf{yy}}^{a}$ as given by Eq. (2.58) and Eq.(2.59). We will now derive the analysis scheme for EnKF using the ensemble covariances as defined by Eq. (2.60) and Eq. (2.61).

As was shown by Burgers et al.(1998) it is essential that the observations are treated as random variables having a distribution with mean equal to the first guess

observations and covariance equal to $\mathbf{C}_{\varepsilon\varepsilon}$. Thus, we can define an ensemble of observations by

$$\mathbf{d}_j = \mathbf{d}_{obs} + \boldsymbol{\varepsilon}_j \tag{2.62}$$

where j ranges from 1 to the total number of ensemble members N_e. It is ensured that the generated random measurement errors have mean equal to zero. Next we define the ensemble covariance matrix of the measurement error as

$$\mathbf{C}_{\varepsilon\varepsilon}^{e} = \overline{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathbf{T}}}. \tag{2.63}$$

The analysis step in the EnKF consists of the following updates performed on each of the ensemble model state, which is based on the Kalman Equation derived in Eq.(2.26) as follows,

$$\mathbf{y}_j^a = \mathbf{y}_j^f + \left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^f \mathbf{H}^{\mathbf{T}} (\mathbf{H}\left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^f \mathbf{H}^{\mathbf{T}} + \mathbf{C}_{\boldsymbol{\varepsilon\varepsilon}}^{e})^{-1} (\mathbf{d}_j - \mathbf{H}\mathbf{y}_j^f) \tag{2.64}$$

with j denotes ensemble model number. With finite number of the ensemble model, this equation will be an approximation. Averaging Eq.(2.64) over the ensemble model, we get

$$\overline{\mathbf{y}^a} = \overline{\mathbf{y}^f} + \left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^f \mathbf{H}^{\mathbf{T}} (\mathbf{H}\left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^f \mathbf{H}^{\mathbf{T}} + \mathbf{C}_{\boldsymbol{\varepsilon\varepsilon}}^{e})^{-1} (\bar{\mathbf{d}} - \mathbf{H}\overline{\mathbf{y}^f}), \tag{2.65}$$

where $\bar{\mathbf{d}} = \mathbf{d}_{obs}$, therefore the relation between the analyzed and predicted ensemble mean state is identical to the relation between the analyzed and predicted one in the standard Kalman filter, except for the use of $\left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^{f}$ and $\mathbf{C}_{\boldsymbol{\varepsilon\varepsilon}}^{e}$ instead of $\mathbf{C}_{\mathbf{yy}}^{f}$ and $\mathbf{C}_{\boldsymbol{\varepsilon\varepsilon}}$.

Now, we derive the analyzed error covariance estimate resulting from the analysis scheme given above. From Eq. (2.64) and (2.65), we can obtain

$$\mathbf{y}_j^a - \overline{\mathbf{y}^a} = (\mathbf{I} - \mathbf{K}^e\mathbf{H})(\mathbf{y}_j^f - \overline{\mathbf{y}^f}) + \mathbf{K}^e(\mathbf{d}_j - \bar{\mathbf{d}}), \tag{2.66}$$

where we have used the definition of the Kalman Gain as

$$\mathbf{K}^e = \left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^{f}\mathbf{H}^{\mathbf{T}}(\mathbf{H}\left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^{f}\mathbf{H}^{\mathbf{T}} + \mathbf{C}_{\boldsymbol{\varepsilon\varepsilon}}^{e})^{-1}. \tag{2.67}$$

The derivation of the analysis error covariance is below,

$$\begin{aligned}
\left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^{a} &= \overline{\left(\mathbf{y}_j^a - \overline{\mathbf{y}^a}\right)\left(\mathbf{y}_j^a - \overline{\mathbf{y}^a}\right)^{\mathbf{T}}} \\
&= \overline{\left((\mathbf{I} - \mathbf{K}^e\mathbf{H})(\mathbf{y}_j^f - \overline{\mathbf{y}^f}) + \mathbf{K}^e(\mathbf{d}_j - \bar{\mathbf{d}})\right)\left((\mathbf{I} - \mathbf{K}^e\mathbf{H})(\mathbf{y}_j^f - \overline{\mathbf{y}^f}) + \mathbf{K}^e(\mathbf{d}_j - \bar{\mathbf{d}})\right)^{\mathbf{T}}} \\
&= (\mathbf{I} - \mathbf{K}^e\mathbf{H})\overline{(\mathbf{y}_j^f - \overline{\mathbf{y}^f})(\mathbf{y}_j^f - \overline{\mathbf{y}^f})^{\mathbf{T}}}(\mathbf{I} - \mathbf{K}^e\mathbf{H})^{\mathbf{T}} + \mathbf{K}^e\overline{(\mathbf{d}_j - \bar{\mathbf{d}})(\mathbf{d}_j - \bar{\mathbf{d}})^{\mathbf{T}}}\mathbf{K}^{e\,\mathbf{T}} \\
&= (\mathbf{I} - \mathbf{K}^e\mathbf{H})\left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^{f}(\mathbf{I} - \mathbf{K}^e\mathbf{H})^{\mathbf{T}} + \mathbf{K}^e\mathbf{C}_{\boldsymbol{\varepsilon\varepsilon}}^{e}\mathbf{K}^{e\,\mathbf{T}} \\
&= \left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^{f} - \mathbf{K}^e\mathbf{H}\left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^{f} - \left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^{f}\mathbf{H}^{\mathbf{T}}\mathbf{K}^{e\,\mathbf{T}} + \mathbf{K}^e\left(\mathbf{H}\left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^{f}\mathbf{H}^{\mathbf{T}} + \mathbf{C}_{\boldsymbol{\varepsilon\varepsilon}}^{e}\right)\mathbf{K}^{e\,\mathbf{T}} \\
&= \left(\mathbf{I} - \mathbf{K}^e\mathbf{H}\right)\left(\mathbf{C}_{\mathbf{yy}}^{e}\right)^{f}
\end{aligned} \tag{2.68}$$

Note that the derivation is the same as the Kalman filter in Eq. (2.28) in the sec.2.1.3. This implies that the EnKF in the limit of an infinite ensemble size will give the exactly

the same result as KF and EKF. Finally, it should be noted that the EnKF analysis scheme is sub-optimal for non-Gaussian prior PDF for **y** because it does not fully take into account non-Gaussian contributions in the linear updating**.** Thus, the solution is something between a linear Gaussian update and a full Bayesian update for the non Gaussian prior PDFs.

2.4 EnKF for History Matching in Reservoir Problem

In the previous sections, we formulate the EnKF in the scope of the state estimation problem. But in the reservoir characterization framework, our model parameters (e.g. permeability and porosity) are poorly known. We need to estimate the parameters as well as state variables (e.g. Pressure and Saturation). This problem falls in the joint parameter and state estimation problem or combined parameter and state estimation problem. Here we show how we apply the EnKF for the reservoir history matching problem in the context of the combined parameter and state estimation problem.

2.4.1 History Matching Terminology for EnKF

Let's clarify terminology that we use in the reservoir history matching problem for EnKF as follows.

Model Parameters

These are variables that are uncertain and time independent. These variables are also called ***static model variables***. They include rock properties such as porosity ϕ and

absolute permeability K. For EnKF updating, the estimate of these properties changes as data are integrated although the parameters themselves should not be interpreted to be changing with time. We denote the static parameters as $\mathbf{m}^s$.

State Variables

These are variables that are uncertain and time dependent and define the state of the system. Because of the time dependence, they are also called ***dynamic model variables***. The uncertainty in these variables results from the uncertainty in the model parameters and some other uncertain factors such as initial conditions. For reservoir problems, state variables could include phase pressures (p_o, p_g, p_w) and phase saturations (S_o, S_g, S_w). These variables are the solutions of systems of the governing 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. We denote the dynamic model variables as $\mathbf{m}^d$.

Data

These are observation data related directly or indirectly (linearly or nonlinearly) to the state variables and the model parameters. For reservoir problems, data ranges from the well measurement data including surface flow rates (production rate or injection rate), Ratio of the rate (Water Cut (WWCT) and Gas Oil Ratio (GOR)) or bottom-hole pressure (WBHP) to a spatial observation data such as seismic data. In reality, the observation data always have some unknown level of error or noise associated with them. We denote the observation data as $\mathbf{d}_{obs}$.

Forward Model

This is the dynamical model which advances the state of the system in time. Essentially it is the reservoir simulator which is solving the governing partial differential equations in the descretized time and space domain. In this study we use the commercial streamline reservoir simulator ***FrontSim*** by *Schlumberger*. We denote the Forward model as $g(\circ)$, which is a nonlinear operator for the model static and dynamic variables. If the forward model is perfect, we can relate the true model parameters to the observed data as

$$\mathbf{d}_{obs} = g(\mathbf{m}^{s}_{true}, \mathbf{m}^{d}_{true}) + \boldsymbol{\varepsilon} \tag{2.69}$$

where $\boldsymbol{\varepsilon}$ is the unknown measurement noise. It is assumed to be unbiased and Gaussian, $\boldsymbol{\varepsilon} \sim N(0, \mathbf{C_D})$ (i.e. $E[\boldsymbol{\varepsilon}] = 0$, and $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}] = \mathbf{C_D}$). $\mathbf{C_D}$ is the measurement error covariance matrix which is a diagonal matrix if the measurement errors are uncorrelated.

State Vector

State vector consists of model parameters $\mathbf{m}^{s}$, state variables $\mathbf{m}^{d}$ and calculated data $\mathbf{d}_{cal}$. The calculated data are the derived model responses from the simulation with the model parameters and state variables. This can be described as

$$\mathbf{d}_{cal} = g(\mathbf{m}^{s}, \mathbf{m}^{d}). \tag{2.70}$$

We define the state vector at time k as the augmented vector

$$\mathbf{y}_k = \begin{Bmatrix} \mathbf{m}_k^{\mathbf{s}} \\ \mathbf{m}_k^{\mathbf{d}} \\ \mathbf{d}_{cal,k} \end{Bmatrix} . \tag{2.71}$$

The calculated data in time *k* can be related with measurement matrix **H** as follows,

$$\mathbf{d}_{cal,k} = \mathbf{H}\mathbf{y}_k \tag{2.72}$$

where matrix **H** is simply define by the null matrix **0** and the identity matrix **I** as

$$\mathbf{H} = [\mathbf{0} \quad \mathbf{I}]. \tag{2.73}$$

This indicates the matrix **H** extracts the model responses $\mathbf{d}_{cal,k}$ from the augmented state vector $\mathbf{y}_k$ by the matrix multiplication.

2.4.2 Ensemble of the State Vector

EnKF uses an ensemble of state vectors instead of a single state vector. The statistics (mean and covariance) are then computed from the ensemble; the ensemble of state vectors can be represented by the equation below.

$$\mathbf{\Psi}_k^p = \left\{ \mathbf{y}_{k,1}^p \quad \mathbf{y}_{k,2}^p \quad \cdots \quad \mathbf{y}_{k,N_e}^p \right\} \tag{2.74}$$

where the superscript p denotes prior and subscript denotes time k and the index of the ensemble member and N_e is the ensemble size. Each initial state vector represents an initial model of an infinite ensemble of possible states that are consistent with initial measurement from core, well logs and seismic. The ensemble of initial models can be generated using any of the standard geostatistics techniques like sequential Gaussian simulation or indicator simulation.

2.4.3 Forecast Step and Update Step in EnKF

EnKF has two main steps: a forecast step and an update step. In this thesis, the forecast step is carried out by a commercial reservoir streamline simulator ***FrontSim***. This step can be represented as

$$\begin{Bmatrix} \mathbf{m}_k^d \\ \mathbf{d}_{cal,k} \end{Bmatrix} = g(\mathbf{m}_{k-1}^s, \mathbf{m}_{k-1}^d) \tag{2.75}$$

where the forward model operator $g(\circ)$ represents a numerical solution of the porous media fluid flow equation moving forward from time step k-1 to time step k. In the sec.2.3.2, we derive the optimal Kalman update equation in matrix from is given by Eq.(2.64). We rewrite the Eq. (2.64) by using the ensemble of state vector notation in Eq. (2.74) as

$$\boldsymbol{\Psi}_k^u = \boldsymbol{\Psi}_k^p + \mathbf{K}_k\left(\mathbf{D}_{obs,k} - \mathbf{H}\boldsymbol{\Psi}_k^p\right) \tag{2.76}$$

The superscript *u* denotes updated and *p* denotes prior.Here the matrix $\mathbf{K}_k$ is known as the Kalman gain, matrix $\mathbf{D}_{obs,k}$ represents an ensemble of perturbed observations as it is defined by equations as follows

$$\mathbf{D}_{obs,k} = \left\{\mathbf{d}_{obs,k,1} \quad \mathbf{d}_{obs,k,2} \quad \ldots \quad \mathbf{d}_{obs,k,N_e}\right\} \tag{2.77}$$

where

$$\mathbf{d}_{obs,k,j} = \mathbf{d}_{obs,k} + \boldsymbol{\varepsilon}_j \tag{2.78}$$

and $\mathbf{d}_{obs,k}$ represents a vector of any type of production data measured at time '*k*', and $\boldsymbol{\varepsilon}_j$ represent the noise in the observation data for member '*j*'. The Kalman gain matrix $\mathbf{K}_k$ is also rewritten from Eq. (2.67) as follows,

$$\mathbf{K}_k = \mathbf{C}_{\Psi,k}^p\mathbf{H}^{\mathbf{T}}\left(\mathbf{H}\mathbf{C}_{\Psi,k}^p\mathbf{H}^{\mathbf{T}} + \mathbf{C}_{\mathbf{D}}\right)^{-1} \tag{2.79}$$

where $\mathbf{C}_{\Psi,k}^p$ represents an estimate of the state vector covariance matrix at time '*k*'; and $\mathbf{C}_{\mathbf{D}}$ represents observation error covariance matrix; typically we assume the errors in the observation are not correlated, therefore, $\mathbf{C}_{\mathbf{D}}$ is simply a diagonal matrix. Since the true

state vector is unknown, as we described in sec.2.3.1, we approximate it with the mean of the ensemble using Eq. (2.60); the covariance matrix $\mathbf{C}_{\Psi,k}^{p}$ is rewritten as

$$\mathbf{C}_{\Psi,k}^{p}=\frac{1}{N_e-1}\sum_{i,j=1}^{N_e}(\mathbf{y}_{k,i}^{p}-\overline{\mathbf{y}}_{k}^{p})(\mathbf{y}_{k,j}^{p}-\overline{\mathbf{y}}_{k}^{p})^{\mathbf{T}} \tag{2.80}$$

where

$$\overline{\mathbf{y}}_{k}^{p}=\frac{1}{N_e}\sum_{j=1}^{N_e}\mathbf{y}_{k,j}^{p} \tag{2.81}$$

Since in the Kalman gain equation, this covariance matrix $\mathbf{C}_{\Psi}^{p}$ is always multiplied by matrix **H**, in practice there is not need to compute the whole covariance matrix but only a small portion of it. This is described in the following equations below. Inside the covariance matrix of the state vector, $\mathbf{C}_{\Psi,k}^{p}$ reveals

$$\mathbf{C}_{\Psi,k}^{p}=\begin{pmatrix}\mathbf{C}_{\mathbf{m}^s} & \mathbf{C}_{\mathbf{m}^s,\mathbf{m}^d} & \mathbf{C}_{\mathbf{m}^s,\mathbf{d}_{cal}}\\ \mathbf{C}^{\mathbf{T}}_{\mathbf{m}^d,\mathbf{m}^s} & \mathbf{C}_{\mathbf{m}^d} & \mathbf{C}_{\mathbf{m}^d,\mathbf{d}_{cal}}\\ \mathbf{C}^{\mathbf{T}}_{\mathbf{d}_{cal},\mathbf{m}^s} & \mathbf{C}^{\mathbf{T}}_{\mathbf{d}_{cal},\mathbf{m}^d} & \mathbf{C}_{\mathbf{d}_{cal}}\end{pmatrix}_{k}^{p} \tag{2.82}$$

where $\mathbf{C}_{\mathbf{m}^s}$ is the covariance matrix of the static variables, the size of this matrix is $M\times M$ where M is the number of grid blocks, and matrix $\mathbf{C}_{\mathbf{m}^d}$ is the covariance

matrix of the dynamic variables, the size of the matrix is $M \times M$ for one dynamic variable per one grid block; $\mathbf{C}_{\mathbf{d}_{cal}}$ is the covariance matrix of the calculated data variable, the size of the matrix is $N_d \times N_d$ where N_d is the number of measurement at a given time. Non-diagonal elements in the matrix $\mathbf{C}_{\Psi}^{p}$ such as $\mathbf{C}_{M^s,d}$ and $\mathbf{C}_{M^d,d}$ are the cross covariance matrix between the static variable and the calculated data. Thus, the term $\mathbf{C}_{\Psi}^{p}\mathbf{H}^T$ and $\mathbf{H}\mathbf{C}_{\Psi}^{p}\mathbf{H}^T$ in the Kalman Gain Matrix K is expressed in sub-matrices as

$$\mathbf{C}_{\Psi,k}^{p}\mathbf{H}^{\mathbf{T}} = \begin{pmatrix} \mathbf{C}_{\mathbf{m}^s} & \mathbf{C}_{\mathbf{m}^s,\mathbf{m}^d} & \mathbf{C}_{\mathbf{m}^s,\mathbf{d}_{cal}} \\ \mathbf{C}^{\mathbf{T}}_{\mathbf{m}^d,\mathbf{m}^s} & \mathbf{C}_{\mathbf{m}^d} & \mathbf{C}_{\mathbf{m}^d,\mathbf{d}_{cal}} \\ \mathbf{C}^{\mathbf{T}}_{\mathbf{d}_{cal},\mathbf{m}^s} & \mathbf{C}^{\mathbf{T}}_{\mathbf{d}_{cal},\mathbf{m}^d} & \mathbf{C}_{\mathbf{d}_{cal}} \end{pmatrix}_k^p \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{\mathbf{m}^s,\mathbf{d}_{cal}} \\ \mathbf{C}_{\mathbf{m}^d,\mathbf{d}_{cal}} \\ \mathbf{C}_{\mathbf{d}_{cal}} \end{pmatrix}_k^p \tag{2.83}$$

$$\mathbf{H}\mathbf{C}_{\Psi,k}^{p}\mathbf{H}^{\mathbf{T}} = \mathbf{C}_{\mathbf{d}_{cal},k}^{p} \tag{2.84}$$

If we replace the terms $\mathbf{C}_{\Psi,k}^{p}\mathbf{H}^T$ and $\mathbf{H}\mathbf{C}_{\Psi,k}^{p}\mathbf{H}^T$ into update equation (2.76), then for one model state vector $\mathbf{y}_k$ we can rewrite the equation as

$$\mathbf{y}_k^u = \begin{Bmatrix} \mathbf{m}_k^s \\ \mathbf{m}_k^d \\ \mathbf{d}_{cal,k} \end{Bmatrix}^u = \begin{Bmatrix} \mathbf{m}_k^s \\ \mathbf{m}_k^d \\ \mathbf{d}_{cal,k} \end{Bmatrix}^p + \begin{pmatrix} \mathbf{C}_{\mathbf{m}^s,\mathbf{d}} \\ \mathbf{C}_{\mathbf{m}^d,\mathbf{d}} \\ \mathbf{C}_{\mathbf{d}_{cal}} \end{pmatrix}_k^p \left(\mathbf{C}_{\mathbf{d}_{cal},k}^{p} + \mathbf{C}_d\right)^{-1}\left(\mathbf{d}_{k,obs} - Hg\left(\mathbf{y}_k^p\right)\right) \tag{2.85}$$

This equation indicates that we don't need to compute the whole covariance matrix elements in $\mathbf{C}_{\Psi,k}^{p}$ to update the state variables for each ensemble member in the update step of the EnKF. This significantly reduces the computational burden and the implementation becomes relatively easy. And posterior covariance matrix can be computed by

$$\mathbf{C}_{\Psi,k}^{u} = \mathbf{C}_{\Psi,k}^{p} - \mathbf{C}_{\Psi,k}^{p}\mathbf{H}^{\mathbf{T}}(\mathbf{H}\mathbf{C}_{\Psi,k}^{p}\mathbf{H}^{\mathbf{T}} + \mathbf{C}_{d})^{-1}\mathbf{H}^{\mathbf{T}}\mathbf{C}_{\Psi,k}^{p}\,. \tag{2.86}$$

2.5 Hybrid Ensemble Kalman Filter Formulations

Our proposed algorithm proceeds along the same lines as the conventional EnKF with an additional step as discussed below. The procedure is implemented as follows: Eq. (2.76) is the basis of the first step in the hybrid EnKF approach proposed here. We calculate intermediate posterior state vector $\tilde{\mathbf{y}}_{i}^{u}$ as below,

$$\tilde{\mathbf{y}}_{k,j}^{u} = \mathbf{y}_{k,j}^{p} + \mathbf{C}_{\Psi,k}^{p}\mathbf{H}^{\mathbf{T}}(\mathbf{H}\mathbf{C}_{\Psi,k}^{p}\mathbf{H}^{\mathbf{T}} + \mathbf{C}_{D})^{-1}(\mathbf{d}_{obs,j} - \mathbf{H}\mathbf{y}_{k,j}^{p}) \tag{2.87}$$

Then, the ensemble updated mean is specified by

$$\mathbf{y}_{mean}^{u} = \sum_{i=1}^{N_e}\tilde{\mathbf{y}}_{i}^{u} \quad . \tag{2.88}$$

Simultaneously the parameter estimation is carried via inversion for the ensemble updated mean model static parameter (e.g. permeability), $\mathbf{m}^{s}_{mean}$ of the state vector $\mathbf{y}^{u}_{mean}$. This ensemble mean is then used as a starting model in a deterministic inverse formulation to integrate the production history from the beginning (time zero). This is an important distinction because unlike the conventional EnKF, the inversion step here explicitly ensures that the updated model is consistent with all the previous observations.

The non-linear inversion is carried out by iterative minimization of an augmented misfit function, $F(\boldsymbol{\delta}\mathbf{m})$ to find the elements of the update vector $\boldsymbol{\delta}\mathbf{m}$. The details of the inversion procedure can be found elsewhere (Yoon et al. 1999; He et al. 2002; Tarantola 2005; Cheng et al. 2006; Oyerinde 2007). Specifically, we utilize a generalized travel time inversion (GTTI) technique for the inversion of the water-cut responses and a low-frequency pressure inversion algorithm (Vasco and Karasaki, 2006) to integrate the bottom-hole pressure data. Critical to our approach is the efficiency of the inversion algorithm. This is facilitated by a strteamline-based analytic computation of the water-cut sensitivities (Yoon et al. 1999) and a low frequency asymptotic solution of the diffusivity equation for the bottom-hole pressure sensitivity calculations (Kim et al., 2009). The linearized misfit function for the inversion can be written in terms of the sensitivity matrix **S**, as

$$F(\boldsymbol{\delta}\mathbf{m}) = \left\|\boldsymbol{\delta}\mathbf{d} - \mathbf{S}(\boldsymbol{\delta}\mathbf{m})\right\| + \beta_1\left\|\boldsymbol{\delta}\mathbf{m}\right\| + \beta_2\left\|\mathbf{L}(\boldsymbol{\delta}\mathbf{m})\right\| \tag{2.89}$$

$$\mathbf{m}^{k+1}_{\text{mode}} = \mathbf{m}^{k}_{\text{mode}} + \mu\boldsymbol{\delta}\mathbf{m}\,. \tag{2.90}$$

In Eq. (2.89), the first term on the RHS represents the data misfit, the second term is the 'norm' constraint which minimizes deviations from the prior model and the third term is the 'roughness' constraint that ensures large-scale continuity in the geologic model. The 'norm' and 'roughness' terms can be weighted appropriately based on our prior information and there is some guidelines in the literature to this regard (Parker 1994). An iterative least squares solver (LSQR) (Golub and Van Loan 1989) is used to solve Eq. (2.89) and (2.90) to obtain the updates and the updated solution or the posterior mode, $\mathbf{m}_{\text{mode}}$. Through the inversion process, we also ensure consistency between the dynamic state variables (e.g. pressure and phase saturation) and the updated model parameters (e.g. permeability). We then construct the state-vector associated with the inversion solution as $\mathbf{y}_{\text{mode}} = [\mathbf{m}^{s}_{\text{mode}}, \mathbf{m}^{d}_{\text{mode}}, \mathbf{d}_{\text{mode}}]^{\mathbf{T}}$ which replaces the ensemble mean in the EnKF formulation. This step ensures that the EnKF trajectory is centered on the posterior mode which is a better representation of the true state, particularly for non-linear model dynamics.

In this thesis, we propose two approaches for the hybrid EnKF: (1) Swapping of mean and mode estimation and (2) Coarse scale permeability constraint as discussed below.

2.5.1 Swapping of Mean and Mode Estimation

The Kalman Gain calculation in Eq.(2.79) again is

$$\mathbf{K} = \mathbf{C}_{\Psi}^{p}\mathbf{H}^{\mathbf{T}}(\mathbf{H}\mathbf{C}_{\Psi}^{p}\mathbf{H}^{\mathbf{T}} + \mathbf{C}_{D})^{-1}$$

where the covariance matrix estimation is from Eq.(2.80)

$$\mathbf{C}_{\Psi}^{p}\mathbf{H}^{\mathbf{T}} = \frac{1}{N_e - 1}\sum_{i,j=1}^{N_e}(\mathbf{y}_i^p - \overline{\mathbf{y}}^p)(\mathbf{H}\mathbf{y}_j^p - \mathbf{H}\overline{\mathbf{y}}^p)^{\mathbf{T}} \tag{2.91}$$

$$\mathbf{H}\mathbf{C}_{\Psi}^{p}\mathbf{H}^{\mathbf{T}} = \frac{1}{N_e - 1}\sum_{i,j=1}^{N_e}(\mathbf{H}\mathbf{y}_i - \mathbf{H}\overline{\mathbf{y}}^p)(\mathbf{H}\mathbf{y}_j - \mathbf{H}\overline{\mathbf{y}}^p)^{\mathbf{T}} \tag{2.92}$$

In our approach, the mean of the state variables (e.g. Permeability, Pressure and Saturation) is replaced by the mode values from the inversion results as follows.

$$\overline{\mathbf{y}}^p = \mathbf{y}_{\text{mode}} \tag{2.93}$$

$$\mathbf{H}\overline{\mathbf{y}}^p = \mathbf{H}\mathbf{y}_{\text{mode}} \tag{2.94}$$

These operations indicate that we assume the true estimation of the state and parameter variables are closer to the mode estimation than the ones from the ensemble averaging estimation. Especially from the parameter (e.g. permeability) estimation perspective, even for a non- Gaussian distribution such as multi modal distributions, the ensemble mean estimation tends to be more Gaussian because of the central limit theorem. Moreover, for the state (e.g. Pressure and Saturation) estimation perspective, by using the mode values from the inversion process, the parameter values and state values are consistent in the governing equation of the system while there is no physical consistency

between the ensemble mean parameters and the ensemble mean state variables for the nonlinear dynamical system.

2.5.2 Coarse Scale Permeability Constraint

This approach involves a coarse-scale constraint on the individual ensemble members based on the solution from the inversion. This is designed to drive the individual ensemble members towards the posterior mode. Based on the solution from the inversion process, we first conduct an upscaling of the model parameters to generate a coarse scale permeability 'data' as follows,

$$\mathbf{d}_{\ln k,k} = \mathbf{U}\mathbf{y}_{\text{mode}} \tag{2.95}$$

Similar to the measurement matrix $\mathbf{H}$ in Eq. (2.73), the matrix $\mathbf{U}$ is expressed as

$$\mathbf{U} = \left[\mathbf{u}_{N_s} \quad \mathbf{0}_{N_d} \quad \mathbf{0}_{M}\right]. \tag{2.96}$$

Note that the upscaling operator $\mathbf{u}$ is applied only for the model parameters, $\mathbf{m}^{s}_{\text{mode}}$. We adopt a flow-based upscaling method for $\mathbf{u}$ (See the Appendix A). The coarse-scale permeability derived from the inversion are treated as 'data' for a second assimilation sequence that further updates the ensemble members by minimizing a misfit function based on coarse-scale permeability data,

$$\mathbf{y}_{k,j}^{u} = \tilde{\mathbf{y}}_{k}^{u} + \tilde{\mathbf{C}}_{\Psi,k}^{u}\mathbf{U}^{\mathbf{T}}(\mathbf{U}\tilde{\mathbf{C}}_{\Psi,k}^{u}\mathbf{U}^{\mathbf{T}} + \mathbf{R})^{-1}(\mathbf{d}_{\ln k,k} - \mathbf{U}\tilde{\mathbf{y}}_{k,j}^{u}). \tag{2.97}$$

The precision of the coarse-scale observations is contained in the covariance matrix, **R**. By choosing large **R**, the impact of the second stage assimilation can be minimized, if needed, to prevent an ensemble collapse. In Eq. (2.97), $\mathbf{U}\tilde{\mathbf{y}}_{k,j}^{u}$ is the upscaled permeability derived from the fine-scale ensemble members after the first step assimilation and application of the flow-based upscaling. Similar to the first assimilation step, each covariance term is evaluated as follow,

$$\tilde{\mathbf{C}}_{\Psi,k}^{u} = \frac{1}{N_e - 1}\sum_{i,j=1}^{N_e}\left[\tilde{\mathbf{y}}_{k,i}^{u} - \mathbf{y}_{\text{mode}}\right]^{T}\left[\tilde{\mathbf{y}}_{k,j}^{u} - \mathbf{y}_{\text{mode}}\right] \tag{2.98}$$

$$\tilde{\mathbf{C}}_{\Psi,k}^{u}\mathbf{U}^{\mathbf{T}} = \frac{1}{N_e - 1}\sum_{i,j=1}^{N_e}(\tilde{\mathbf{y}}_{k,i}^{u} - \mathbf{y}_{\text{mode}})(\mathbf{U}\tilde{\mathbf{y}}_{k,j}^{u} - \mathbf{U}\mathbf{y}_{\text{mode}})^{\mathbf{T}} \tag{2.99}$$

$$\mathbf{U}\tilde{\mathbf{C}}_{\Psi,k}^{p}\mathbf{U}^{\mathbf{T}} = \frac{1}{N_e - 1}\sum_{i,j=1}^{N_e}(\mathbf{U}\tilde{\mathbf{y}}_{k,i}^{u} - \mathbf{U}\mathbf{y}_{\text{mode}})(\mathbf{U}\tilde{\mathbf{y}}_{k,j}^{u} - \mathbf{U}\mathbf{y}_{\text{mode}})^{\mathbf{T}} \tag{2.100}$$

It is worthwhile to note that the covariance calculation is now based not on the ensemble mean $\bar{\mathbf{y}}$ but the posterior mode from inversion, $\mathbf{y}_{\text{mode}}$. And similar to Eq. (2.86), the posterior covariance is calculated as,

$$\mathbf{C}_{\Psi,k}^{u} = \tilde{\mathbf{C}}_{\Psi,k}^{u} - \tilde{\mathbf{C}}_{\Psi,k}^{u}\mathbf{U}^{\mathbf{T}}(\mathbf{U}\tilde{\mathbf{C}}_{\Psi,k}^{u}\mathbf{U}^{\mathbf{T}} + \mathbf{R})^{-1}\mathbf{U}^{\mathbf{T}}\tilde{\mathbf{C}}_{\Psi,k}^{u} \tag{2.101}$$

In our approach, we chose to carry out the inversion on the ensemble mean directly in the fine-scale and impose the coarse-scale permeability (as opposed to the fine-scale inversion results) on the ensemble members for several reasons. First, we can reduce the dimensions of the inverted matrix in the Eq. (2.97) while calculating the Kalman gain in the second step assimilation. This could be important, particularly for large-scale problems. Second, we capture the large-scale features of the geological model contributing the flow dynamics in the underlying problem without over-constraining the ensemble members. Finally, the fine-scale inversion solution can be easily upscaled from the high resolution to the low resolution in the cross covariance estimates in Eq. (2.99). As an alternative, one could carry out the inversion in the coarse-scale directly. However, this will require downscaling of the coarse-scale estimates to the fine-scale for application of Eq. (2.99). In our approach, we favored upscaling as opposed to downscaling.

CHAPTER III

CHARACTERISTICS OF ENSEMBLE KALMAN FILTER

Prior to the application of the proposed hybrid EnKF approaches, we first need to be aware of the characteristics of EnKF performance in general. This chapter aims to depict the some of the important key parameters in EnKF which can affect the EnKF updating significantly. We demonstrate the sensitivity of the key parameters through a synthetic example and a field scale application.

3.1 Application to a Synthetic Case

To generate the reference permeability model and the permeability of the initial members in the ensemble, the sequential Gaussian simulation utility of GSLIB[6] is utilized. A total of 99 different permeability realizations were generated; each realization uses the same variogram and was conditioned to the permeability at the well position. One of the realizations was taken to be the true or the reference permeability model. The generated permeability realizations and the reference permeability fields have a non-Gaussian spatial histogram and this was chosen to highlight some of the difficulties in the application of the traditional EnKF. The reference model for this synthetic case is shown in the figure 3.1 below.

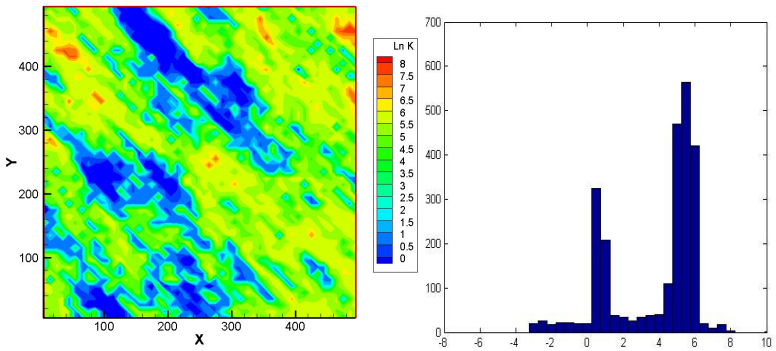


Figure 3.1 Reference model log permeability and its spatial histogram

Description of the EnKF parameters for this case is listed below.

EnKF parameters

- 3 producers and 1 injector
- The number of the ensemble is 99
- Favorable mobility Ratio (Oil viscosity =0.3 cp, Water viscosity =0.3 cp)
- Assimilation WWCT observation Data from 800 days to 3200 days (every 100days from 800 days to 2000 days and every 50 days for the rest)
- WWCT measurement error is 10%
- After assimilation, run all the ensemble from time 0 to 4000 days
- Injection Rate 300 RESV/ 3 Production rates 100,150,70 RESV
- State variables { lnK, P, Sw, WWCT}

The history matching results are shown in the figures 3.2 and 3.3 below. Both results show that updated model responses in field cumulative oil production total (FOPT) and water cut (WWCT) become closer to the reference one than the initial model responses in terms of the reduction in the spread around the true response. However, there is still a substantial (-15%~+15%) misfit in WWCT responses.

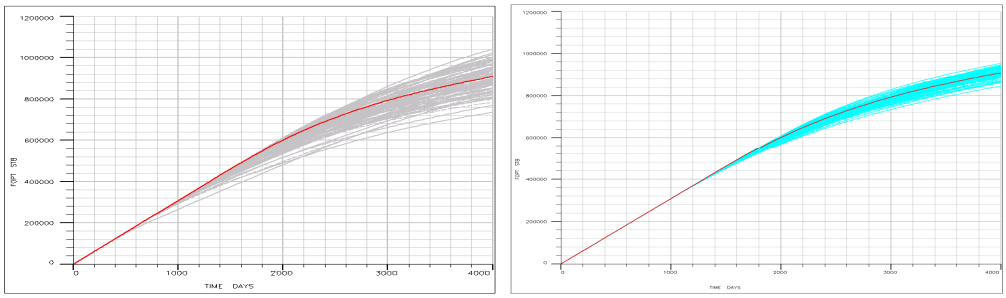


Figure 3.2 Field cumulative oil production responses from the ensemble models; the reference model is in red line; (a) is the initial ensemble models, (b) is the EnKF updated models.

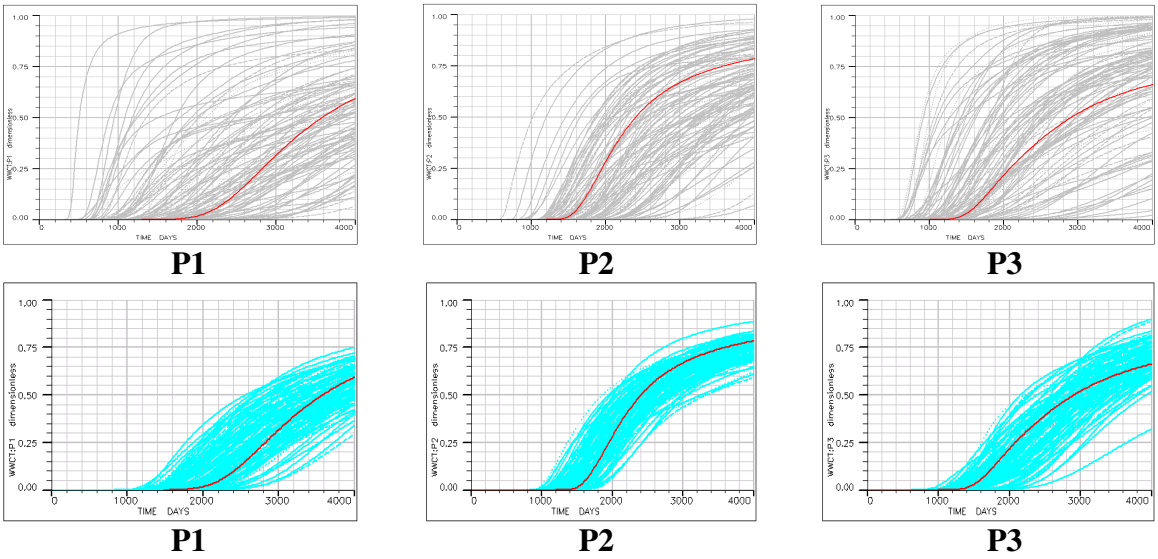
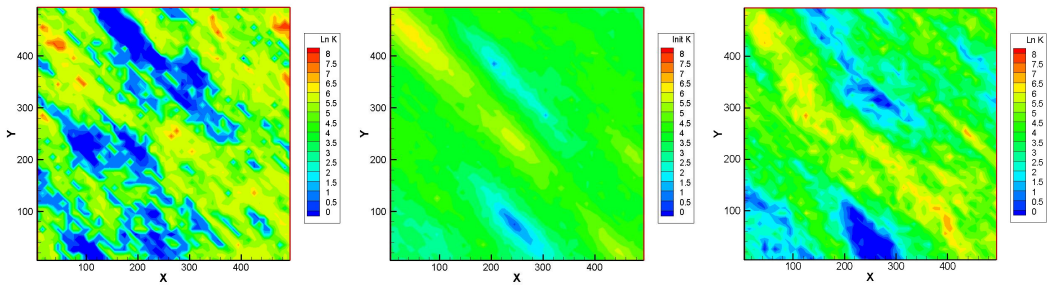


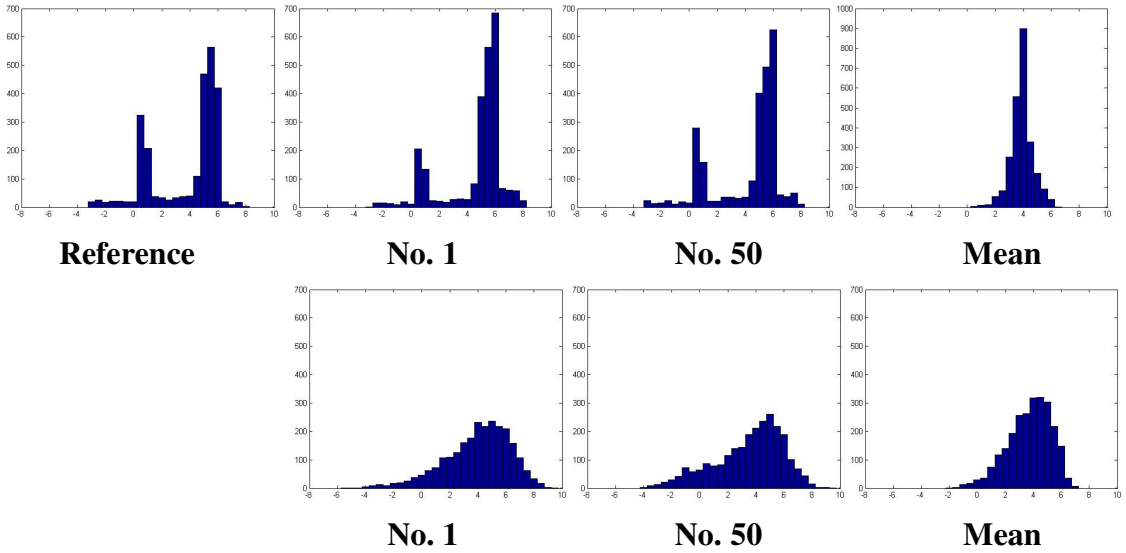
Figure 3.3 Well water cut responses from the ensemble models; the reference model is in red line; first row is the initial ensemble models, the second is the EnKF updated models.

In terms of the permeability fields, Figure 3.4 shows the mean of the ensemble permeability field and a comparison between initial mean model and the final updated mean model.



Reference **Initial** **EnKF**

Figure 3.4 Mean of the ensemble permeability fields; reference model in the left, the initial mean of the ensemble models in the center, and the mean of the final updated ensemble models.



Reference **No. 1** **No. 50** **Mean**

No. 1 **No. 50** **Mean**

Figure 3.5 Spatial histograms of the log permeability values comparison initial models and updated models; in the first row and from the left, reference model, the ensemble No. 1 model, and the ensemble No. 50, the mean of the ensemble models in the second row from the left, the ensemble No. 1 model, and the ensemble No. 50, the mean of the ensemble models.

Since the Kalman updating equation is optimal for Gaussian distribution which can be described by mean and covariance, but for non-Gaussian PDF like this case, the EnKF updating makes the spatial histogram of permeability Gaussian distribution. Based on the results that we obtained from this case, we examine two methods to assist EnKF updating. One is Streamline-based covariance localization and the other is Normal Score transformation of the parameters. Following are the description of the methods and the results applied for the same synthetic problem.

3.2 Streamline based Covariance Localization

In the literature, there are many studies to improve the estimation of the covariance calculation from the ensemble model. One of them is called the covariance localization which modifies the calculation of the sample covariance by weighting values. Especially for small ensemble sizes with noisy and possibly erroneous cross-covariances, unrealistic updates and degraded EnKF performance is often observed. (Gu and Oliver 2006; Arroyo et al. 2006; Devegowda et al. 2007)

The aim of most covariance localization schemes is to eliminate spurious terms in the cross-covariance matrix arising due to sampling errors caused by finite and small ensemble sizes and to increase the effective number of ensemble members (Hamill et al. 2001). Distance based covariance localization schemes (Houtekamer and Mitchell 2001; Hamill et al. 2001) are predicated on the assumption that the correlation between model grid cells and well data is associated with certain length scales beyond which the

correlation can be assumed to be zero. Mathematically, the localized EnKF can be expressed with the same notation as Chapter II as

$$\mathbf{y}_j^a = \mathbf{y}_j^f + \boldsymbol{\rho} \circ \left(\mathbf{C}_{\mathbf{yy}}^e\right)^f \mathbf{H}^{\mathbf{T}} (\mathbf{H}\left(\mathbf{C}_{\mathbf{yy}}^e\right)^f \mathbf{H}^{\mathbf{T}} + \mathbf{C}_{\boldsymbol{\varepsilon\varepsilon}}^e)^{-1} (\mathbf{d}_j - \mathbf{H}\mathbf{y}_j^f) \tag{3.1}$$

where the localizing function $\boldsymbol{\rho}$ operates on the Kalman gain matrix. The operator '$\circ$' is an element-by-element multiplication also called a Schur product. Various valid correlation functions are discussed further in Gaspari and Cohn (1996). However the assumption that correlation is based on the distances between two points is not always valid for the reservoir environment where the build-in heterogeneity dominates the flow dynamics. Arroyo et al. (2006) proposed the streamline base localization method in the reservoir problem application. The basic idea is that the streamline trajectory from the reservoir simulation is utilized to identify the influential zones at the particular time and under the well configuration settings. All ensemble model flow geometries are all analyzed and stacked up to identify the common influential area. Then the spurious values in the covariance calculations can be removed by assigning 0's and 1's for the Schur product to eliminate the place where there is no streamline passing through. We denote this localization method as streamline trajectory localization in this thesis. Further, Devegowda et al. (2007) extend the streamline trajectory localization by weighting the relative impact of the common influential area by the magnitude of the sensitivity values derived from the streamline simulation. Now the calculation of $\boldsymbol{\rho}$ in Eq. (3.1) becomes

$$\rho_{i,j} = \left| \frac{S_{i,j}}{\max_j \left(S_{i,j} \right)} \right| \tag{3.2}$$

where S refers to the sensitivity values formed by summing the sensitivities over all the realizations and subscript i and j denotes the number of the grid block and the number of observation data respectively. Essentially, Eq. (3.2) implies that relative impact of the influential zone is calculated as the sensitivity values normalized by the maximum sensitivity of a grid block inside the influential zone with respect to the corresponding observation data. The way to calculate the sensitivity is out of scope of this thesis and is referred to the several works in the literature (Yoon et al. 1999; He et al. 2002; Cheng et al. 2006; Oyerinde 2008). We denote this localization method as streamline sensitivity localization in this thesis. Examples of the streamline trajectory localization and streamline sensitivity localization are shown in Figures 3.6 and 3.7 below.

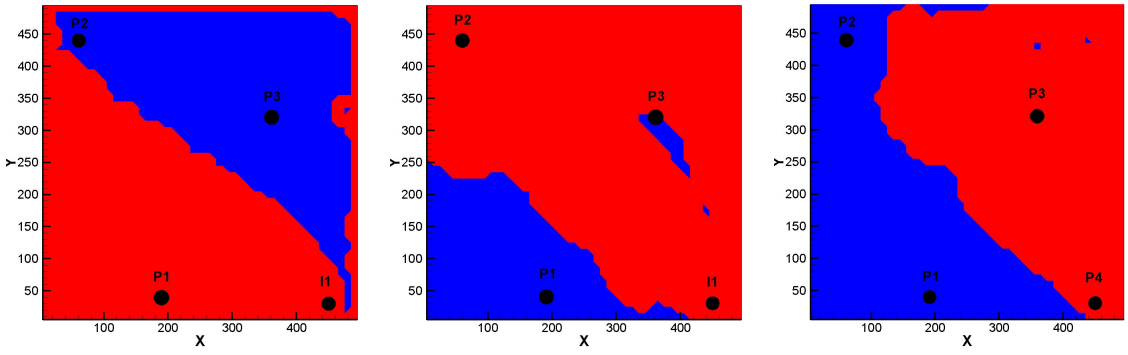


Figure 3.6 An example for the streamline trajectory localization; the localizing function is plotted for each producer, from left P1, P2 and P3. The red color is 1's and blue color is 0's.

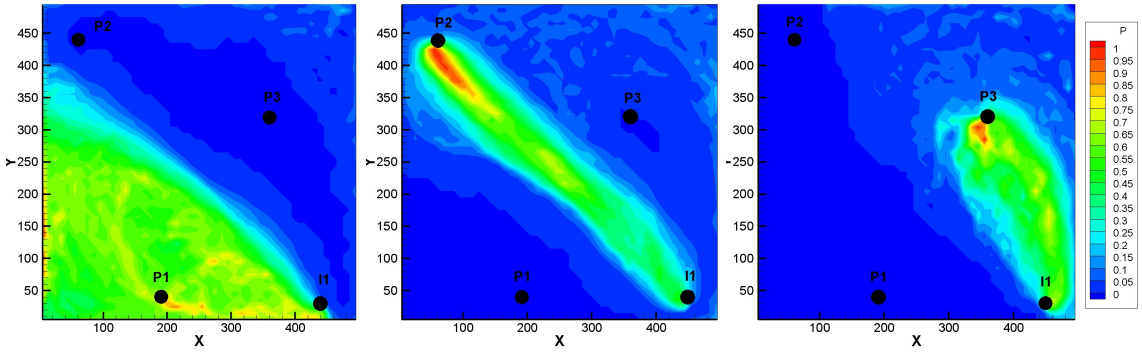
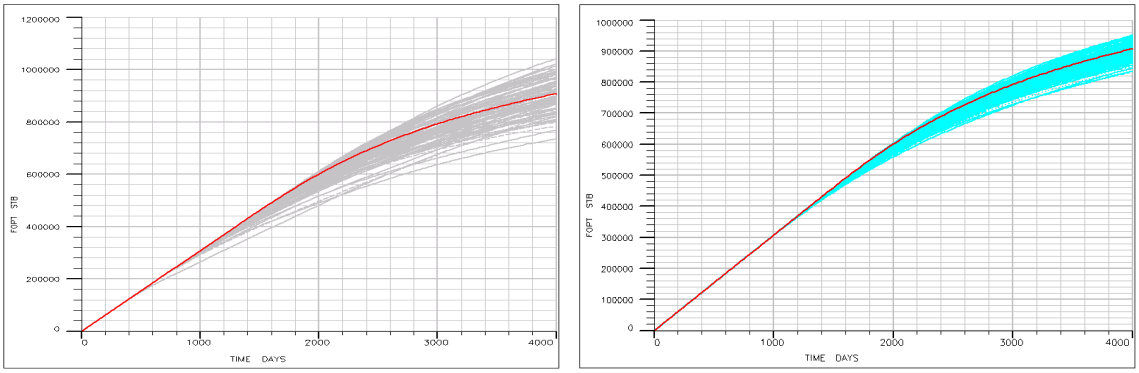


Figure 3.7 An example for the streamline sensitivity localization; the localizing function is plotted for each producer, from left P1, P2 and P3.

As you can see in the Figure 3.6, the influential zone for each producer is identified in red color and the regions in color blue are not correlated to the WWCT observation data. Comparing between the figures 3.6 and 3.7, the streamline trajectory localization captures the common influential zone and streamline sensitivity localization identify the relative influence inside the common influential zone clearly.

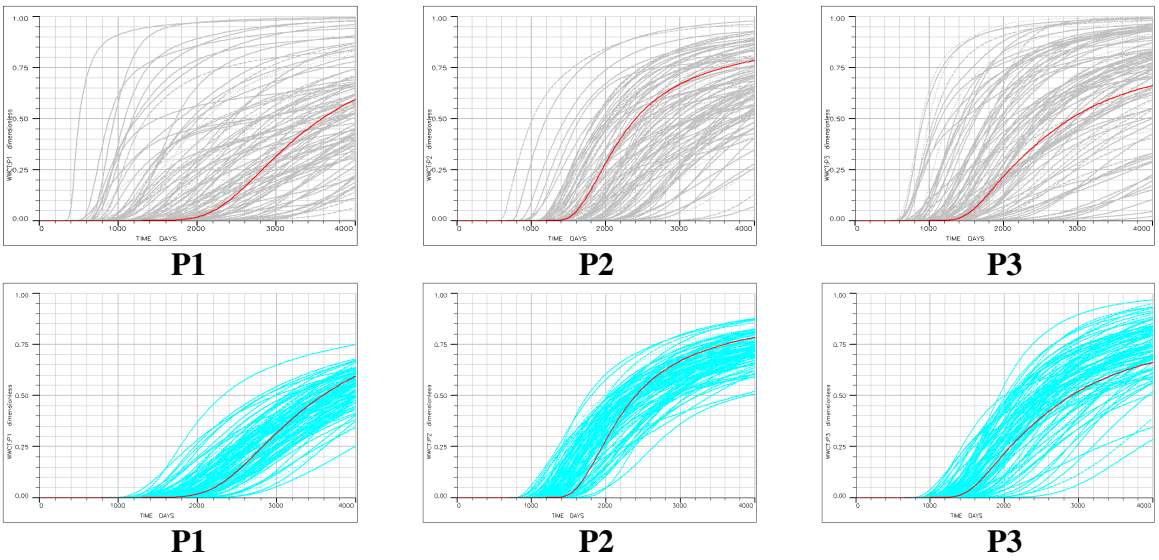
An experiment was conducted for the same synthetic problem by using the streamline trajectory localization method. The history matching results are shown in the figures 3.8 and 3.9. The quality of matching FOPT is almost same as the result in figure 3.2. As for the WWCT matching, the producer No. 1 and No. 2 (P1 and P2) results are similar to the results in figure 3.3, but the producer No. 3 (P3) matching is degraded compared to the result in figure 3.3. P3 is located in the right corner of the domain. If one looks at the updated permeability filed in figure 3.10, the EnKF with localization results show that the high permeability zone is created in the diagonal region, same as in

the EnKF without localization result. But the updated values are more overshooting (around lnK=8) compared to the EnKF without localization, and the low permeability



(a) ?(b)

Figure 3.8 Field cumulative oil production responses from the ensemble models; the reference model is in red line; (a) is the initial ensemble models, (b) is the EnKF updated models with streamline trajectory localization

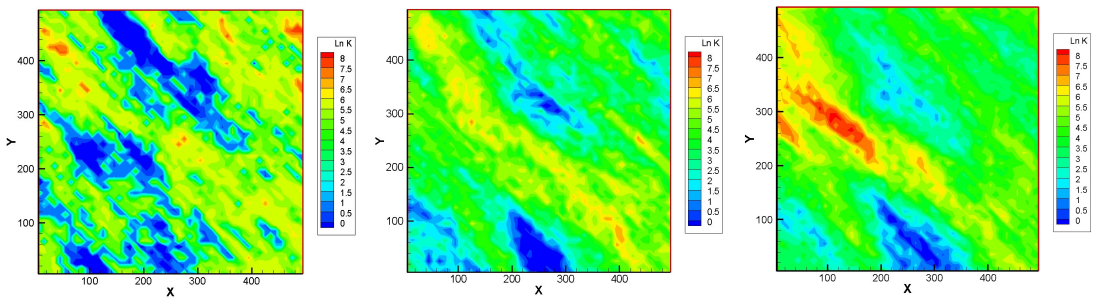


P1 P2 P3

P1 P2 P3

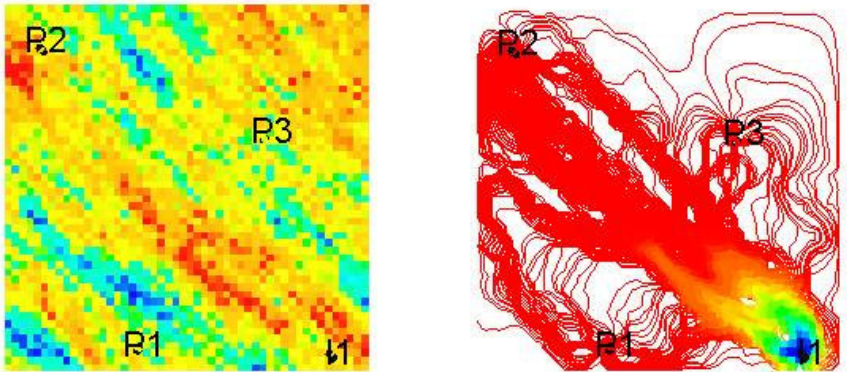
Figure 3.9 Well water cut responses from the ensemble models; the reference model is in red line; first row is the initial ensemble models, the second is the EnKF updated models with streamline trajectory localization.

barrier in the top center is not captured as well as the EnKF without localization result. This results in the communication between P3 and P2 which causes the worse WWCT matching in P3.



Reference **EnKF** **EnKF-ST**

Figure 3.10 Mean of the ensemble permeability fields; reference model in the left, the mean of the EnKF final updated ensemble models in the center, and the mean of the EnKF updated ensemble models with streamline trajectory localization.



(a) **(b)**

Figure 3.11 (a)Well configuration of the synthetic problem from an initial model and (b)streamline trajectory of the model.

A reason for the high values of the updated model is that streamline localization magnifies the correlation between the parameters and the WWCT in the diagonal

influential zone and results in increasing the values over the sequence of the updating. Figure 3.11 show the well configuration and the streamline trajectory from an initial model. The streamline trajectory shows bundles of streamline are clustered in the diagonal region exclusively, and it indicates the change that we make through EnKF is only for the region. As is shown in the Figure 3.12, the spatial log permeability distribution has bi-modal distribution in EnKF with localization compared to the EnKF without localization. It is because the localization restricts EnKF updating to the influential zones identified the streamline trajectory through the ensemble model and preserves the prior model more compared to the EnKF without localization. However, it still makes the spatial distribution more Gaussian compared to the initial model distribution (see Figure 3.5).

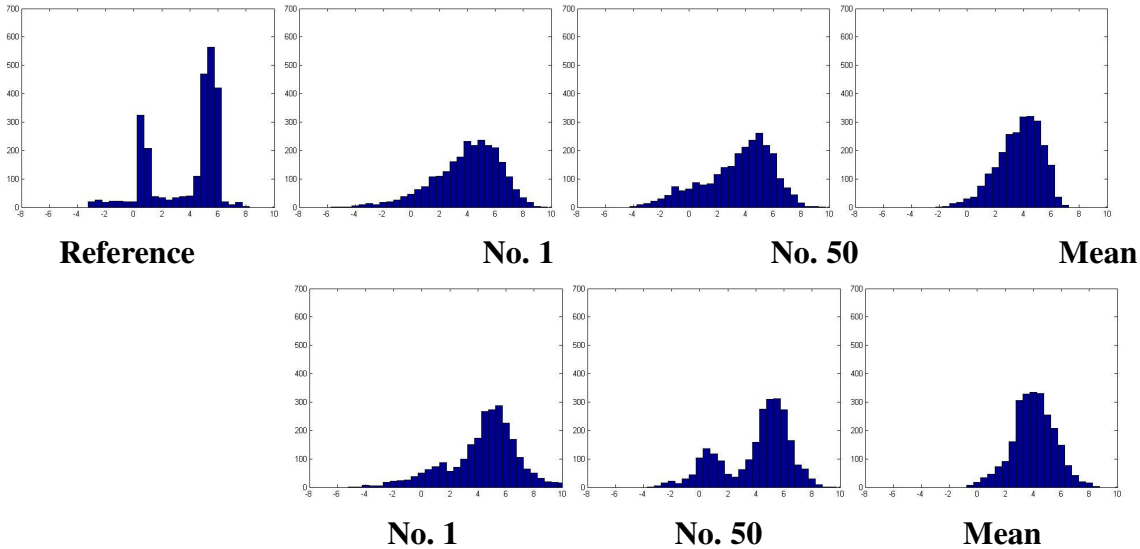


Figure 3.12 Spatial histograms of the log permeability values comparison EnKF updated models and EnKF updated models with streamline trajectory localization; in the first row and from the left, reference model, the ensemble No. 1 model, and the ensemble No. 50, the mean of the ensemble models in the second row from the left, the ensemble No. 1 model, and the ensemble No. 50, the mean of the ensemble models.

3.3 Normal Score Transformation of the Parameter

One of the main assumptions in EnKF is that Gaussian error statistics in the derived equations. This implies that the prior Gaussian distribution is suitable in the updating both the state and parameter estimations. However this requirement is not necessarily acceptable for the most of the reservoir characterization problems, because the petro physical properties such as porosity and permeability are often characterized in a multi-facies and multimodal distribution in a real field environment. Gu. and Oliver (2004) shows the benefit to use the normal score transformation on the state variables (Saturations), which can be bi-modal distribution near the water front, to prevent the non-physical updated values from EnKF.

In this study we apply the normal score transformation on the parameter variable (Permeability) which has a bi modal distribution and update the normal scored values in EnKF and back transformed to the physical values after the assimilation step. The crucial point for this approach is that we construct the transformation table for the permeability values based on the prior model values. In this way, we can preserve the prior parameter distribution and geological realism. An example of the normal score transformation is demonstrated in Figure 3.13.

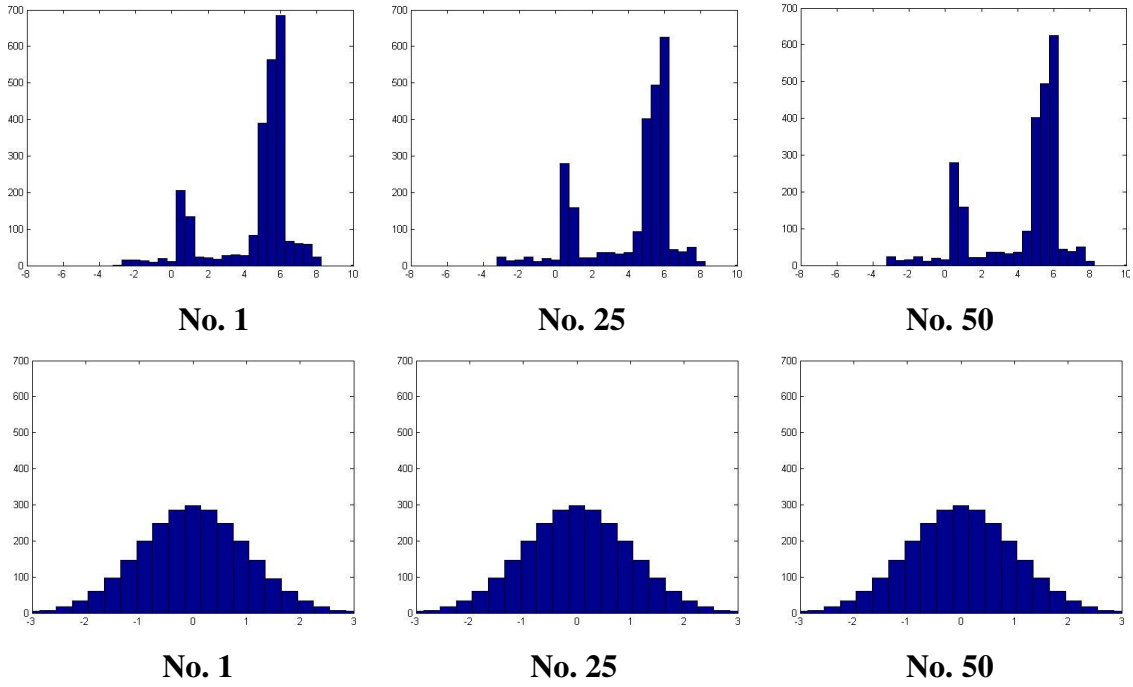
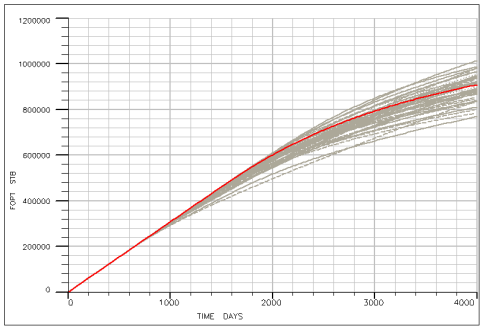


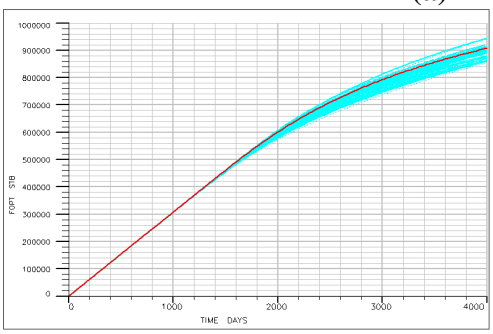
Figure 3.13 Spatial histograms of the log permeability values comparison initial ensemble models and normal scored ensemble models; In the first row and from the left, initial permeability spatial histogram of ensemble No. 1,No. 25 and No. 50 models, in the second row and from left, the normal scored permeability spatial histogram of ensemble No. 1, No. 25, and No. 50 models.

As the Figure 3.13 shows, after the normal score transformation, the spatial permeability distribution is all standard normal distribution. So the prior parameter distribution can be described by the mean and covariance which is suitable for EnKF updating.

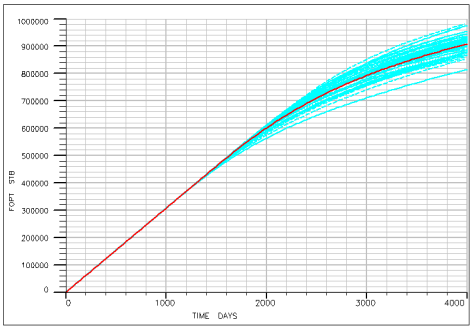
So we apply the normal score transformation for the permeability values for the same synthetic case as in sec.3.1. We change the ensemble size to be 50 and compare the result with the EnKF without normal score transformation (Plain EnKF) results below. Figures 3.14 and 3.15 show the comparison between Plain EnKF and the EnKF with normal score transformation of the permeability (Normal EnKF) in FOPT and WWCT.



(a) Initial Model



(b)EnKF



(c)Normal EnKF

Figure 3.14 Field cumulative oil production responses from the ensemble models; the reference model is in red line; (a) is the initial ensemble models , (b) is the EnKF updated models, (c) is EnKF updated models with the normal score transformation of the permeability.

Both results show the better matching for the Plain EnKF than the Normal EnKF. This indicates that the normal score transformation does not produce the better matching quality, however if you look at the mean of the updated ensemble permeability field in Figure 3.16 and ensemble model spatial permeability histograms in Figure 3.17, normal score transformation prevents the overshoot and undershoot problems in the permeability values and preserve the prior permeability bi-modal distribution after the sequence of the updating as we expected.

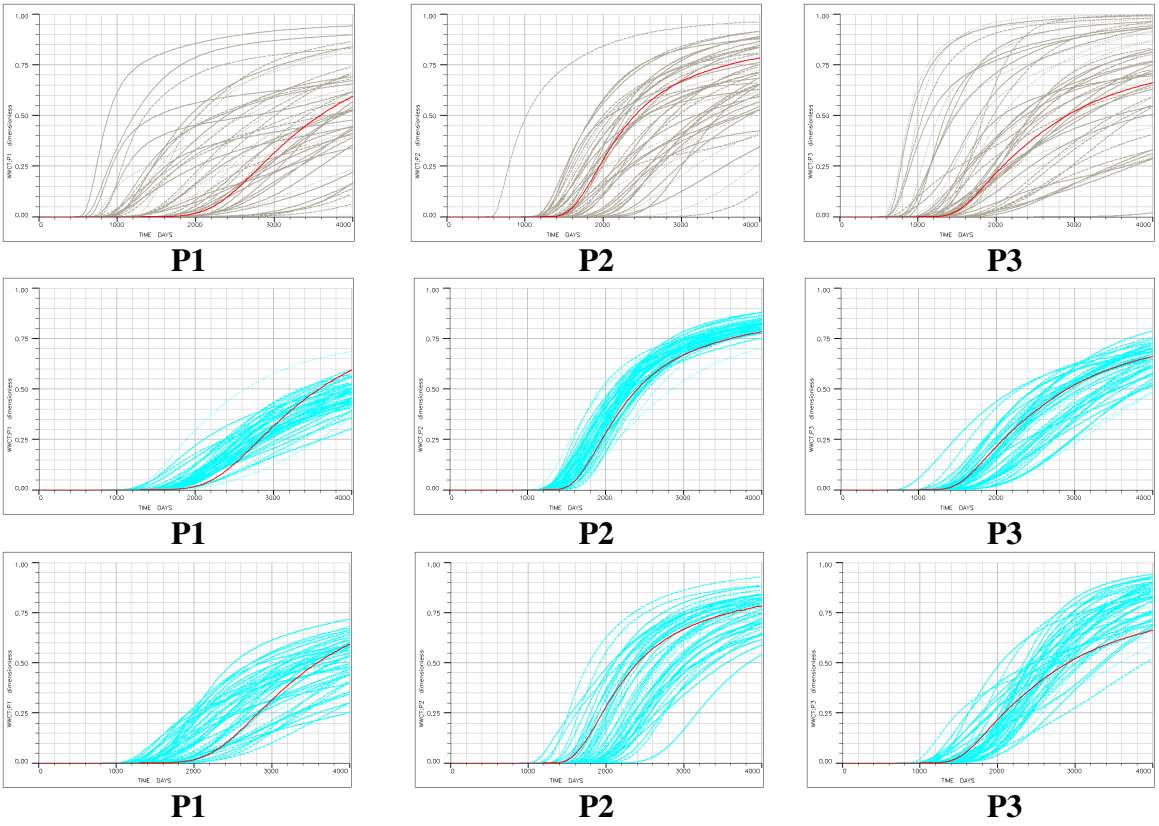


Figure 3.15 Well water cut responses from the ensemble models; the reference model is in red line; first row is the initial ensemble models, the second row is the EnKF updated models, and the third row is the EnKF updated models with normal score transformation.

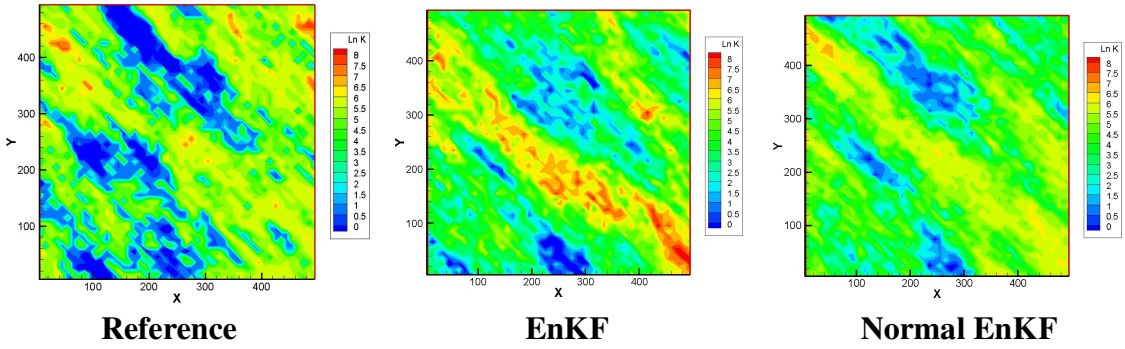


Figure 3.16 Mean of the ensemble permeability fields; reference model in the left, the mean of the EnKF final updated ensemble models in the center, and the mean of the EnKF final updated ensemble models with normal score transformation.

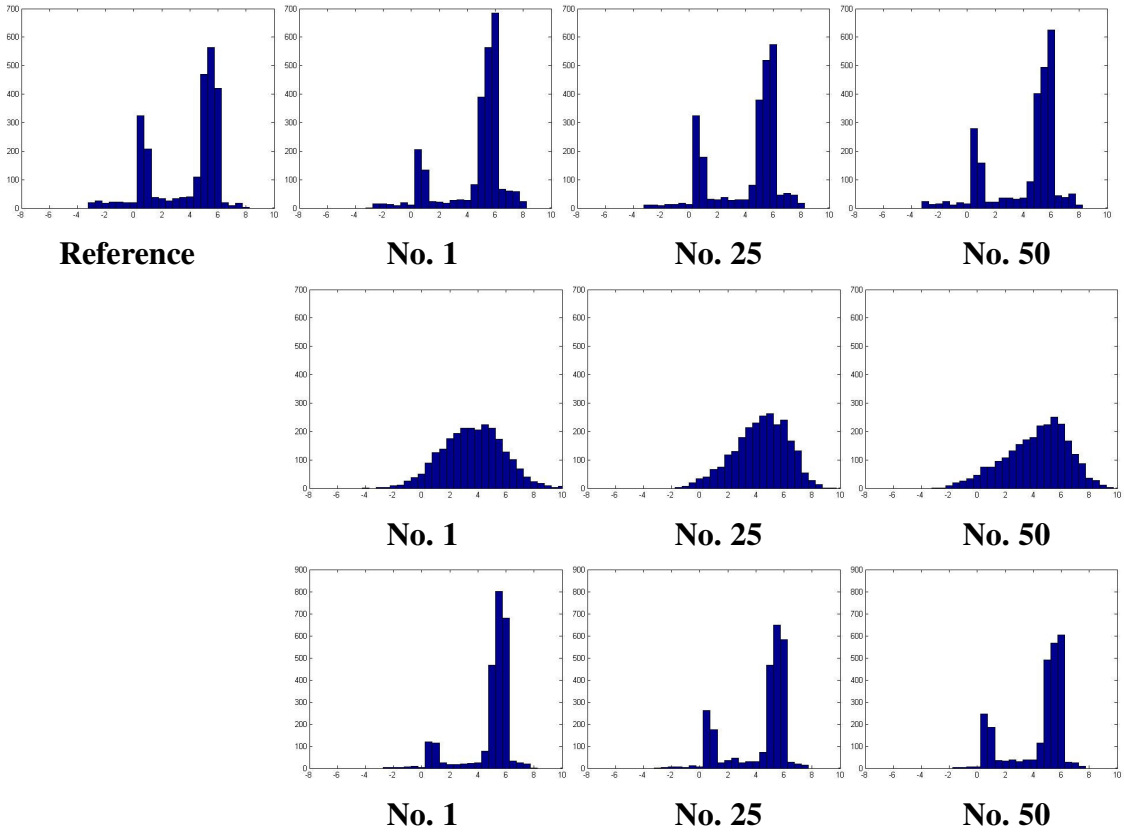


Figure 3.17 Spatial histograms of the log permeability values comparison; in the first row and from the left, initial permeability spatial histograms of ensemble No. 1,No. 25 and No. 50 models, in the second row and from left, EnKF updated permeability spatial histograms of ensemble No. 1, No. 25, and No. 50 models and in the third row and from left, Normal EnKF updated permeability spatial histogram of ensemble No. 1, No. 25,and No. 50 models.

3.4 Goldsmith Field Application

In this section, we demonstrate the applicability of the EnKF for a field example, the Goldsmith CO_2 pilot project study (He at al. 2002; Cheng et al. 2005). We used an ensemble size of 50 realizations of porosity and permeability fields conditioned to well data and secondary seismic attributes. The pilot area comprises 9 inverted 5-spot patterns covering around 320 acres and the average thickness of the formation is 100 ft.

Significant water-cut data for 20 years of production prior to the start of the CO_2 flood is available at 9 production wells and these were used to condition the permeability fields. By the location of these wells in the center of the field, it is expected that most of the changes to the permeability distribution should be concentrated in this region. Figure 3.18 is an areal plot of the location of the Goldsmith CO_2 pilot study within the extended study area showing the location of the producers (in yellow) and the injectors (in blue). The simulation model is shown in Figure 3.19 below which has dimensions of 58 by 53 by 10 in corner point grid system.

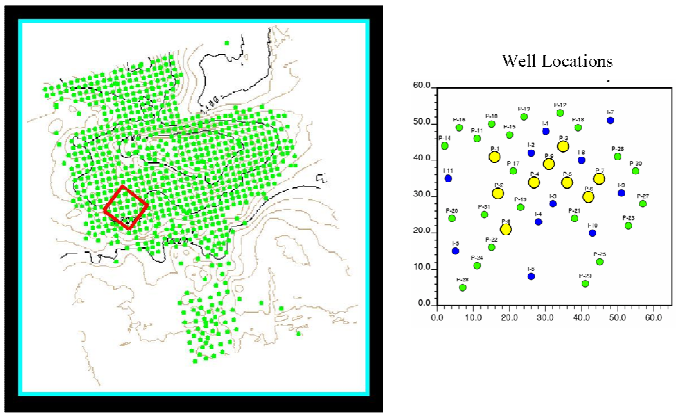


Figure 3.18 Gold Smith well configuration map

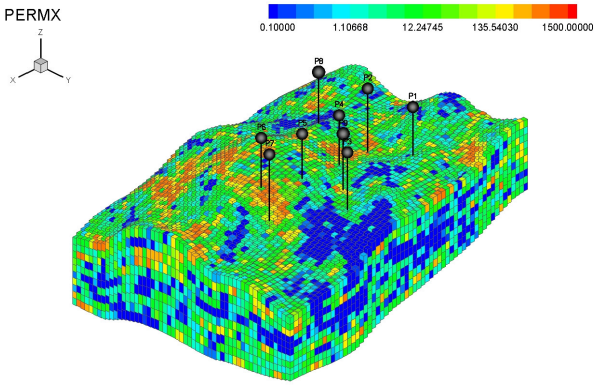


Figure 3.19 Gold Smith simulation model

50 initial model WWCT responses are shown in the Figure 3.20.

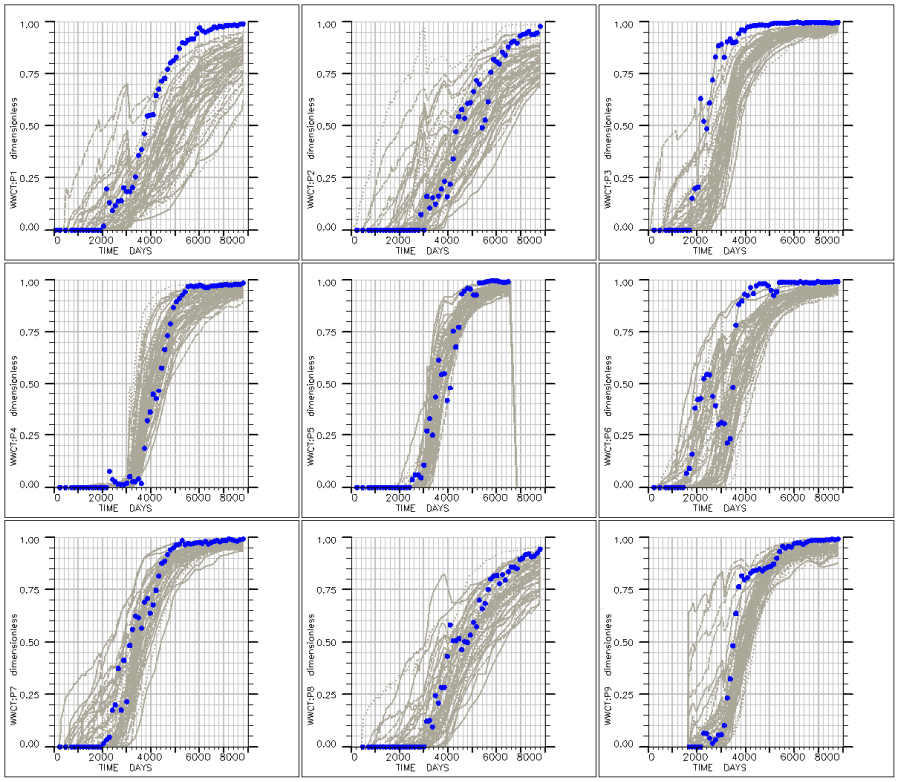


Figure 3.20 Initial 50 ensemble model WWCT responses; the observation data in blue points and the initial ensemble model responses are in grey lines, in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5,and P6, and in the third row, from left, P7, P8 and P9.

As you can see the initial model WWCT spread is very large (at largest about 50% difference). We need to reduce the uncertainty of the ensemble model and build better performance models with reliable forecasts through history matching. Here we found the key parameters to adjust for EnKF to get the better results and the characteristics of EnKF for the filed application. Initial permeability fields are shown in Figure 3.21. Default parameters that we use for this study are listed below.

EnKF Parameters

- 9 producers for WWCT history mach
- 58×53×10 grids
- The number of the ensemble is 50
- Assimilation WWCT observation Data from 0 days to 3840 days and prediction runs to 7800 days
- WWCT measurement error is 10%
- After assimilation, run all the ensemble from time 0 to 7800 days
- State variables { lnK, P, Sw, WWCT}

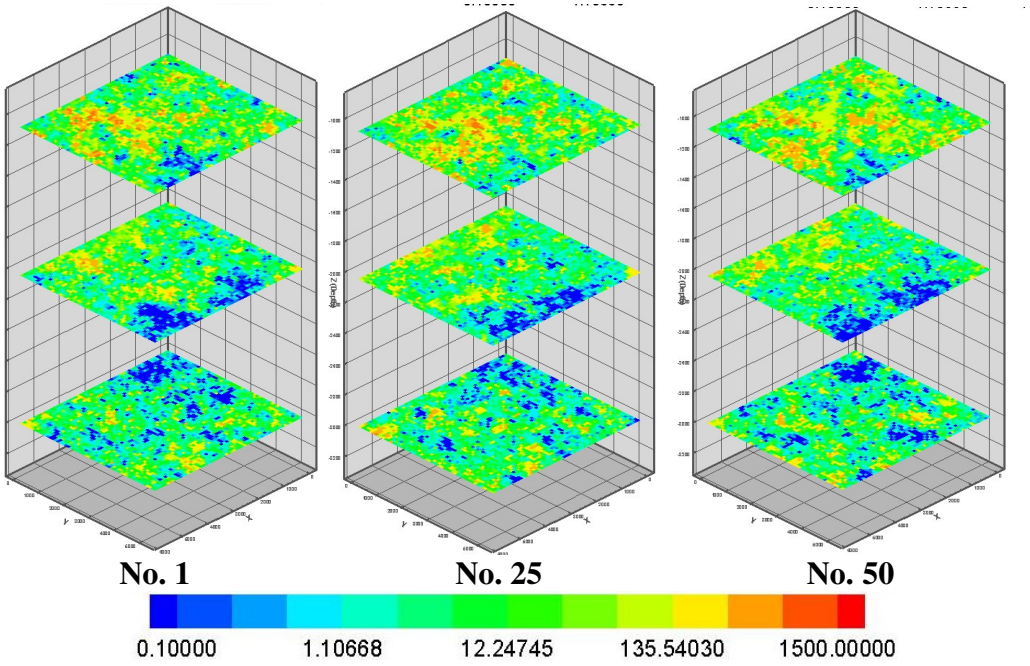


Figure 3.21 Initial ensemble permeability fields; from the left, ensemble No. 1, No. 25 and No. 50 at the depth of 1080 ft, 2042 ft and 3004 ft from the top.

3.4.1 Sensitivity with Updating Limit Constraint

From this field application study, we found that it is very important to limit the updated variables through EnKF. We first let the limit open lnK[-12 12] and conduct Plain EnKF. Figure 3.22 is the result of WWCT matching from the Plain EnKF. As you

can see that the spread of the WWCT responses are reduced from the initial model results in Figure 3.20, but the matching the observation history is worse in the updated model responses than the initial model ones.

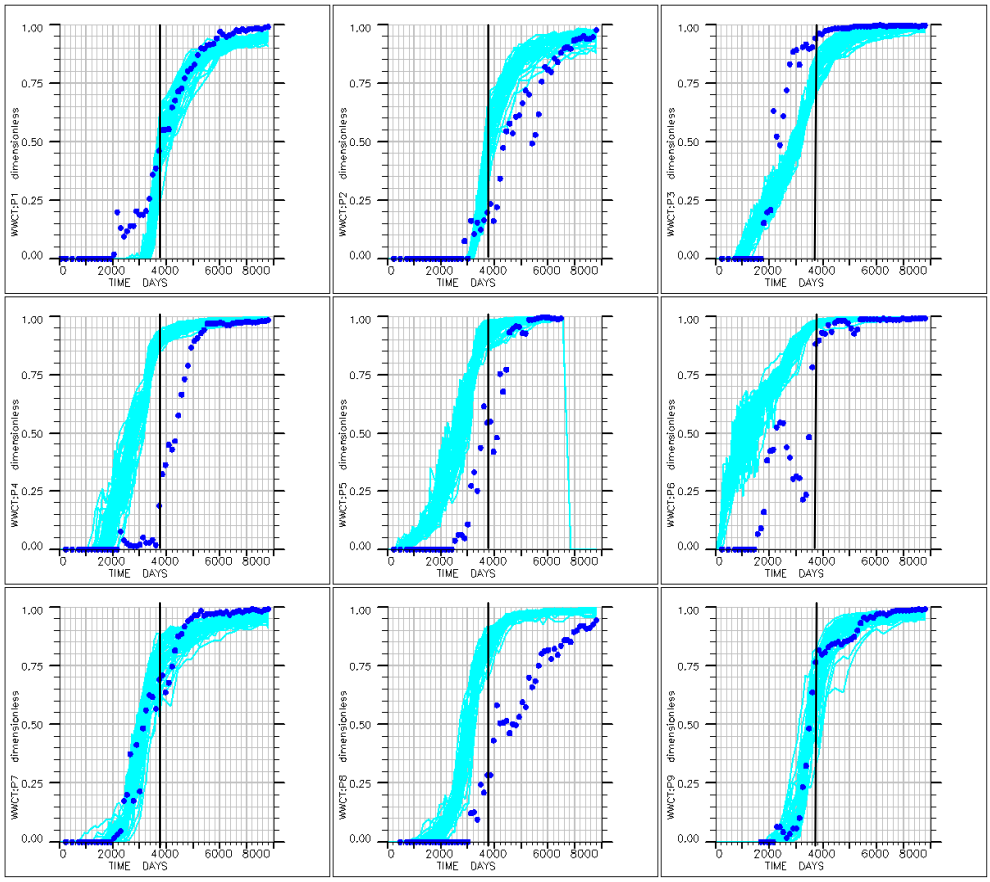
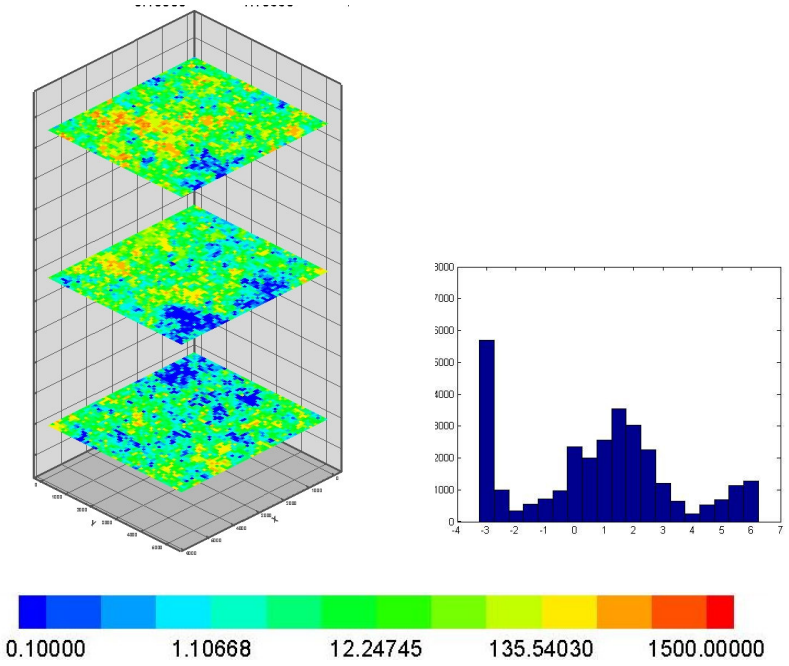
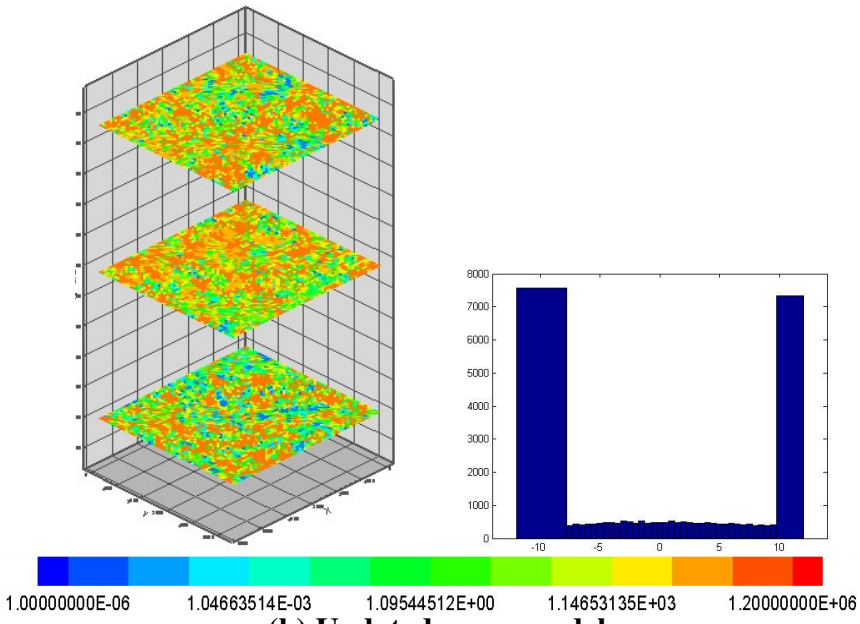


Figure 3.22 EnKF final 50 updated ensemble model WWCT responses without limit value constraints; the observation data in blue points and the ensemble model responses are in light blue lines; in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5,and P6, and in the third row, from left, P7, P8 and P9.

From the updated permeability results in Figure 3.23, EnKF updated model has severe over and under shooting values reached to the maximum and minimum limit.



(a) Initial mean model



(b) Updated mean model

Figure 3.23 The ensemble No. 1 permeability field and spatial distribution comparison between (a) the mean of the initial ensemble models and the mean of the updated ensemble models.

This is totally destroying the geological realism and consequently the matching WWCT is also not successful. Based on the prior knowledge of the permeability distribution, we can define the minimum and maximum allowable limit values. This is also one of the major uncertainties on the reservoir characterization. For this study following, I define the maximum and minimum permeability value range as lnK[-4 7] from the prior initial model distributions. We conduct the EnKF again with the limit of the updated permeability values.

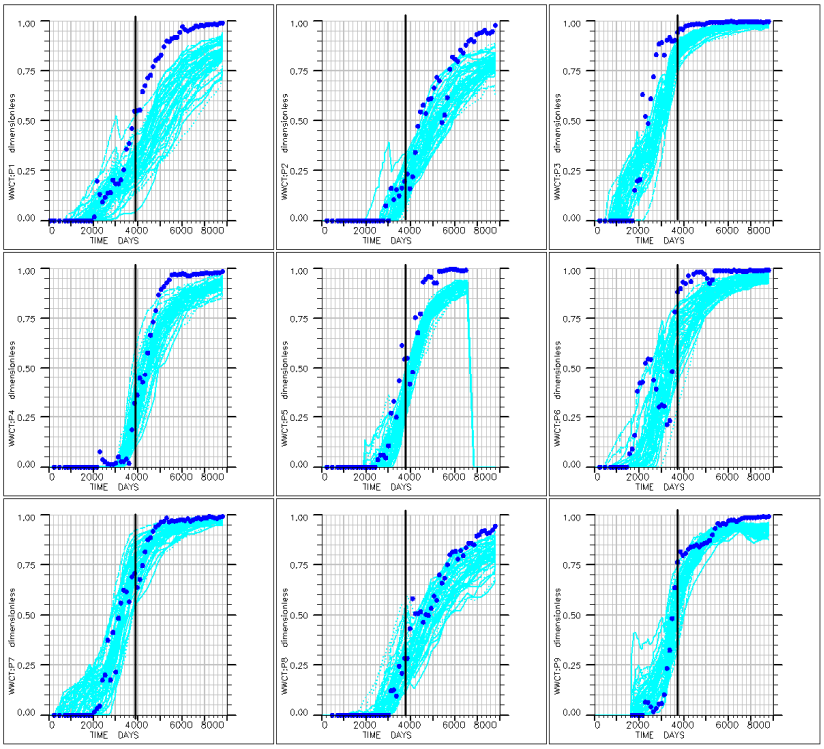


Figure 3.24 EnKF final 50 updated ensemble model WWCT responses with limit value constraints; the observation data in blue points and the ensemble model responses are in light blue lines; in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5,and P6, and in the third row, from left, P7, P8 and P9. Vertical line shows the last assimilation time step.

The figure 3.24 shows the WWCT matching results and it shows significant improvement from the results in figure 3.20 in terms of the matching to the observation data trend. But we still see discrepancy between the simulation responses and the observation data especially for the water break through time in P1, P3 and P7, P8 and prediction part in the P1, P2 and P5 and non-monotonic behavior in WWCT P6. Updated ensemble permeability fields are shown in Figure 3.25. Compared to the initial ensemble model permeability fields in Figure 3.21, we can see the overshoot and undershoot problems in the updated models. the high permeability and low permeability region, respectively. As we can expect, the spatial model distribution becomes more Gaussian after the sequence of the updating shown in Figure 3.26.

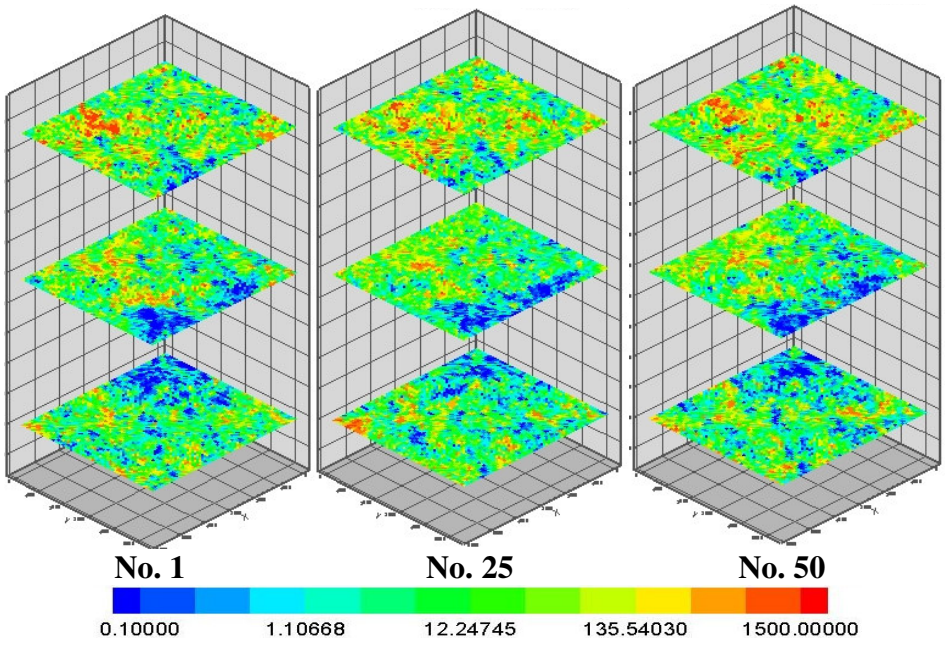


Figure 3.25 Updated ensemble permeability fields; from the left ensemble No. 1, No. 25 and No. 50 at the depth of 1080 ft, 2042 ft and 3004 ft from the top.

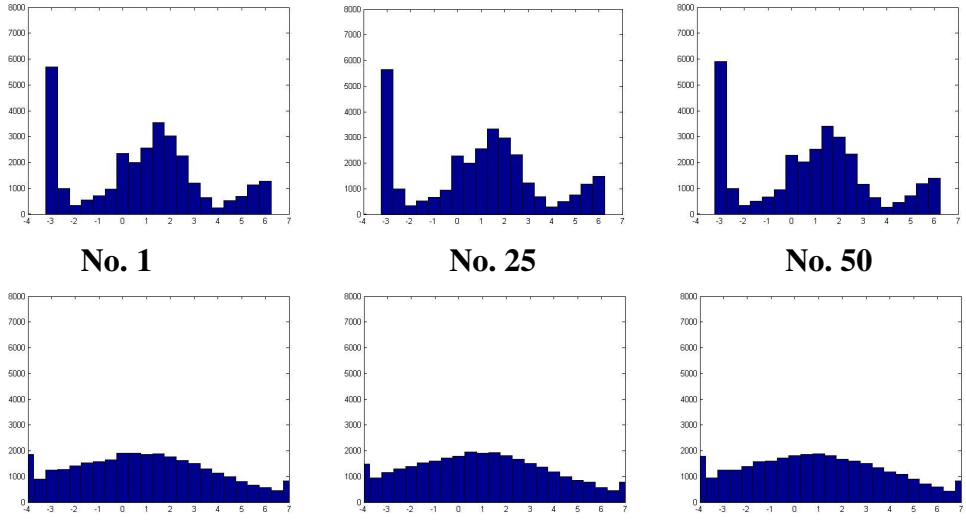


Figure 3.26 Spatial histograms of the log permeability values comparison between initial ensemble models and EnKF updated models; in the first row and from the left, the initial ensemble No. 1, No. 25 and No. 50 models, in the second row from the left, the updated ensemble No. 1, No. 25 and No. 50 models.

3.4.2 Sensitivity with the Measurement Error Variance

Another thing that we want to test for this study is that how the measurement error variance affects the EnKF updating because we do not know the true error variance in a real field case. So We try changing the measurement error values from 0.1 to 0.05 and rerunning the EnKF with limit of the updated permeability values. By reducing the observation error, the spread of the WWCT responses from the updated model is reduced very much shown in Figure 3.27, but the quality of the matching is degraded compare to the previous results. This indicates that the ensemble model collapsed to each other due to the loss of the variance of the data and if you look at the updated permeability profile in Figure 3.28, and its distribution, the updated model has severe

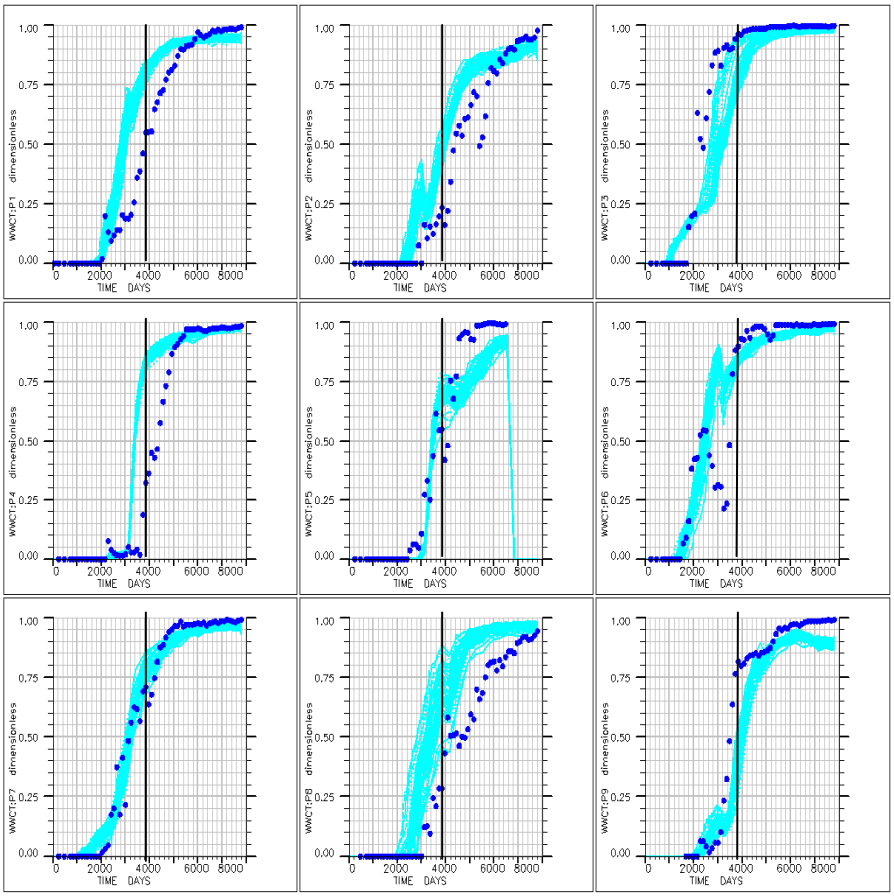


Figure 3.27 EnKF final 50 updated ensemble model WWCT responses with limit value constraints for measurement error variance =5 %; the observation data in blue points and the ensemble model responses are in light blue lines, in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5, and P6, and in the third row, from left, P7, P8 and P9. Vertical line shows the last assimilation time step.

over and undershooting values and totally destroyed the prior model information.

Based on these results, it is necessary to take account for the relative impact of the observation data to the prior information to preserve the built in geological realism and overestimate the observation data.

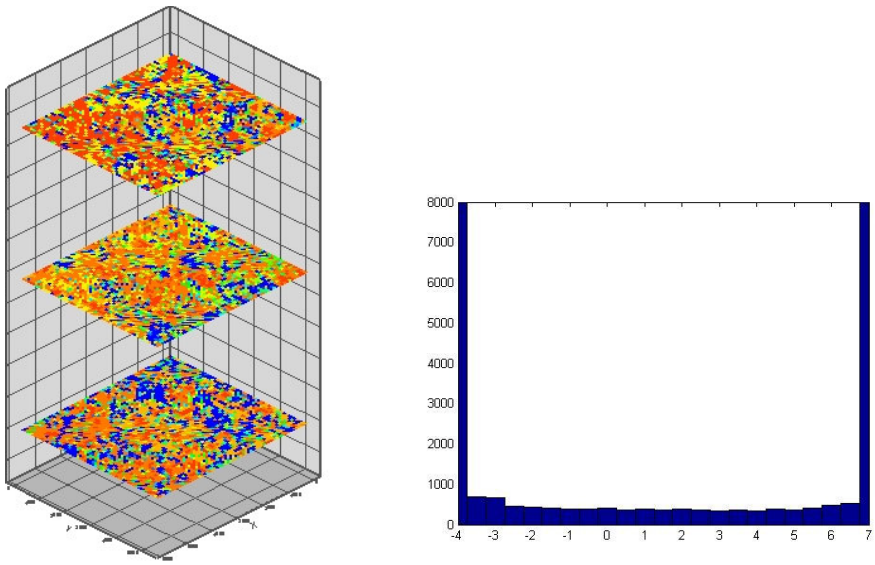
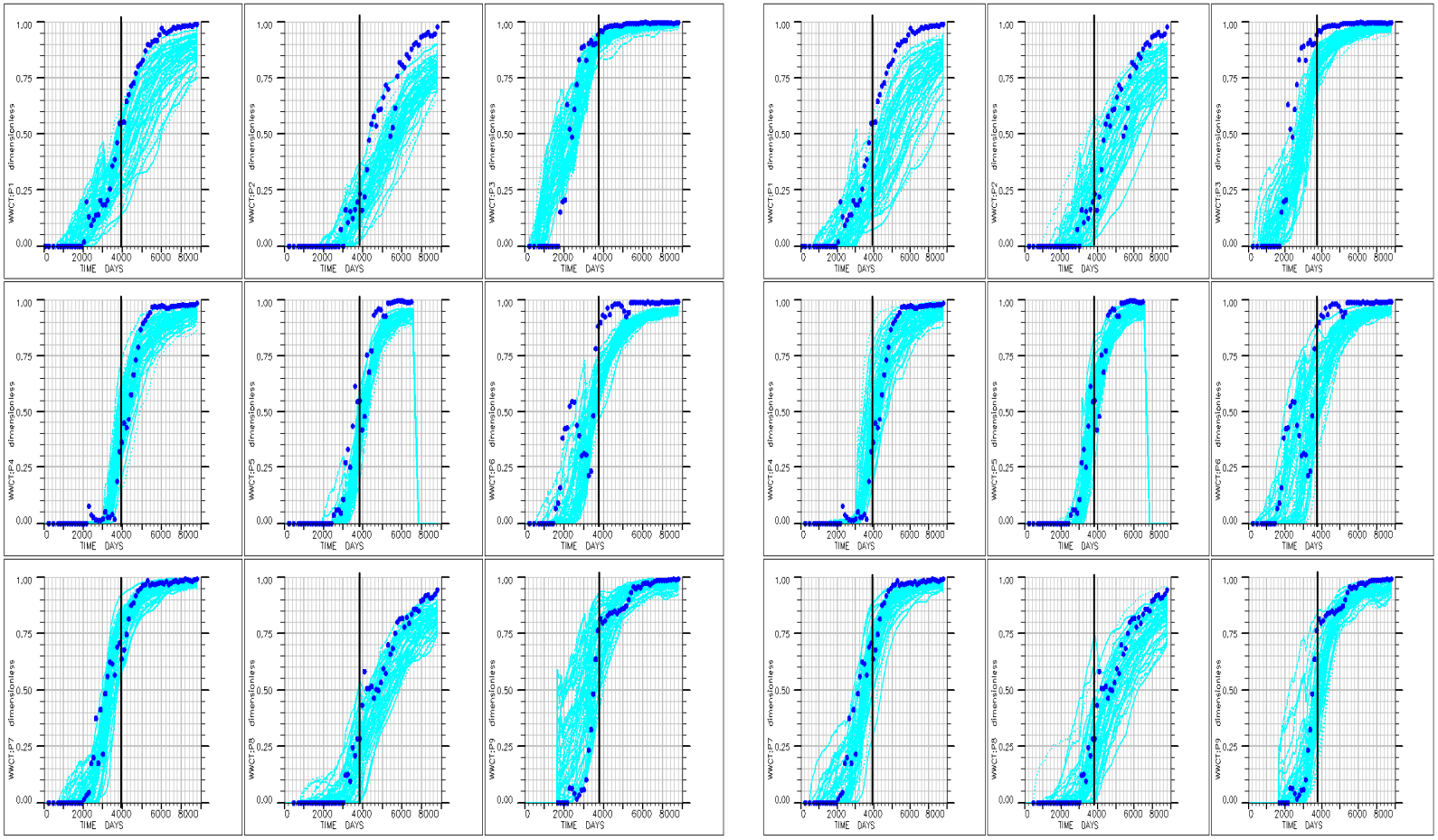


Figure 3.28 The ensemble No. 1 permeability field and spatial distribution

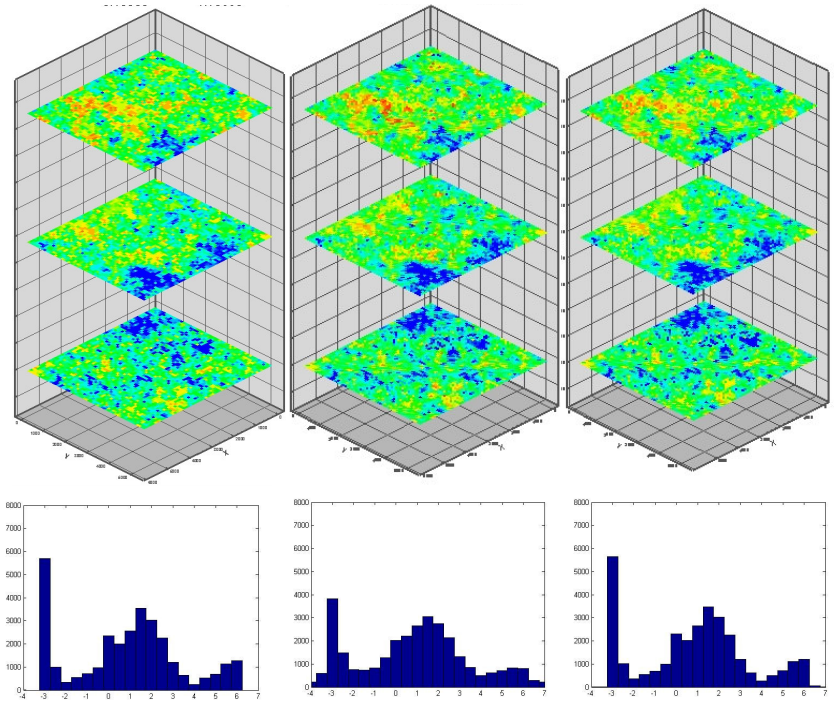
3.4.3 Comparison with Localization Methods

We want to investigate the effect of the covariance localization on EnKF for this case and compare the results from streamline trajectory localization and from sensitivity localization. We denote the streamline trajectory localization results as EnKF-ST and Sensitivity Localization ones as EnKF-SS. Figure 3.29 shows both EnKF-ST and EnKF-SS preserve the prior permeability distribution better than without localization. Precisely, EnKF-SS results prevent the overshooting and undershooting better than the EnKF-ST (See the Figure 3.30). And Streamline sensitivity localization result shows the change from the prior model is minimal and matching WWCT is improved from the initial models.



(a) EnKF-ST **(b) EnKF-SS**

Figure 3.29 EnKF final 50 updated ensemble model WWCT responses comparisons between (a) ENKF with streamline trajectory localization and (b) EnKF with streamline sensitivity localization; the observation data in blue points and the ensemble model responses are in light blue lines, in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5,and P6, and in the third row, from left, P7, P8 and P9. Vertical line shows the last assimilation time step.



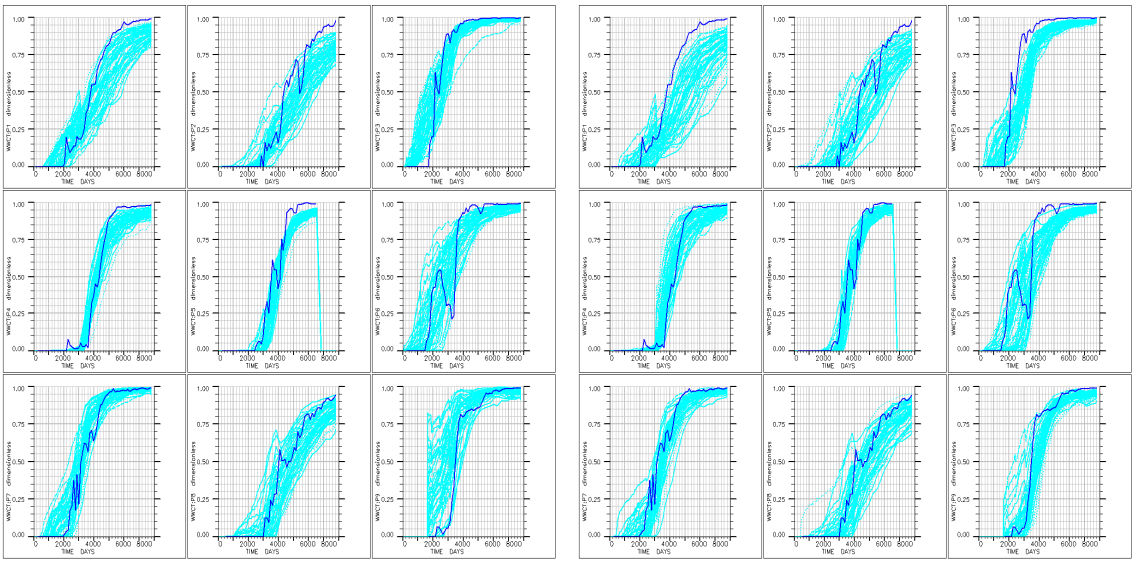
Initial **EnKF-ST** **EnKF-SS**

Figure 3.30 The ensemble No. 1 permeability field and spatial histogram comparison between the initial ensemble model, the updated ensemble model from EnKF-ST and the updated ensemble model from EnKF-SS.

3.4.4 Normal Score Transformation with Localization Methods

On top of the localization, we can use normal score transformation of the permeability values to see if the combination of two provides the better estimation. We denote the EnKF with streamline trajectory localization and normal score transformation as NST-EnKF, and the EnKF with streamline sensitivity localization and normal score transformation as NSS-EnKF in the following result figures. Figure 3.31 shows the comparison of history matching results between NST-EnKF and NSS-EnKF. Over all the matching quality between two are similar except for P3 and P9. As Figures 3.32 and

3.33 show, normal score transformation can assist preserving the prior permeability distribution better and with localization, the change from the initial model is localized and minimal. Precisely with streamline sensitivity localization, the change from the initial model is more localized near the well locations and smaller than one with streamline trajectory localization.



(a) NST-EnKF **(b) NSS-EnKF**

Figure 3.31 EnKF final 50 updated ensemble model WWCT responses comparisons between (a) normal scored ENKF with streamline trajectory localization and (b) normal scored EnKF with streamline sensitivity localization; the observation data in blue points and the ensemble model responses are in light blue lines, in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5,and P6, and in the third row, from left, P7, P8 and P9. Vertical line show the last assimilation time step.

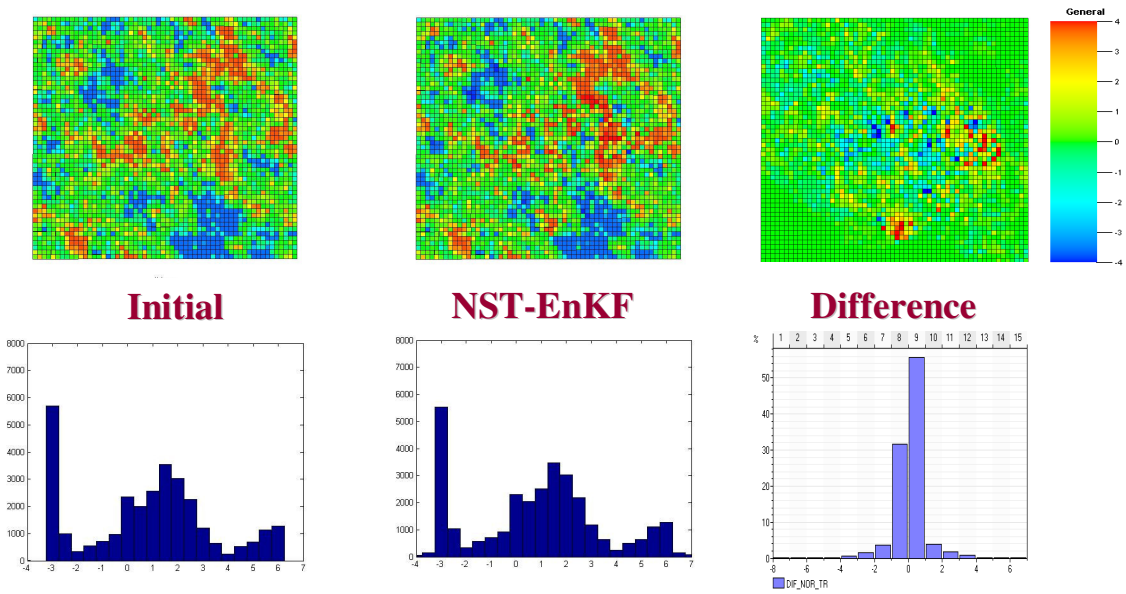


Figure 3.32 The ensemble No. 1 permeability field and spatial histogram and changes from the initial model No. 1; from the left the initial ensemble model No1, the updated ensemble model No. 1 from NST-EnKF and the change that we make from NST-EnKF updating from the initial model No. 1.

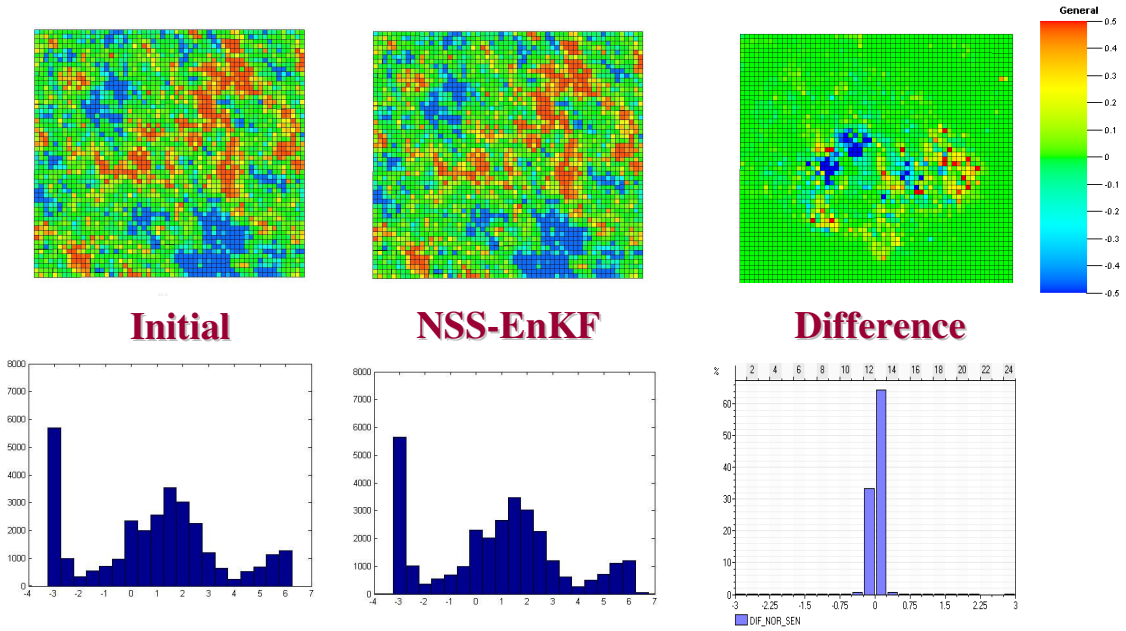


Figure 3.33 The ensemble No. 1 permeability field and spatial histogram and changes from the initial model No. 1; from the left the initial ensemble model No1, the updated ensemble model No. 1 from NST-EnKF and the change that we make from NST-EnKF updating from the initial model No. 1.

3.5 Summary of Chapter III

We demonstrate the characteristics of the EnKF in many sensitivity comparison works. These are the findings to mentions.

- Limit of the updating values should be considered to constrain the parameter estimation solutions
- Measurement Error variance should be assigned properly to maintain the relative contribution of the data to the prior model information to prevent the overconfident on the data and ensemble model collapse after the assimilation.
- Streamline based localization can assist the updating in terms of mitigating overshoot and undershoot problems and maintain the model change small and localized in the influential zone of the flow dynamics
- Normal Score transformation can assist preserving the prior non-Gaussian model parameter distribution through the updating.

In the following chapter, I will approach for these issues by introducing hybrid EnKF approaches.

CHAPTER IV

A HYBRID ENSEMBLE KALMAN FILTER APPLICATION*

In this chapter, we develop two hybrid Ensemble Kalman Filter algorithms; swapping of mean and mode estimation and coarse scale permeability constrained EnKF updating. In the former approach, we calculate the covariance based on a solution from the inversion process. The latter is a sequential approach where updating step in EnKF is divided into 2 steps. First, the ordinary EnKF updating is applied. We then use the mean of the updated model in the inversion algorithm as the initial model and integrate the all history of the observation data from the beginning to the most current time. Then based on the results from the inversion process, we upscale the model and generate the coarse scale permeability data. By using this generated permeability data, we can conduct the second step updating in EnKF to impose the inversion results on the previous EnKF updating results. The requirement of this approach is the good quality of the efficient inversion algorithm and controlling the constraining. We adapt the efficient streamline based Generalized Travel Time Inversion (GTTI) for the WWCT data and low

* Part of this chapter is reprinted with permission from "A Hybrid Ensemble Kalman Filter With Coarse Scale Constraint for Nonlinear Dynamics" by Watanabe S., Datta-Gupta, A., Efendiev, E., Devegowda, D. Paper SPE-124826-MS presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, 4-7 October.

frequency pressure inversion for the WBHP inversion process and demonstrate the advantages of this method through some synthetic examples and a field application.

4.1 Swapping Mean and Mode Estimation Hybrid Approach

As I described in Chapter II sec 2.5.1, the aim for the swapping mean and mode estimation is that we calculate the cross covariance between the model parameters and the model responses based on the mode solution from the inversion process. It is crucial to get a closer solution from the inversion process than the mean of the ensemble model in terms of the parameter solutions (i.e. permeability field). So first we try the same synthetic case as the previous Chapter III, but make the reference case as the average of the initial ensemble model shown in Figure 4.1 ,because the inversion algorithm search the solution from the ensemble mean model. We expect the inversion algorithm, GTTI, captures the true permeability fields by this assumption. We refer to the simple EnKF without hybrid approaches as Plain EnKF and Swapping mean and mode hybrid EnKF as Hybrid-SMM-EnKF in this section.

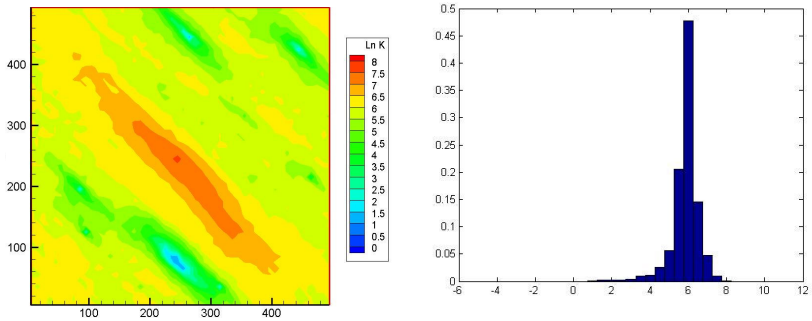


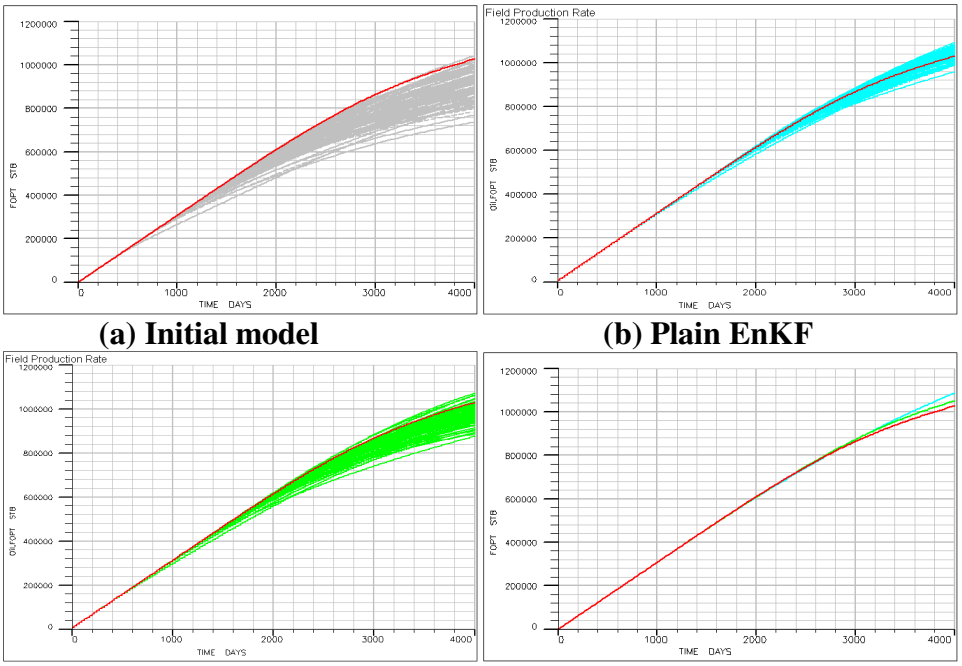
Figure 4.1 Reference model log permeability and its spatial histogram

EnKF parameters for this case are listed below.

EnKF Parameters

- 3 producers and 1 injector
- The number of the ensemble is 99
- Assimilation WWCT observation Data from 800 days to 3200 days (every 100days from 800 days to 2000 days and every 50 days for the rest)
- No covariance localization
- After assimilation, run all the ensemble from time 0 to 4000 days
- Injection Rate 300 RESV/ 3 Production rates 100,150,70 RESV
- State variables { lnK, P, Sw, WWCT}

History matching results are shown in Figures 4.2 and 4.3.



(a) Initial model **(b) Plain EnKF**

(c) Hybrid-SMM-EnKF **(d) Mean model**

Figure 4.2 Field cumulative oil production responses from the ensemble models; the reference model is in red line, Plain EnKF ensemble model responses in light blue, and Hybrid-SMM ones in light green; (a) is the initial ensemble models, (b) is Plain EnKF updated models, and (c) is Hybrid-SMM-EnKF updated models and (d) is the comparison between the mean of Plain EnKF updated ensemble models and the mean of Hybrid-SMM- EnKF updated ensemble models.

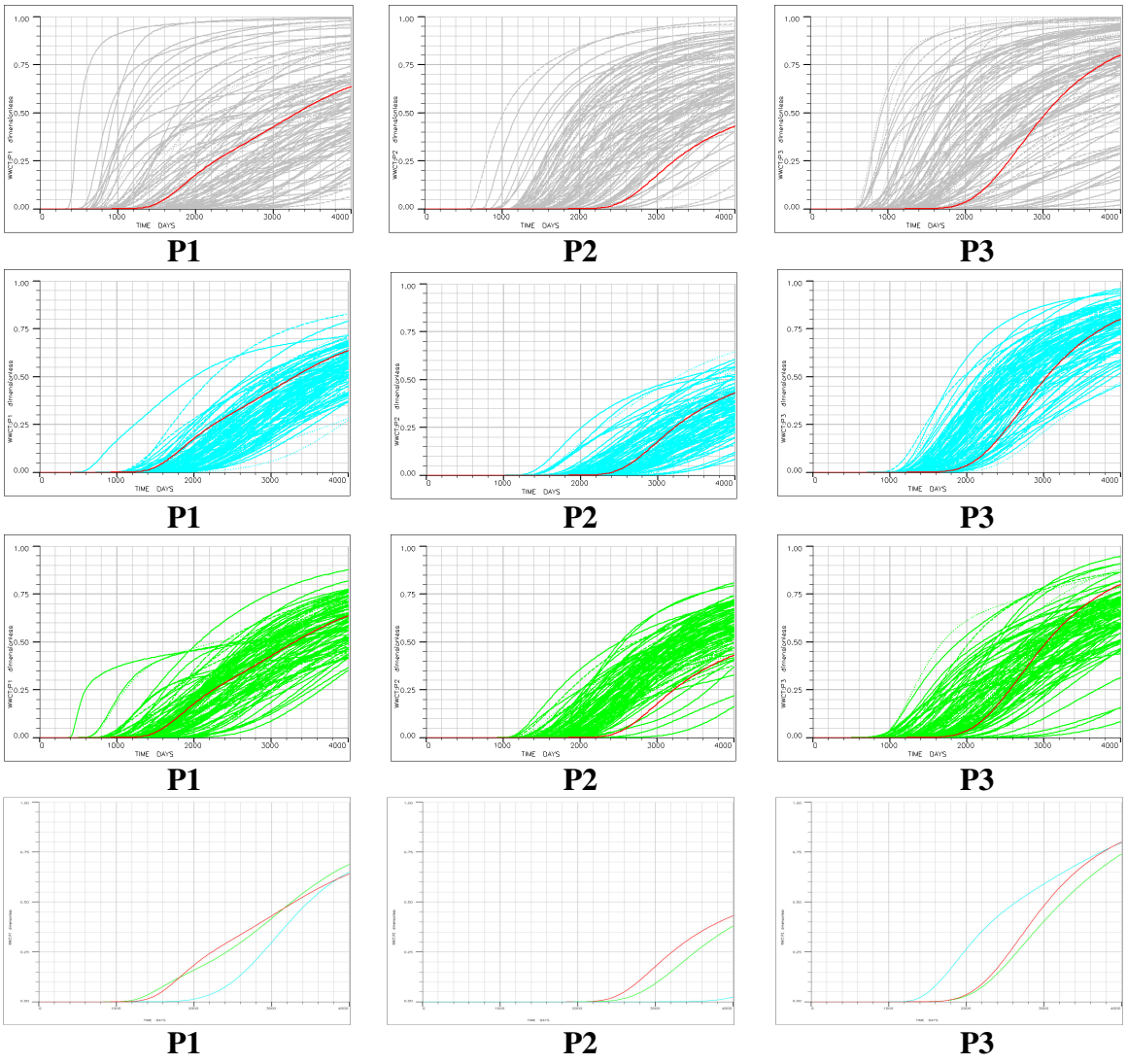
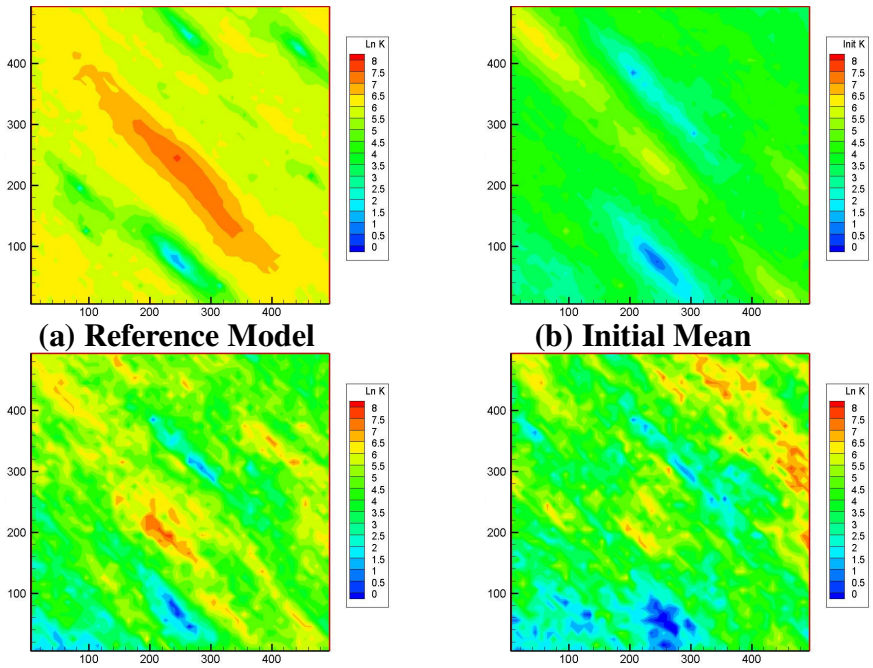


Figure 4.3 Well water cut responses from the ensemble models; the reference model is in red line; the first row is the initial ensemble models, the second row is Plain EnKF updated ensemble models, and the third row is Hybrid-SMM-EnKF-updated ensemble models, the forth row is the comparison between the mean of the updated ensemble models from Plain EnKF and Hybrid-SMM-EnKF.

Both Plain EnKF and Hybrid-SMM-EnKF assimilate the observation data, and spread of the uncertainty in WWCT is reduced from the initial model responses. To be precisely, the Hybrid-SMM-EnKF matching is not as good as the plain EnKF (see the

well P2 in Figure 4.3). The mean of the permeability profile results in figure 4.4 show that the Hybrid-SMM-EnKF captures the high permeability trend in the diagonal line better than Plain EnKF one. And if you look at the prediction from the mean of the final model responses, Hybrid-SMM-EnKF result is better than Plain EnKF one, which implies that the mean estimate was improved by hybrid approach, but the individual model responses did not improve from the better mean estimate. However, the permeability profile results in Figure 4.4 show that Plain EnKF updating suffers from the overshooting in the place where parameters are not resolved by the WWCT observation data in terms of the flow geometry. This indicates the necessity of the covariance localization to eliminate the spurious values in the covariance matrix.



(a) Reference Model **(b) Initial Mean**

(c) Hybrid-SMM EnKF **(d) Plain EnKF**

Figure 4.4 Mean of the ensemble permeability fields comparison; (a) reference model, (b) the initial mean of the ensemble models, and (c) the mean of the final updated ensemble models from Hybrid-SMM EnKF, (c) the mean of the final updated ensemble models from Plain EnKF.

4.1.1 Sensitivity with Initial Model Biasness

Because we can see the biasness in the initial model responses in WWCT from the true model responses in the first row of Figure 4.3, we change the true model to have a bi-modal distribution same as sec.3.1 in Chapter III shown in Figure 4.5 and conduct the Hybrid-SMM-EnKF updating with the same parameter as sec.3.1 in Chapter III. We couple EnKF with the inversion process every 400 days from 2000 days (4 times) inversion.

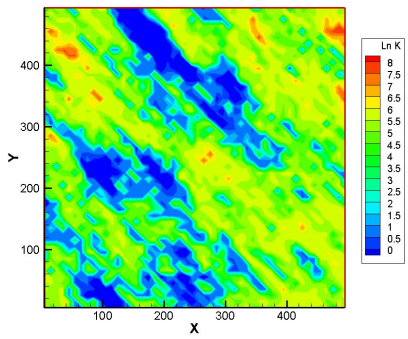
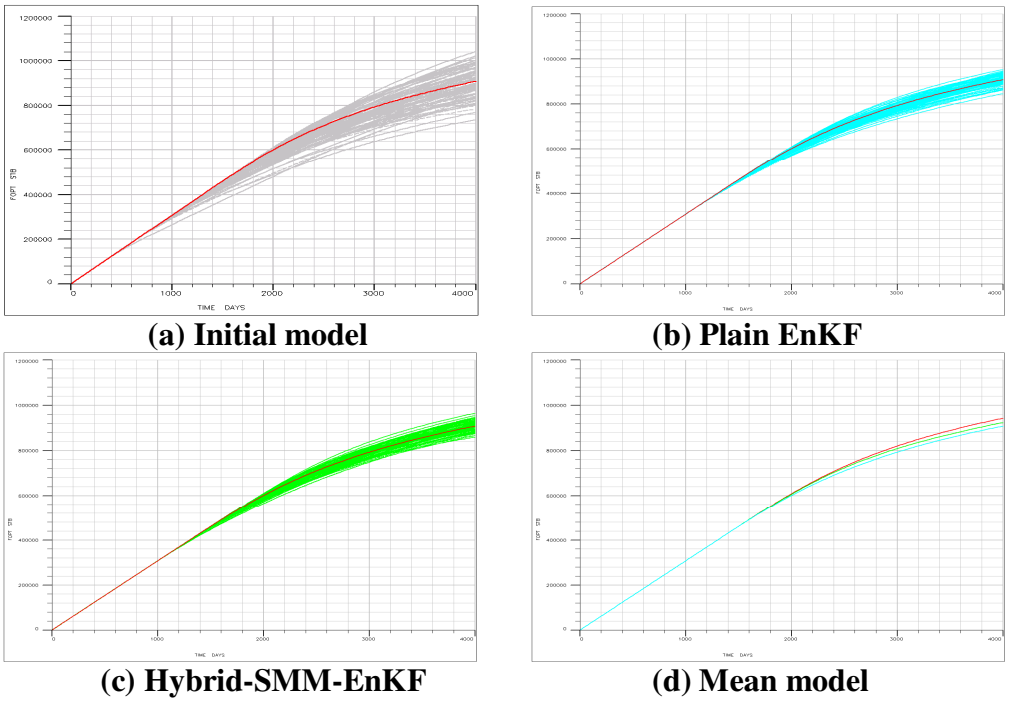


Figure 4.5 Reference model log permeability

The history matching results are shown in Figures 4.6 and 4.7. By coupling the inversion, the matching result for the ensemble mean is improved (see Figure 4.6 (d) and last row of Figure 4.7). As for the updated permeability field comparison, Figure 4.8 shows both Plain EnKF and Hybrid-SMM-EnKF capture the main permeability underlying continuity from the initial mean estimation.



(a) Initial model **(b) Plain EnKF**

(c) Hybrid-SMM-EnKF **(d) Mean model**

Figure 4.6 Field cumulative oil production responses from the ensemble models; the reference model is in red line, Plain EnKF ensemble model responses in light blue, and Hybrid-SMM ones in light green; (a) is the initial ensemble models, (b) is Plain EnKF updated models, and (c) is Hybrid-SMM EnKF and (d) is the comparison between the mean of Plain EnKF updated ensemble models and the mean of Hybrid EnKF updated ensemble models.

In summary for swapping mean and mode estimate hybrid approach, we acknowledge that Hybrid EnKF results depend on the inversion results. We need to develop criteria to judge if a mode solution from the inversion process can be replaced by the ensemble mean model in terms of preserving observation trajectory to achieve better estimation for the cross covariance calculation. And it is noticeable that the better estimation of the mean of the ensemble models doesn't always result in the improvement

on the each ensemble model matching quality. This leads us to the next hybrid approach where we constrain each ensemble model updating toward the better mode estimation by sequential coarse scale permeability data updating.

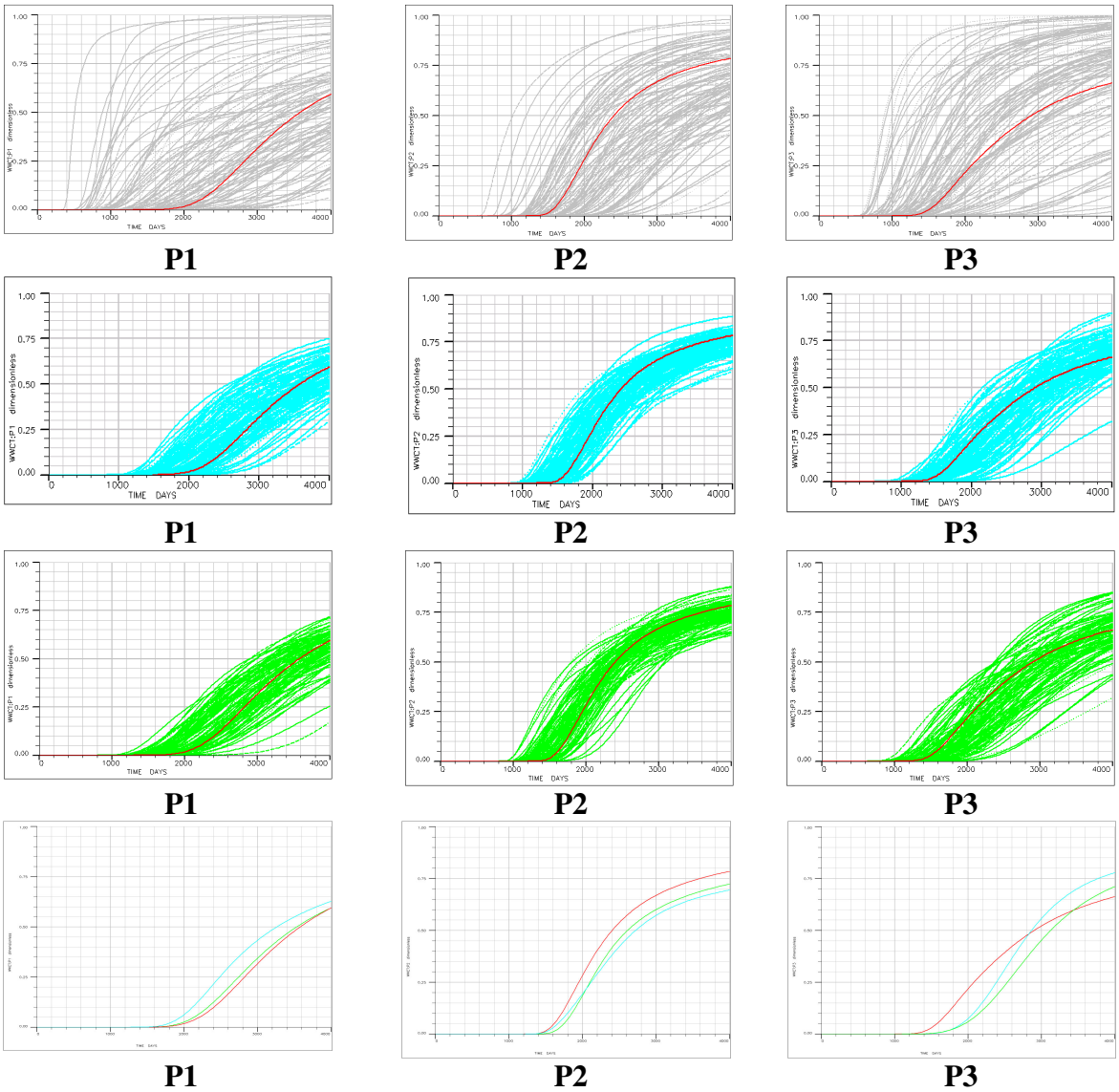
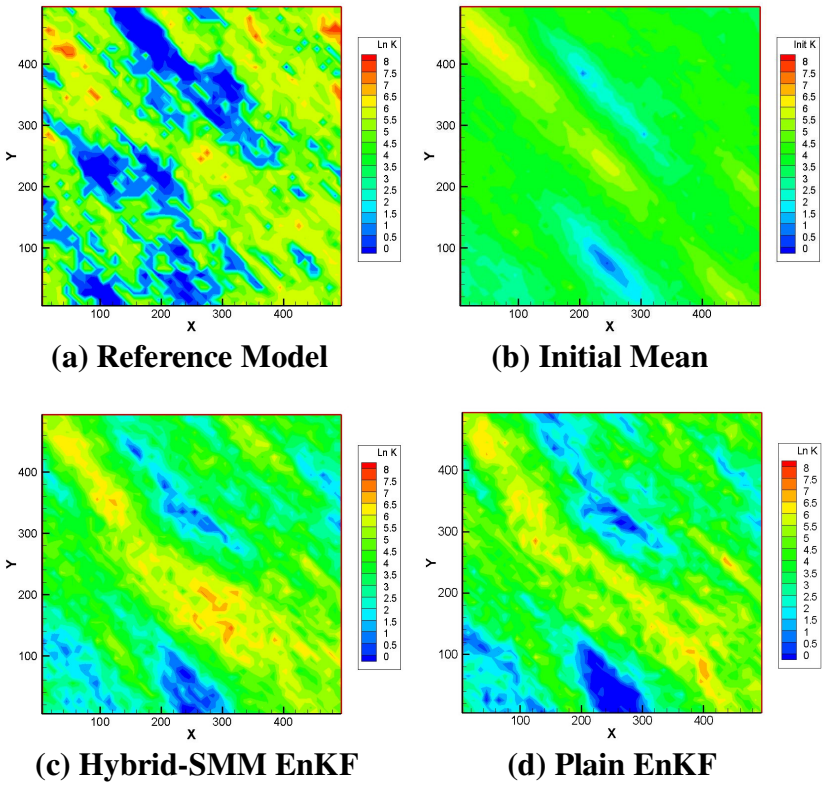


Figure 4.7 Well water cut responses from the ensemble models; the reference model is in red line; the first row is the initial ensemble models, the second row is Plain EnKF updated ensemble models, and the third row is Hybrid-SMM EnKF-updated ensemble models, the forth row is the comparison between the mean of the updated ensemble models from Plain EnKF and Hybrid-SMM-EnKF.



(a) Reference Model **(b) Initial Mean**

(c) Hybrid-SMM EnKF **(d) Plain EnKF**

Figure 4.8 Mean of the ensemble permeability fields comparison; (a) reference model, (b) the initial mean of the ensemble models in the center, and (c) the mean of the final updated ensemble models from Hybrid-SMM EnKF, (d) the mean of the final updated ensemble models from Plain EnKF.

4.2 Hybrid EnKF with Coarse Scale Permeability Constraint

Motivation for this approach is that we know that we can get better parameter estimation from inversion process if the data misfit function or objective function is minimized successfully. And if we can impose that solution on the EnKF updating such that the each model parameter updating is constrained toward the mode solution, we can achieve the better data matching and uncertainty around the mode estimation from the

ensemble model simultaneously. However, we need to investigate the characteristics of this approach with associated key parameters and identify the strength and weakness of the approach. In this section we test the sensitivity of this approach with respect to the following parameters:

1. Upscaling factor
2. Coarse permeability data error variance

We create a 5 spot synthetic case with parameters listed below.

EnKF Parameters

- 50×50×1 synthetic model.
- 1025 realization generated by sgsim
- Oil water 2 phase incompressible model
- Adverse Mobility Ratio (Oil viscosity =10 cp, Water viscosity =1 cp)
- 5 spot pattern well configuration(I1,P1~P4)
- Assimilation Step(1200-3000 days by 200days) 9 step
- Coarse scale permeability constraint assimilation step(1200, 1800, 2400,3000)
- Prediction to 4000 days
- State variables { lnK, P, Sw, WWCT} for the 1st assimilation step
- State variables { lnK, P, Sw, LnK} for the 2 nd assimilation step

The reference model (No. 1015) from the generated initial models is shown in Figure 4.9.

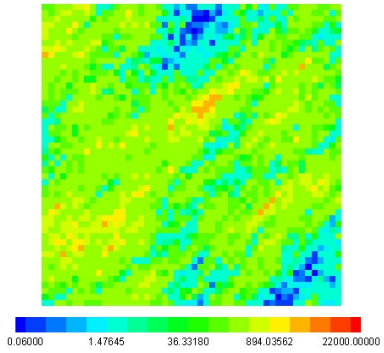
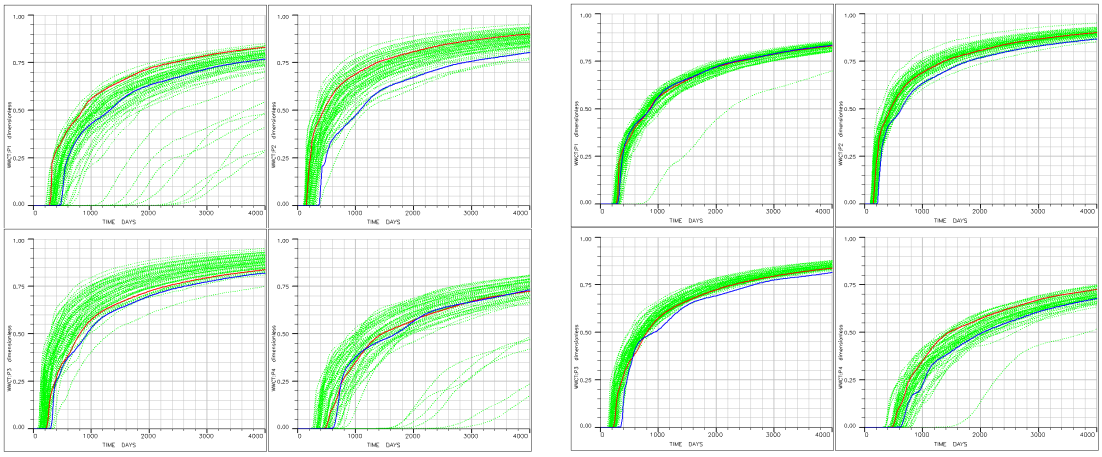


Figure 4.9 Reference model log permeability field



(a) Initial model **(b) Plain EnKF**

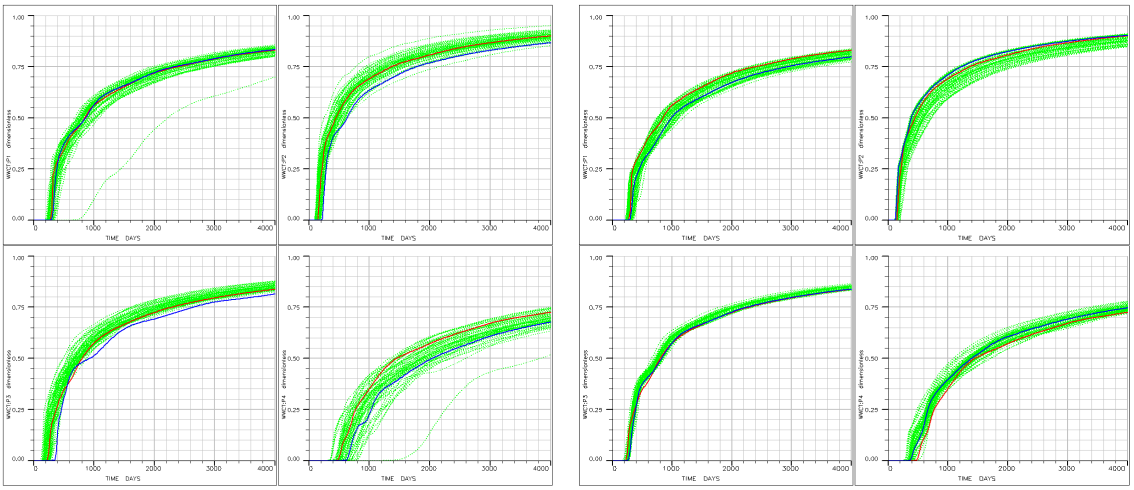
Figure 4.10 Well water cut responses comparison from the ensemble models; the reference model is in red line, ensemble model responses in light green lines and the response from the mean ensemble model in blue line, (a) Initial ensemble models, (b) Plain EnKF updated models, in the top row from left, P1 and P2 and the bottom row from left, P3 and P4.

Based on the reference model, run the forward simulation and generate the WWCT observation data. First, we conduct Plain EnKF for this case and the results are shown in Figure 4.10. we can see a better matching quality after Plain EnKF.

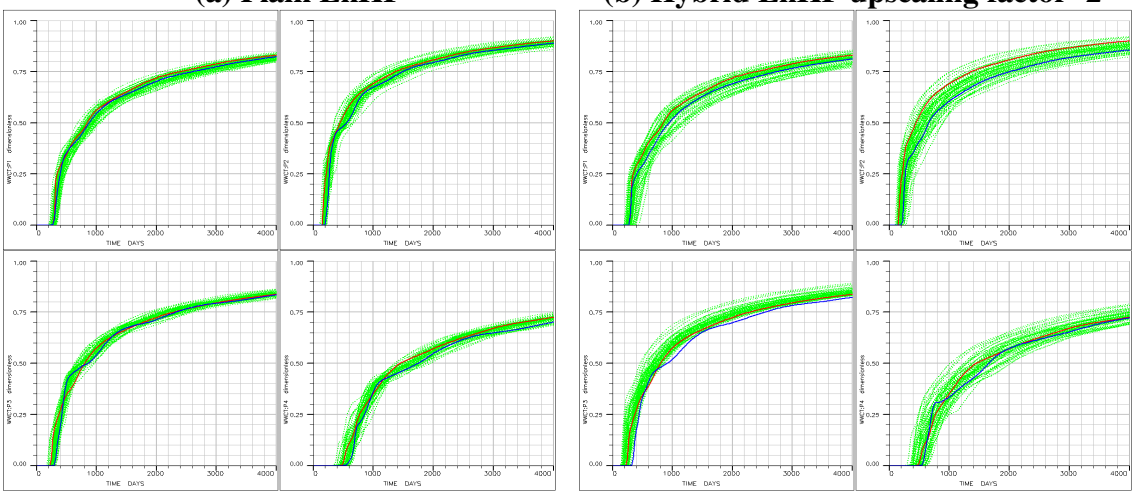
4.2.1 Effect of the Upscaling Factor

The first key parameter that needs to be investigated is upscaling factor which defines the coarse scale dimensions for our approach. We adapt flow based upscaling method to generate the coarse permeability data. For this case, the upscaling is done uniformly. For example, if we assign upscaling factor equal to 2, in this case, upscaled model dimensions become 25 by 25 from 50 by 50 uniformly. The size of the grid block

is upscaled by 2 in both x and y direction. We conduct Hybrid EnKF for this case by changing upscale factors with the fixed other parameters.



(a) Plain EnKF **(b) Hybrid EnKF upscaling factor=2**



(c) Hybrid EnKF upscaling factor=5 **(d) Hybrid EnKF upscaling factor= 10**

Figure 4.11 Well water cut responses comparison from the ensemble models; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line; (a) Plain EnKF updated models, (b) Hybrid EnKF updated models with upscaling factor=2, (b) Hybrid EnKF updated models with upscaling factor=5, and (c) Hybrid EnKF updated models with upscaling factor=10, in the top row from left, P1 and P2 and the bottom row from left, P3 and P4.

Figure 4.11 shows the comparison results in matching WWCT data. It shows the upscaling factor 5 (10 by 10 from 50 by 50) provides the best matching in the updated model responses.

Coarse scale permeability data validity is checked by conducting simulation for the upscaled permeability field to see the discrepancy in water cut response between fine scale model and coarse scale models. Figure 4.12 shows the WWCT responses from the upscaled permeability fields by changing the upscaling factor value.

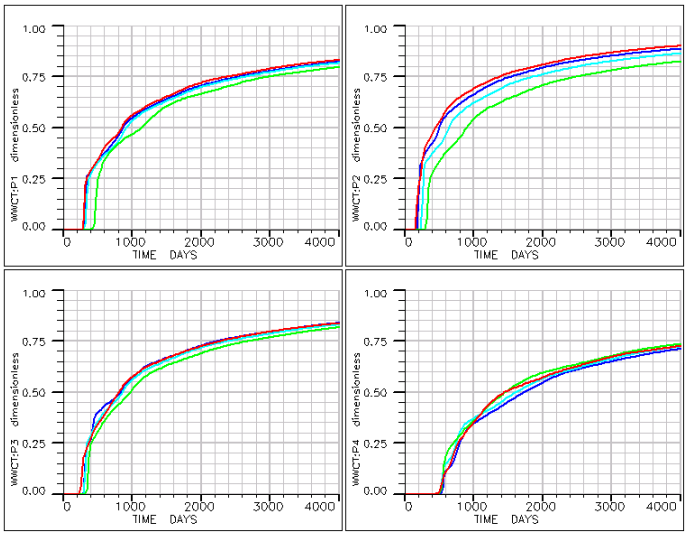


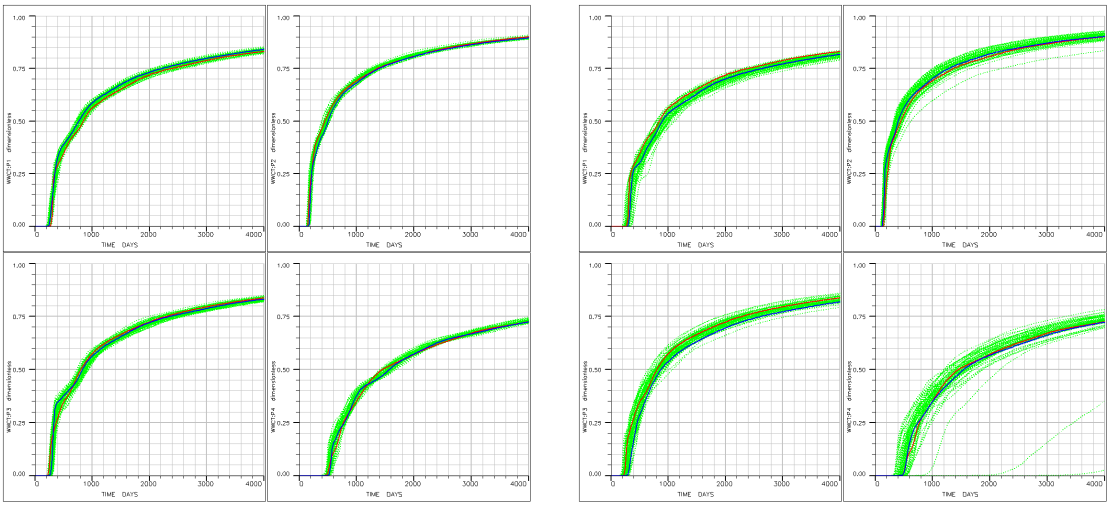
Figure 4.12 Well water cut responses comparison from the upscaled permeability models; the reference model is in red line, the upscaled model with upscaling factor=2 in blue, the upscaled model with upscaling factor=5 in light blue, the upscaled model with upscaling factor=10 in light green, in the top row from left, P1 and P2 and the bottom row from left, P3, P4.

From Figure 4.12, we need to acknowledge that the upscaling scheme introduce the other source of the uncertainty and discrepancy between the fine scale model and coarse scale model. The more the upscaling factor becomes, the more discrepancy can be

seen in the model responses. And the more upscaling factor becomes, the less number of the coarse permeability data are generated. This results in the less reduction of the spread or uncertainty of the updated model responses because of the less constraining of the coarse scale permeability data.

4.2.2 Effect of the Coarse Scale Data Error Variance

We can assign the uncertainty of the coarse scale permeability data by the variance of the data. We conduct the sensitivity of the assimilation by changing the values. The upscaling factor is fixed to be 5 and assign the coarse permeability error standard deviation as lnK=1 and lnK=4.



(a) Error Std lnK=1 **(b) Error Std lnK=4**

Figure 4.13 Well water cut responses comparison from the ensemble models; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line; (a) Updated models with error std lnK=1, (b) Updated models with error std lnK=4.

As Figure 4.13 shows, the more variance of the data is assigned, the final updated model responses are less assimilated in terms of the spread around the observation data. This indicates that the less data error variance causes the ensemble collapse of the updated models and we need to assign the appropriate data variance to reduce the uncertainty and prevent the ensemble collapse.

4.3 Another Synthetic Case

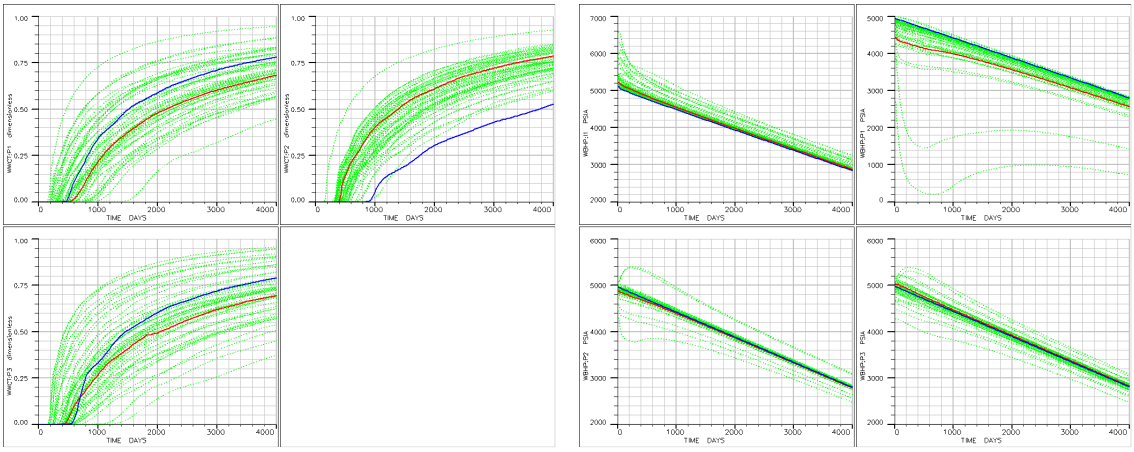
We apply Hybrid EnKF for another synthetic case. This case is set up from the same synthetic case as in sec.3.1 in Chapter III, but we change the fluid properties such as oil and water viscosity to be adverse mobility ratio to make the water breakthrough faster. 50 ensemble initial models are selected arbitrary and every initial model spatial permeability histogram has bi-modal distribution. For this case, we integrate not only WWCT data but also WBHP data. We expect the different kinds of observation improve the parameter estimation and provide the better history matching results. From the work in Chapter III, we apply normal score transformation of the permeability values to preserve the initial bi-modal distribution of the spatial permeability histogram. The default EnKF parameters are listed below.

EnKF Parameters

- 3 producers and 1 injector
- The number of the ensemble is 50
- Adverse mobility Ratio (Oil viscosity =1.5 cp, Water viscosity =0.3 cp)

- Assimilation 3 WWCT observation Data and 4 WBHP data from 1200 days to 3000 days by every 200days
- Measurements error 1 % of the maximum values WWCT 0.01 and WBHP 50 psi
- No covariance localization
- Up scaling factor=2
- Normal score transformation of the permeability values
- After assimilation, run all the ensemble from time 0 to 4000 days
- Injection Rate 300 RESV/ 3 Production rates 100,150,70 RESV
- State variables { lnK, P, Sw, WWCT} for the 1 st assimilation step
- State variables { lnK, P, Sw, LnK} for the 2 nd assimilation step
- Coarse scale permeability constraint assimilation for 3 times (1800, 2400, 3000)

Initial model responses are shown in Figure 4.14. It shows that there is quite large uncertainty in the initial model responses especially for WWCT. And it is important to mention that the mean model prediction in WWCT is far from the reference model response due to both nonlinearity of the problem and the non-Gaussian prior model distribution.

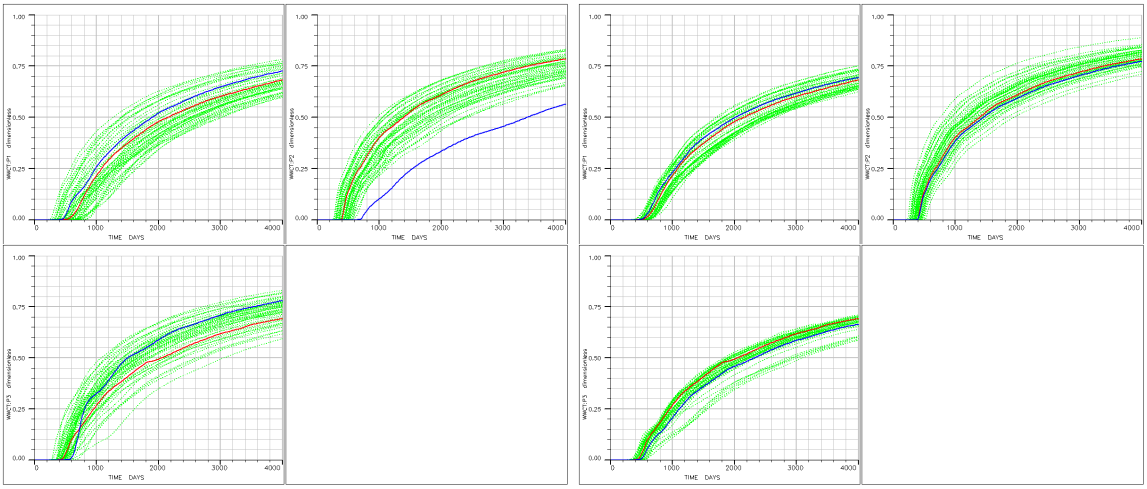


(a) WWCT **(b) WBHP**

Figure 4.14 Initial 50 ensemble model WWCT and WBHP responses; the reference model is in red line, ensemble model responses in light green lines and the response from the initial mean ensemble model in blue line, in the top row from the left, P1, P2, and P3 in the bottom row.

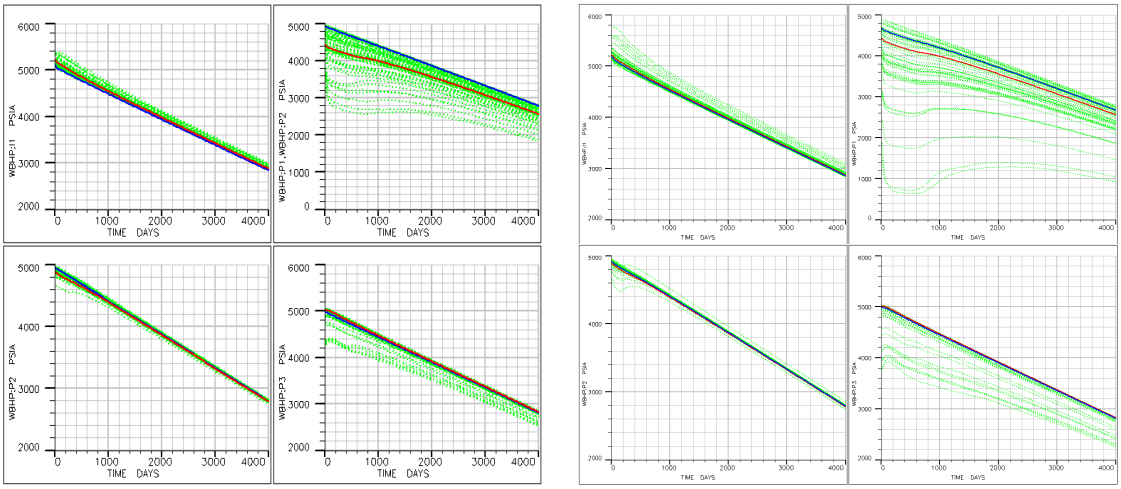
4.3.1 Test of the Validity of the Approach

In order to prove the validity of the hybrid approach, the reference model is used to generate the coarse scale permeability by the flow based upscaling method. Therefore there is no inversion process for this case. We expect that the coarse scale permeability data from the reference model constrain each ensemble model updating in EnKF toward the reference model permeability field. Figures 4.15 and 4.16 show WWCT and WBHP matching comparison between Plain EnKF and Hybrid EnKF. WWCT results support that the reference coarse scale permeability data successfully constrain each ensemble model updating in EnKF and updated model responses match the reference model one very well. And importantly, the prediction from the mean ensemble model also follows the true model trajectory and within the ensemble model responses. As you can see in the permeability profile and distribution results in Figure 4.17 and 4.18, the hybrid results capture the main trend of the permeability fields (low and high regions) quite well and the mean of the ensemble model also capture the underlying geological features and the spatial histogram, which indicates the all ensemble model has driven toward the reference model by imposing the coarse permeability data. That promises the benefit of our hybrid approach as long as we can generate the plausible coarse scale permeability data.



(a) Plain EnKF **(b) Hybrid EnKF**

Figure 4.15 Well water cut responses comparison from the ensemble models; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line; (a) Plain EnKF updated models, (b) Hybrid EnKF updated models with reference coarse scale permeability data, in the top row from the left, P1, P2, and P3.in the bottom row.



(a) Plain EnKF **(b) Hybrid EnKF**

Figure 4.16 Well bottom-hole pressure responses comparison from the ensemble models; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line; (a) Plain EnKF updated models, (b) Hybrid EnKF updated models with reference coarse scale permeability data, in the top row from the left, I1 and P1, and P2 and P3 in the bottom row.

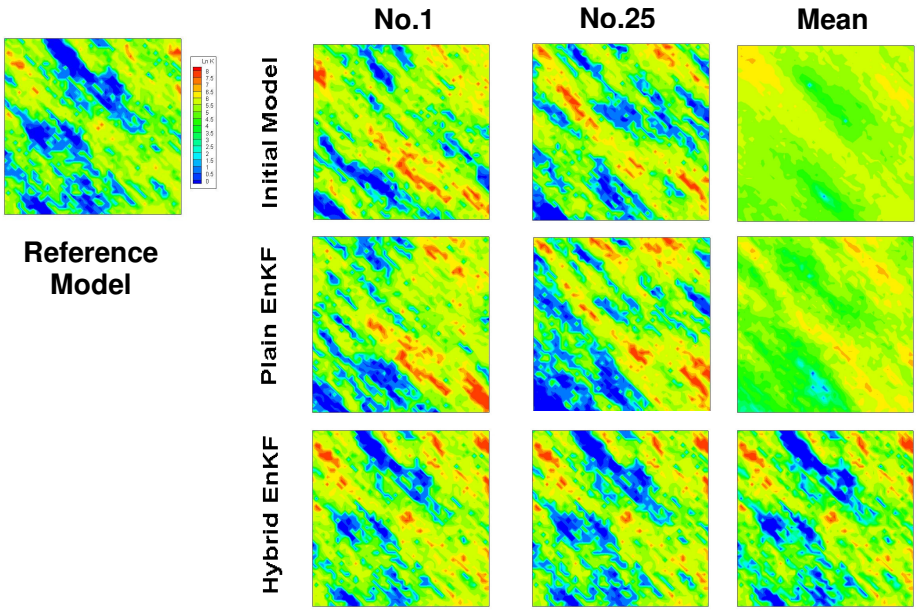


Figure 4.17 The ensemble permeability fields comparison; in the first row from left, the reference model, initial model No. 1and No. 25, the initial mean model, and in the second row from left, Plain EnKF updated model No. 1and No. 25, the updated mean model, and in the third row from left, Hybrid EnKF updated model No. 1and No. 25, updated mean model.

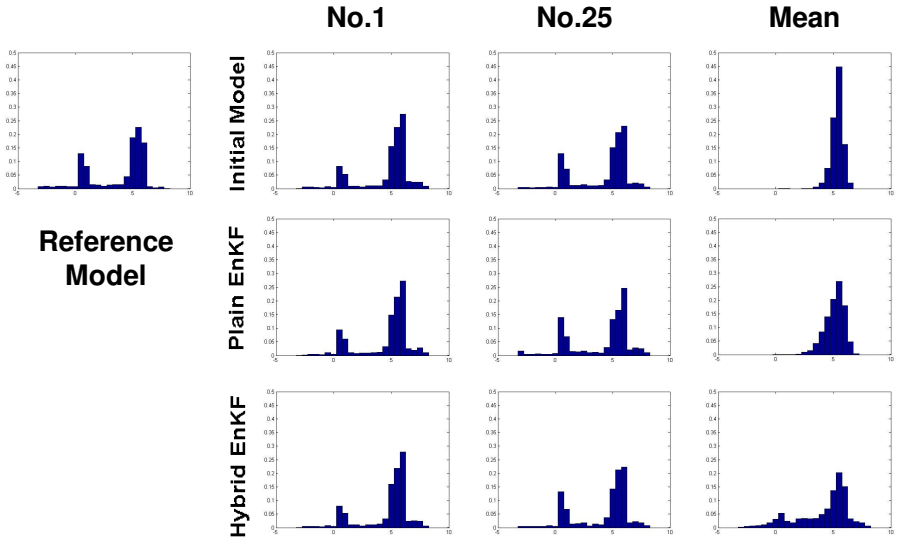


Figure 4.18 The ensemble spatial permeability histogram comparison; in the first row from left, the reference model, initial model No. 1and No. 25, the initial mean model, and in the second row from left, Plain EnKF updated model No. 1and No. 25, the updated mean model, and in the third row from left, Hybrid EnKF updated model No. 1and No. 25, the updated mean model.

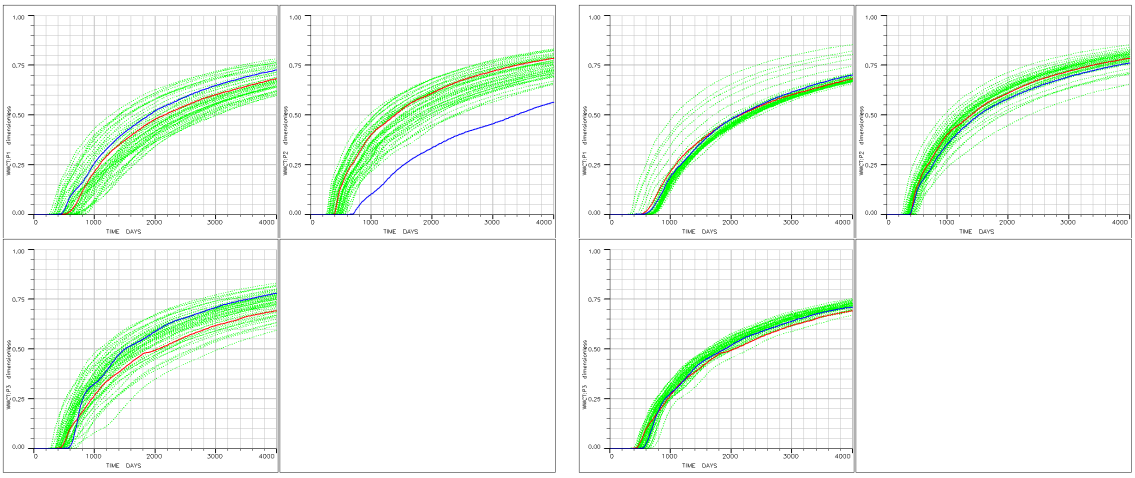
4.3.2 Inversion Coupling Hybrid EnKF

Now we couple the inversion process with hybrid EnKF by coarse scale permeability constraint. We use the same synthetic case as in sec.4.3.1 to see if the inversion process can capture the true permeability field and successfully constrain EnKF updating toward the better mode estimation. Parameters we use for this case are listed below.

Hybrid EnKF Parameters

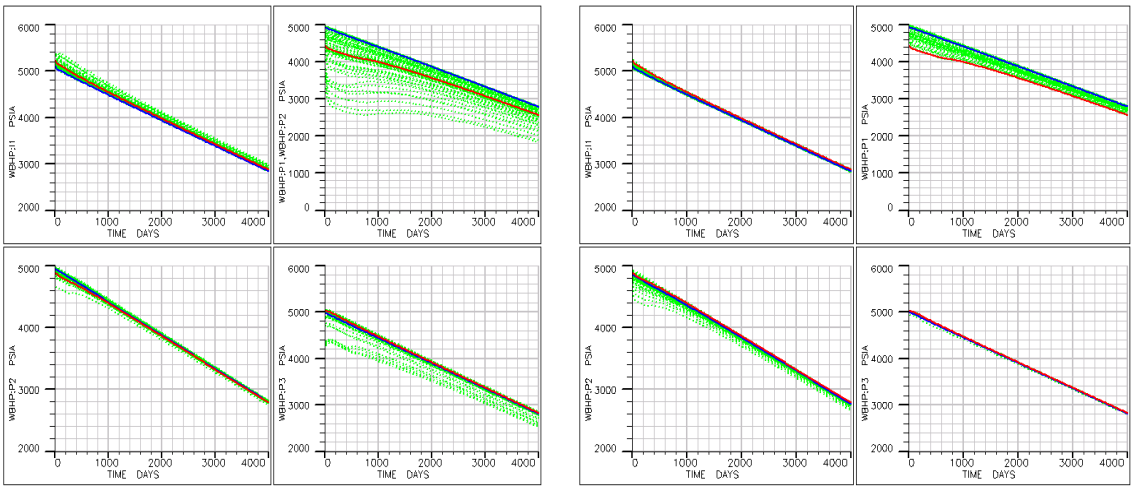
- Assimilation 3 WWCT observation Data and 4 WBHP data from 1200 days to 3000 days by every 200days
- Measurements error 1 % of the maximum values WWCT 0.01 and WBHP 50 psi
- Assign Coarse scale data error lnk=1
- No covariance localization
- Up scaling factor=2
- GTTI inversion for WWCT data
- Normal score transformation of the permeability values
- After assimilation, run all the ensemble from time 0 to 4000 days
- State variables { lnK, P, Sw, WWCT,WBHP} for the 1 st assimilation step
- State variables { lnK, P, Sw, LnK} for the 2 nd assimilation step
- Coarse scale permeability constraint assimilation for 3 times (1800, 2400, 3000)

Figures 4.19 and 4.20 show the history matching results. It shows that Hybrid EnKF outperform Plain EnKF in terms of the spread of the responses around the true model response trajectory although there some models show a deviated WWCT responses in P1. It is clear that all ensemble model responses are constrained to the mode solution from the inversion and the final mean prediction of WWCT follows the true



(a) Plain EnKF **(b) Hybrid EnKF**

Figure 4.19 Well water cut responses comparison from the ensemble models; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line; (a) Plain EnKF updated models, (b) Hybrid EnKF updated models, in the top row from the left, P1, P2, and P3 in the bottom row.



(a) Plain EnKF **(b) Hybrid EnKF**

Figure 4.20 Well bottom-hole pressure responses comparison from the ensemble updated models; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line; (a) Plain EnKF updated models, (b) Hybrid EnKF updated models, in the top row from the left, I1, P1, and P2, P3 in the bottom row.

model response trajectory. But the WBHP matching results does not show a significant improvement.The inversion process worked successfully and its results are shown in Figure 4.21 below. Final model matches the observation data much better than the initial model for the inversion process. Figure 4.22 shows the relative objective function behavior through iteration in the GTTI inversion process. It is monotonically decreasing by 80% in amplitude misfit at the last iteration. Both results support the successful inversion results.

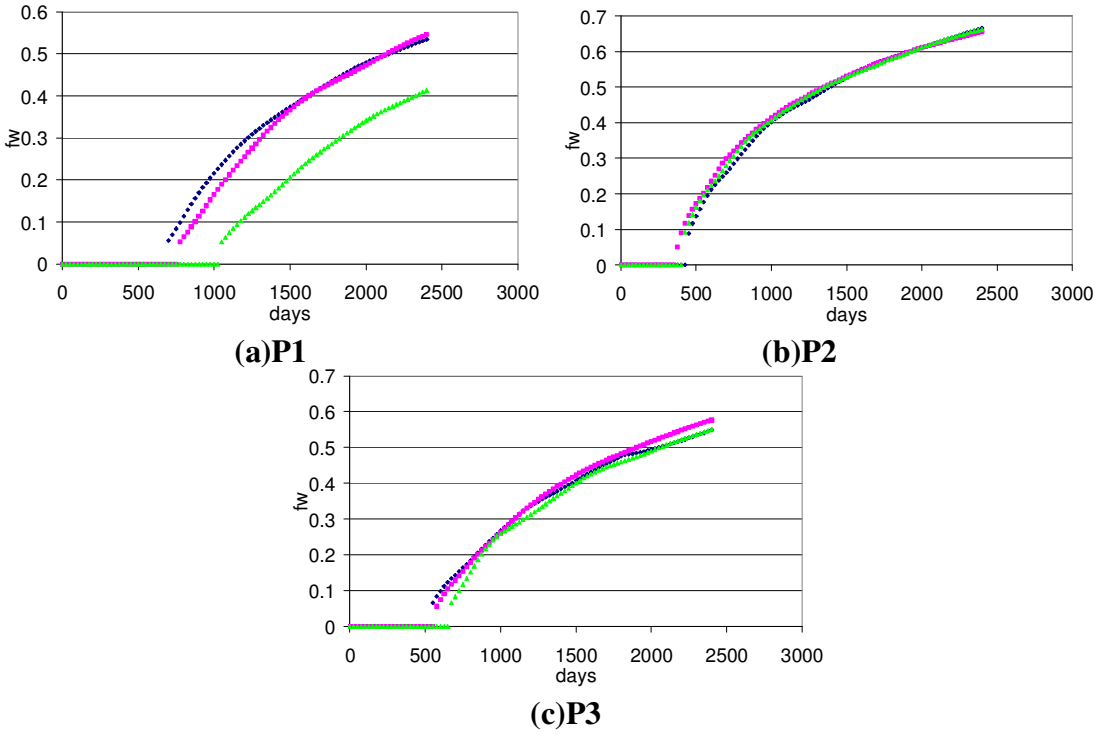


Figure 4.21 Well water cut inversion matching results at 2400 days; the reference model response is in blue point, initial model response in light green triangles and the final updated model response in pink square, (a) P1, (b) P2, and (c) P3 respectively.

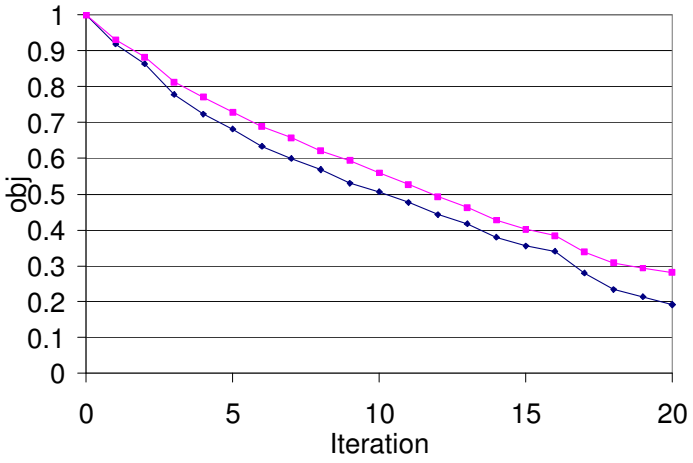


Figure 4.22 Relative objective function of the GTTI inversion process; the GTTI misfit is in blue line, and the amplitude misfit is in pink line.

4.4 Another Nine Spot Synthetic Case

To demonstrate the advantage of Hybrid EnKF approach, we set up a synthetic case where Plain EnKF does not work satisfactorily. Evansen(2004) shows that for a linear model in the case without model errors, the EnKF solution at all times is a combination of the initial ensemble members. Although our model is nonlinear, the updated model has a same trend of the permeability features from the experimental studies shown in Chapter III. The reference permeability model has a long range, high-permeability continuity in the east-west direction as shown in Figure 4.23. A total of 50 initial ensemble members were generated using sequential Gaussian simulation (Deutsche and Journel 1992). To illustrate the difficulties in the conventional EnKF, the initial ensemble members were deliberately chosen to have a north-south orientation of the long range continuity as seen in Figure 4.23. This example is not far removed from

the actual field situations because we seldom have a good prior knowledge of the permeability covariance and its orientation. Our objective is to assess the performance of the conventional and hybrid EnKF formulation in reproducing the large scale features of the reference model. Additionally, the log-permeability histogram was chosen to be bimodal. These model characteristics were chosen to illustrate some of the difficulties associated with the EnKF especially in non-Gaussian settings with inaccurate prior model statistics. For the EnKF application, we first apply a normal score transformation on the model parameters (permeability) and update the normal scored values using the EnKF and back transform to the physical space after the assimilation step. By doing that, we can also preserve the prior model distribution and prevent the overshooting and under shooting through updating.

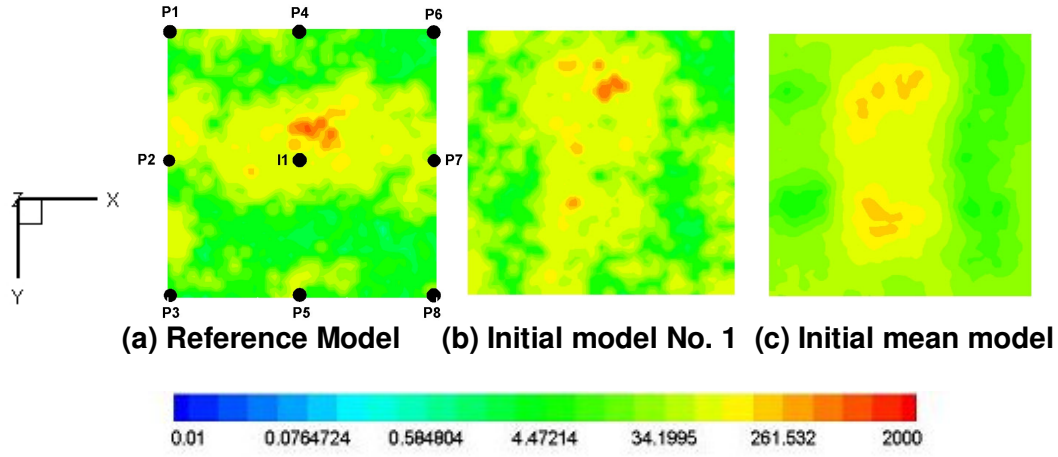


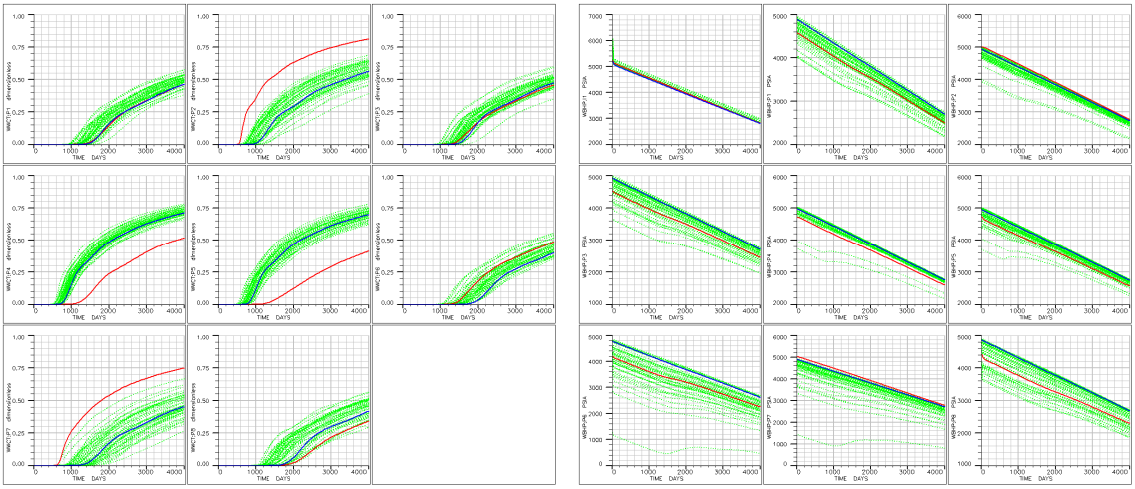
Figure 4.23 Permeability fields; from left (a) reference model, (b) initial model No. 1 and (c) initial mean model

EnKF parameters for this nine spot synthetic case are listed below.

EnKF Parameters

- 50×50×1 Synthetic Case
- 2 Phase oil and water
- 8 producers and 1 injector nine spot
- Depletion well rates control(8 Producers: Equal RSEV 30 rB/day Injector: RSEV 230 rb/day)
- Adverse mobility ratio (Oil Viscosity=1.5 cp, Water Viscosity=0.5 cp)
- The number of the ensemble is 50
- Normal Scare transformation for permeability
- No Localization
- Measurements error 1 % of the maximum values WWCT 0.01 and WBHP 50 psi
- Coarse scale data error lnk=1
- Assimilation 1 WWCT/ 4 WBHP observation Data from 200 days to 3000 days by 200 days for 15 times
- Coarse scale permeability constraints assimilation for 3 times (1400,2200,3000)
- After assimilation, run all the ensemble from time 0 to 4000 days

Initial model WWCT and WBHP responses are shown in Figure 4.24. As we can expect, the initial model responses are significantly different from the reference model, especially in wells P2, P4, P5, P7 because the direction of the permeability continuity is quite different from the reference model and consequently the water front movements have different behavior. This results in large discrepancies in the WWCT data. However, the pressure responses have less discrepancy compared to the reference model because the bottomhole pressure behavior is mostly sensitive to the permeability distribution near the well locations. What we can expect from this initial model construction is that if EnKF updating is biased to the prior models, the updated models from EnKF is far from the reference model in terms of the high permeability continuity distribution.

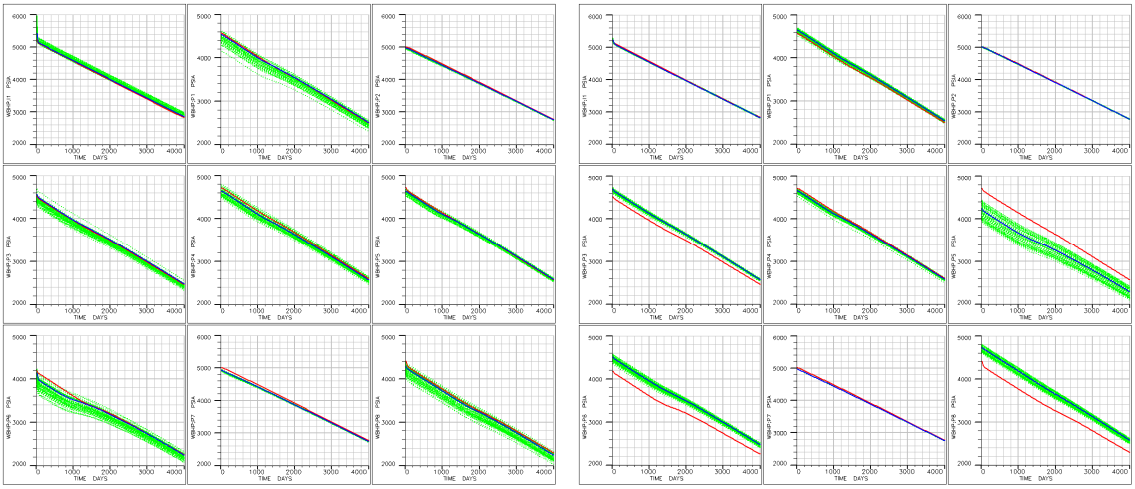


(a) WWCT **(b) WBHP**

Figure 4.24 Initial 50 ensemble model WWCT and WBHP responses; the reference model is in red line, ensemble model responses in light green lines and the response from the initial mean ensemble model in blue line, (a) WWCT responses in the first row from the left, P1, P2, and P3, in the second row P4, P5, and P6, and in the third row from left P7 and P8, (b) WBHP responses in the first row from the left, I1, P1, and P2, in the second row P3, P4, and P5, and in the third row from left P6, P7, and P8.

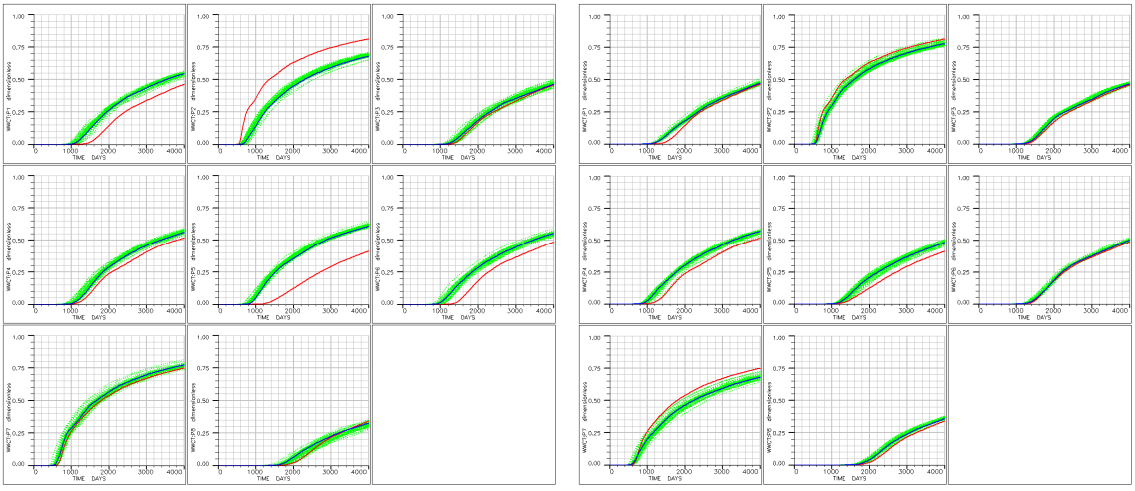
4.4.1 Comparison of Plain EnKF and Hybrid EnKF

We conduct Plain EnKF and Hybrid EnKF for this case and compare the matching results shown in Figures 4.25 and 4.26 below. As it shows in the WWCT data matching results, Hybrid EnKF shows the better matching to the observation data than Plain EnKF ones especially for the wells (P2) located in the x-axis high permeability zone of the reference model and P4 located in low permeability zone of the reference model.



(a) Plain EnKF **(b) Hybrid EnKF**

Figure 4.25 Ensemble updated model WBHP matching results comparison between (a) Plain EnKF (b) Hybrid EnKF; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line, in the first row from the left, I1, and P1, P2, in the second row P3, P4, and P5, and in the third row from left P6, P7 and P8.



(a) Plain EnKF **(b) Hybrid EnKF**

Figure 4.26 Ensemble updated model WWCT matching results comparison between (a) Plain EnKF and (b) Hybrid EnKF; the reference model is in red line, ensemble model responses in light green lines and the response from the mean updated ensemble model in blue line, in the first row from the left, P1, P2, and P3, in the second row P4, P5, and P6, and P7, P8 in the third row from left.

If you look at the permeability field results in Figure 4.27, Plain EnKF updated model suffers overshooting and undershooting values. It seems that overshooting comes from the small size of the ensemble member and from the small observation variance assignment (1%). For Hybrid EnKF updated model, overshooting and undershooting problems was mitigated and they capture the east-west high permeability continuity better than Plain EnKF ones. However, there is the low permeability area clustered near the well P5 which degraded the WBHP matching severely. This problem stems from the inversion results which show the same low permeability area near the well P5. The crucial cause for this artifact is that we conduct the inversion only on WWCT data by

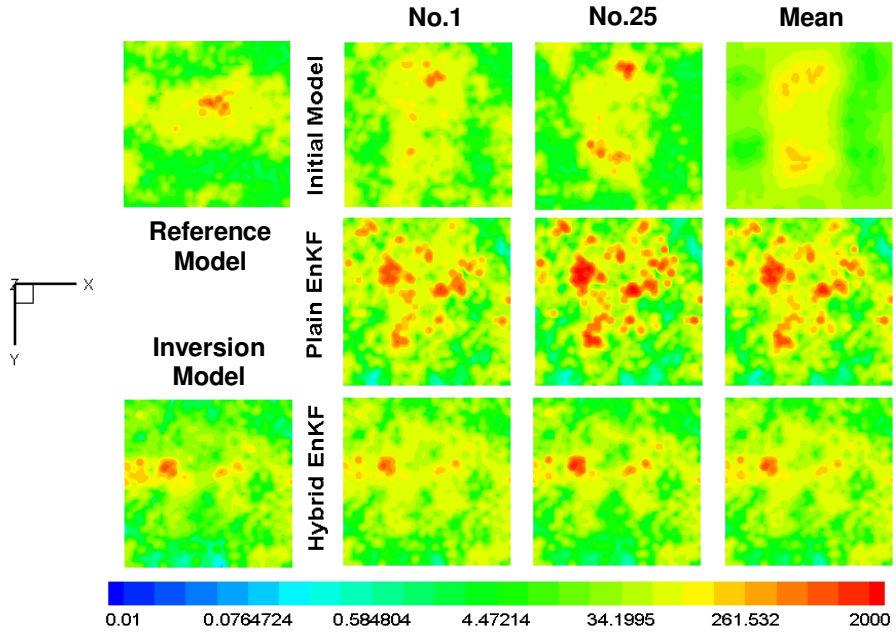


Figure 4.27 Permeability fields comparisons between the initial models, Plain EnKF updated models and Hybrid EnKF updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left Inversion mode model, Hybrid EnKF updated model No. 1, No. 25, and the updated mean ensemble model.

GTTI algorithm which produces the low permeability barrier near the well to delay the water break through effectively. And in the context of Hybrid EnKF approach, we need the simultaneous inversion with WWCT and WBHP data to resort this problem. For the permeability distribution in Figure 4.28, Hybrid EnKF results preserve the initial spatial histogram better than Plain EnKF, and the ensemble mean model also has bi-modal distribution more clearly for Hybrid EnKF than Plain EnKF.

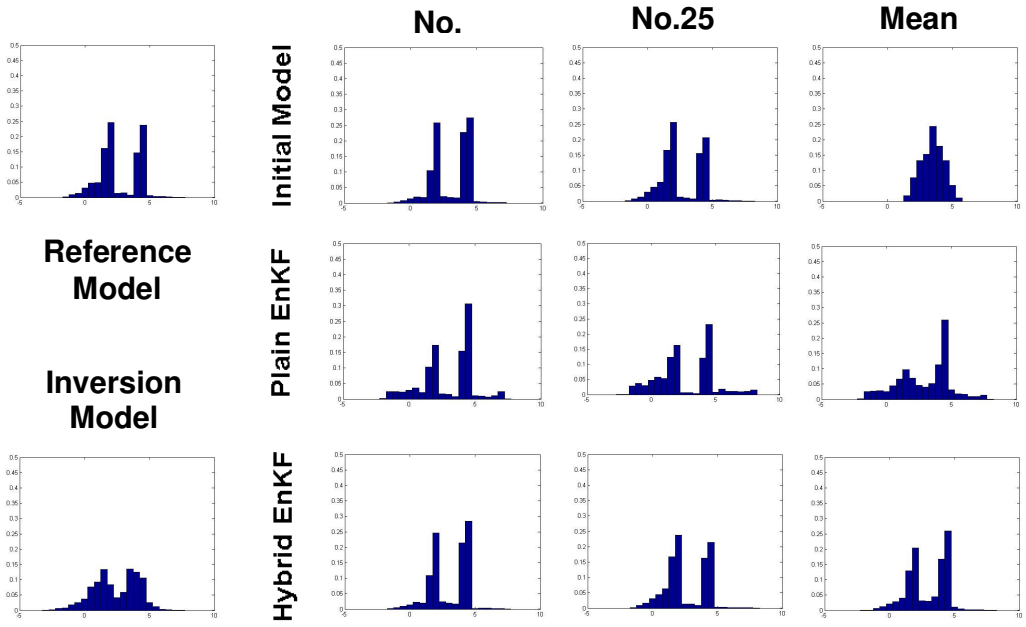
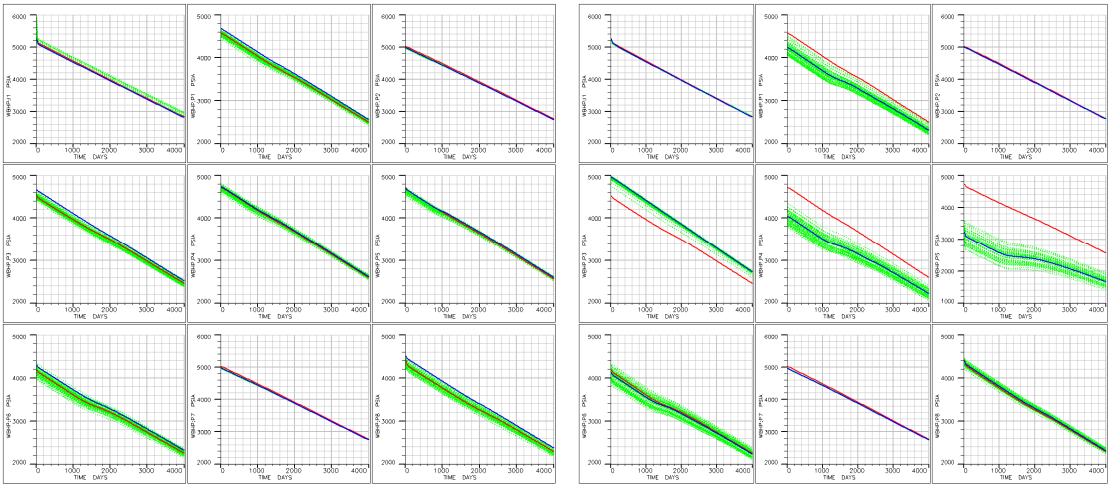


Figure 4.28 Permeability spatial histogram comparisons between the initial models, Plain EnKF updated models and Hybrid EnKF updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left inversion solution model, Hybrid EnKF updated model No. 1, No. 25, and the updated mean ensemble model.

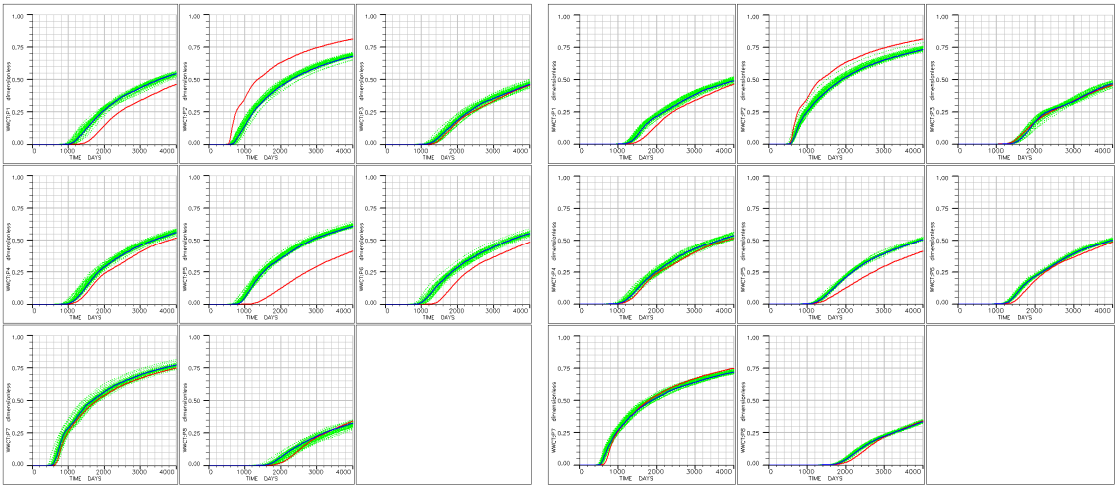
4.4.2 Hybrid EnKF with Streamline Trajectory Localization

We know that streamline trajectory localization prevents overshoot and undershoot problems and preserve the prior model distribution from Chapter III experimental studies. Thus, we test the effect of the covariance localization on top of the coarse scale permeability hybrid approach in this section. As for history matching results shown in Figures 4.29 and 4.30, Hybrid EnKF with localization results are degraded matching both in WWCT and WBHP. Also, Plain EnKF could prevent the overshooting and undershooting problems considerably shown in Figure 4.31. As we discuss in Chapter III, streamline trajectory localization preserve the prior model spatial distribution by identifying the influential zone intersected by streamlines. So the most of the streamline are traced in the initial north-south high permeability region, the model updating are restricted with in that region. Consequently the east-west high permeability regions are not captured by that localization. And the inversion result also is deteriorated from that cause. Figure 4.32 shows the spatial permeability histogram of the models.



(a) Plain EnKF-ST **(b) Hybrid EnKF-ST**

Figure 4.29 Ensemble updated model WBHP matching results comparison between (a) Plain EnKF with streamline trajectory localization and (b) Hybrid EnKF with streamline trajectory localization; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line, in the first row from the left, I1, and P1, P2, in the second row P3, P4, and P5, and in the third row from left P6, P7 and P8.



(a) Plain EnKF-ST (b) Hybrid EnKF-ST

Figure 4.30 Ensemble updated model WWCT matching results comparison between (a) Plain EnKF with streamline trajectory localization and (b) Hybrid EnKF with streamline trajectory localization; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line, in the first row from the left, P1, P2, and P3, in the second row P4, P5, and P6, and in the third row from left P7 and P8.

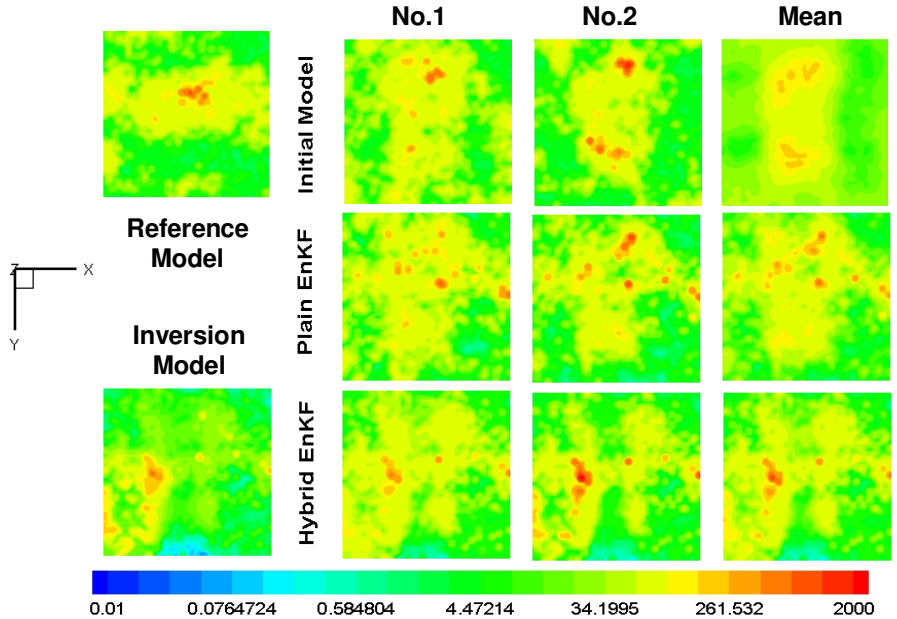


Figure 4.31 Permeability fields comparisons between the initial models, Plain EnKF-ST updated models and Hybrid EnKF-ST updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF-ST updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left inversion solution model, Hybrid EnKF-ST updated model No. 1, and No. 25, the updated mean ensemble model.

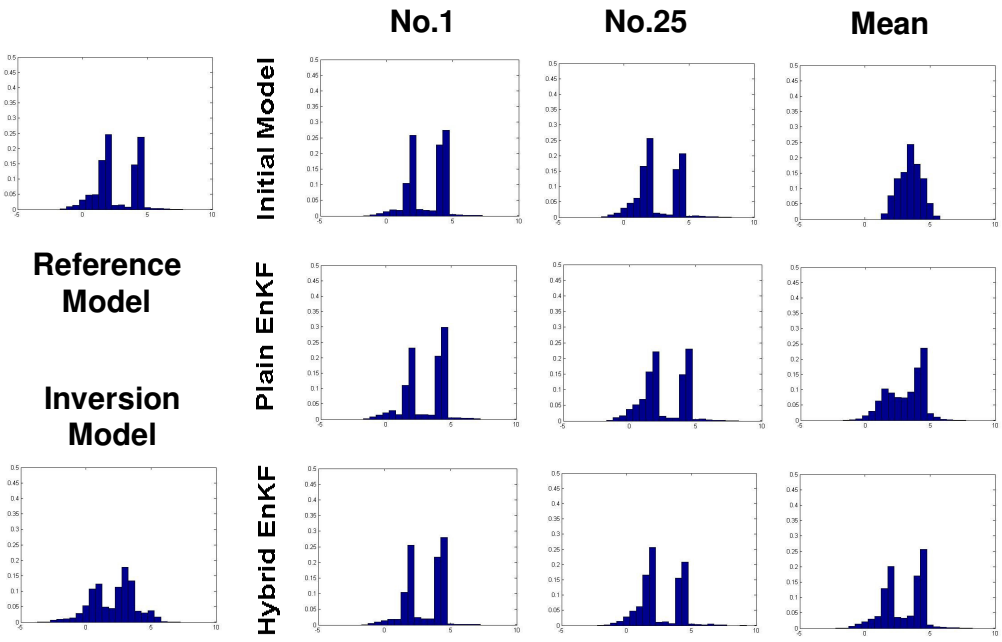


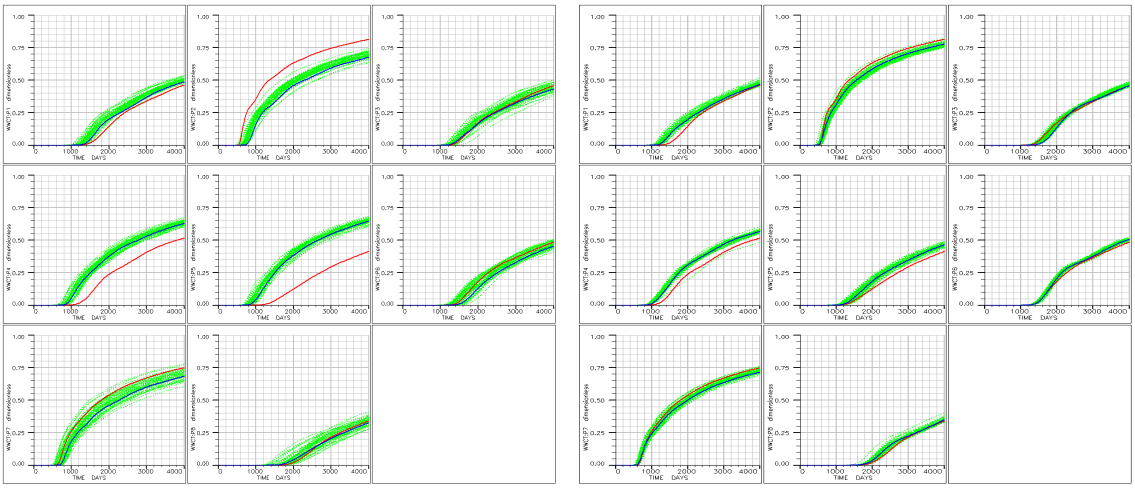
Figure 4.32 Permeability spatial histogram comparisons between the initial models, Plain EnKF-ST updated models and Hybrid EnKF-ST updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF-ST updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left inversion solution model, Hybrid EnKF-ST updated model No. 1, No. 25, and the updated mean ensemble model.

A noticeable thing is that the histogram of the mean updated model shows a clear bimodal distribution which indicates the updated ensemble models are driven toward a mode model. In other words, the distribution of the ensemble mean model is similar to each individual model distribution.

4.4.3 WWCT and WBHP Simultaneous Inversion

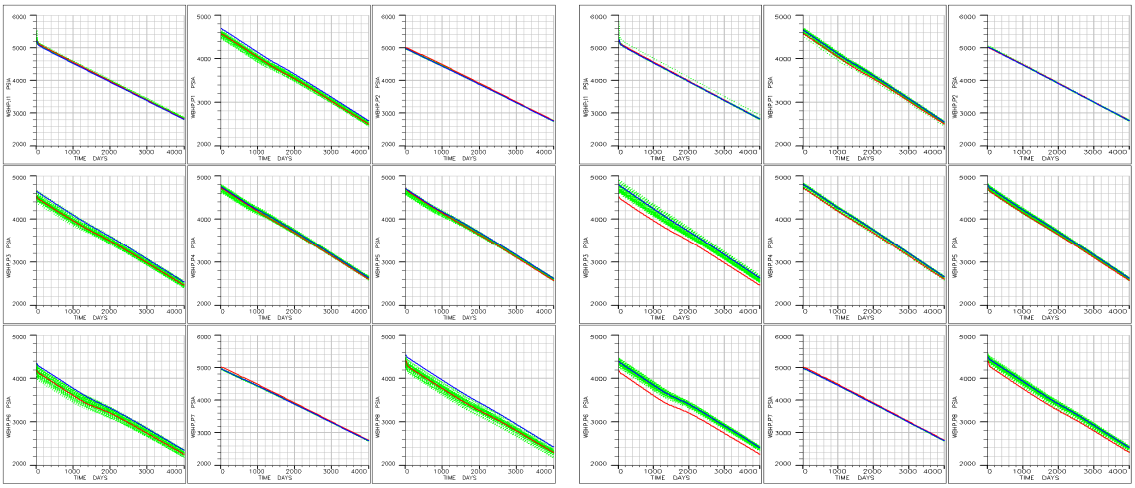
Based on the previous results, we found the necessity to improve the inversion results to achieve the advantage of the hybrid EnKF approach over Plain EnKF. We

carried out a simultaneous inversion of the WWCT and WBHP data starting with the ensemble mean at three different assimilation times as discussed before. For WBHP inversion algorithm, we adapt the low frequency pressure inversion (Vasco and Karasaki 2006). The detail of the algorithm is referred to the paper and is out of the scope of this study. We used streamline-based covariance localization for the conventional EnKF as described in Arroyo et al. (2008). In this approach, the streamline trajectories are used to demarcate regions within the reservoir which will have an influence from the production response and to mitigate spurious correlations through localization. For the hybrid EnKF, no covariance localization was applied. In the results that follow, we denote the hybrid EnKF results as Hybrid EnKF, while the conventional EnKF results as Plain EnKF for no covariance localization and Plain EnKF-ST for streamline-based covariance localization respectively. The WWCT and WBHP history matching results are shown in Figure 4.33 and Figure 4.34 respectively. Hybrid EnKF shows a better matching of the observation data than the conventional EnKF, especially for the well P2 and P4 which are located in the high and low permeability zones of the reference model respectively. The updated permeability field comparisons are shown in Figure 4.35 and clearly indicate improved resolution of the permeability field with the hybrid formulation.



(a) Plain EnKF **(b) Hybrid EnKF**

Figure 4.33 Ensemble updated model WWCT matching results comparison between (a) Plain EnKF and (b) Hybrid EnKF; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line, in the first row from the left, P1, P2, and P3, in the second row P4, P5, and P6, and in the third row from left P7, P8.



(a) Plain EnKF **(b) Hybrid EnKF**

Figure 4.34 Ensemble updated model WBHP matching results comparison between (a) Plain EnKF and (b) Hybrid EnKF; the reference model is in red line, ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line, in the first row from the left, I1, and P1, P2, in the second row P3, P4, and P5, and in the third row from left P6, P7 and P8.

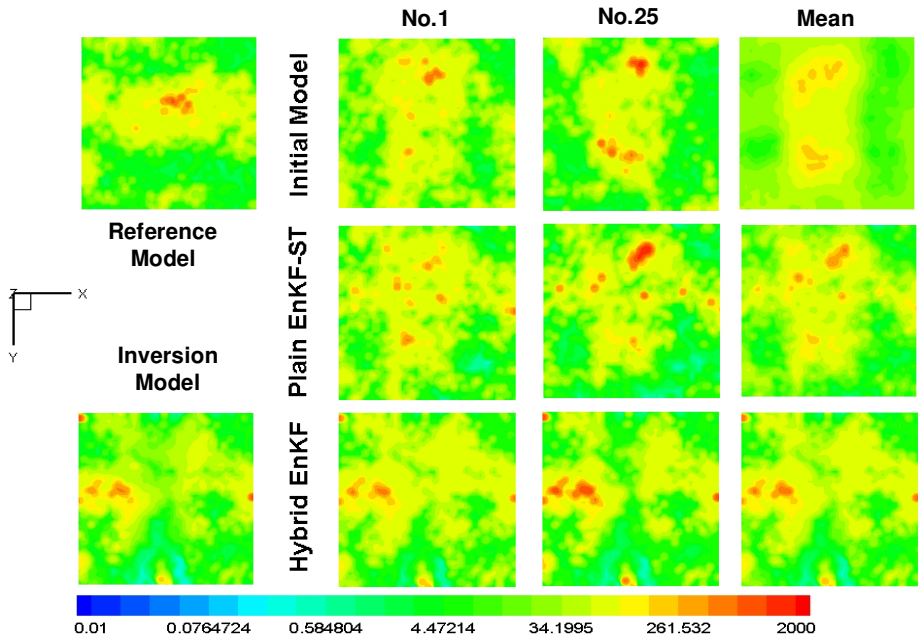


Figure 4.35 Permeability fields comparisons between the initial models, Plain EnKF updated models and Hybrid EnKF updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left inversion solution model, Hybrid EnKF updated model No. 1, and No. 25, the updated mean ensemble model.

Specifically, for Hybrid EnKF, overshooting and undershooting problems are mitigated and the updated models capture the high permeability continuity in the east-west direction much better compared to the conventional EnKF. The results also indicate that for the hybrid approach, there is no need for separate covariance localization as it is already embedded in the inversion solution and the coarse-scale constraints. In terms of the spatial histogram of permeability, Figure 4.36 shows that the bimodal permeability distribution is preserved both in the individual ensemble models and also the mean of the ensemble models for Hybrid EnKF. This is because the non-linear inversion is able to

capture the high contrast in the permeability field and all the ensemble models are constrained to this solution.

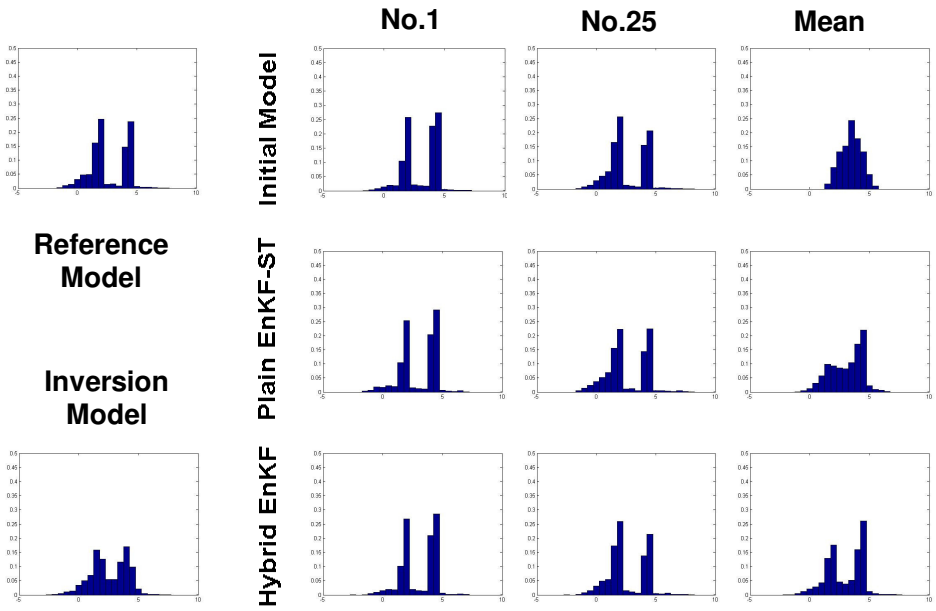


Figure 4.36 Permeability spatial histogram comparisons between the initial models, Plain EnKF updated models and Hybrid EnKF updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left inversion solution model, Hybrid EnKF updated model No. 1, and No. 25, the updated mean ensemble model.

4.5 Uncertainty Quantification Methods for Hybrid EnKF

The previous discussion illustrated the benefits of the hybrid EnKF formulation in terms of reproducing the large-scale features of the reference model. We now extend the discussion to uncertainty quantifications in the posterior model estimates and compare

the results in sec 4.4.3 from the hybrid approach with the conventional EnKF formulation.

4.5.1 Water Front Movement

The orientation of the permeability continuity can easily be inferred from the direction of the movement of the water front at different times as shown in Figures 4.37 and 4.38.

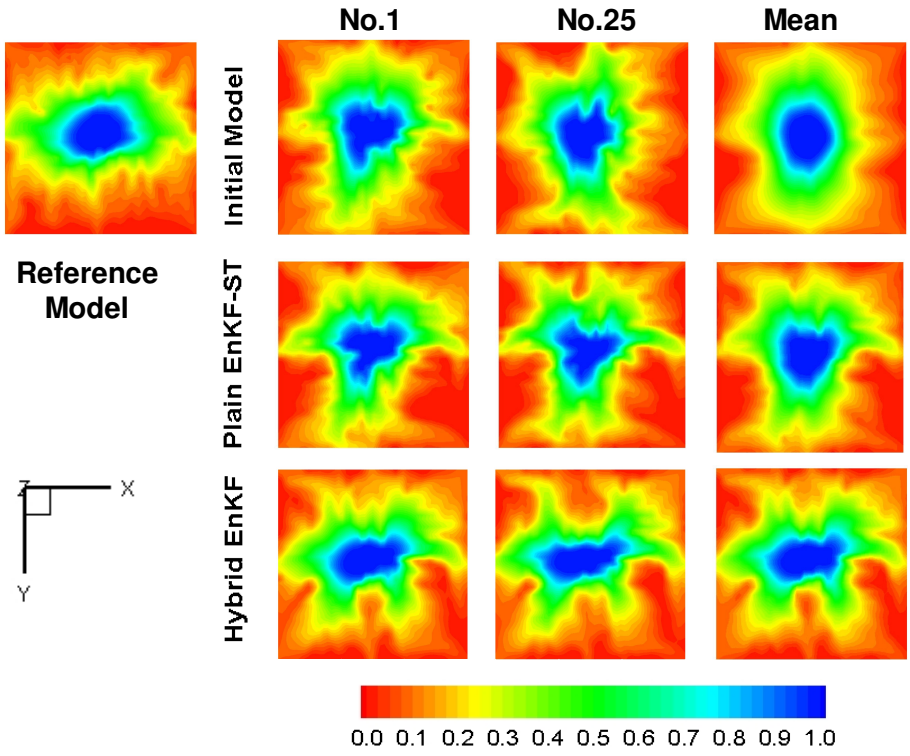


Figure 4.37 Water saturation map at 2500 days comparisons between the initial models, Plain EnKF updated models and Hybrid EnKF updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left inversion solution model, Hybrid EnKF updated model No. 1, and No. 25,the updated mean ensemble model.

The results clearly show the preferential water displacement in the east-west direction for Hybrid EnKF which is consistent with the reference model water front movement. The conventional EnKF, on the other hand, is unable to capture the water front movement in the reference model because of insufficient prior knowledge of the permeability orientation.

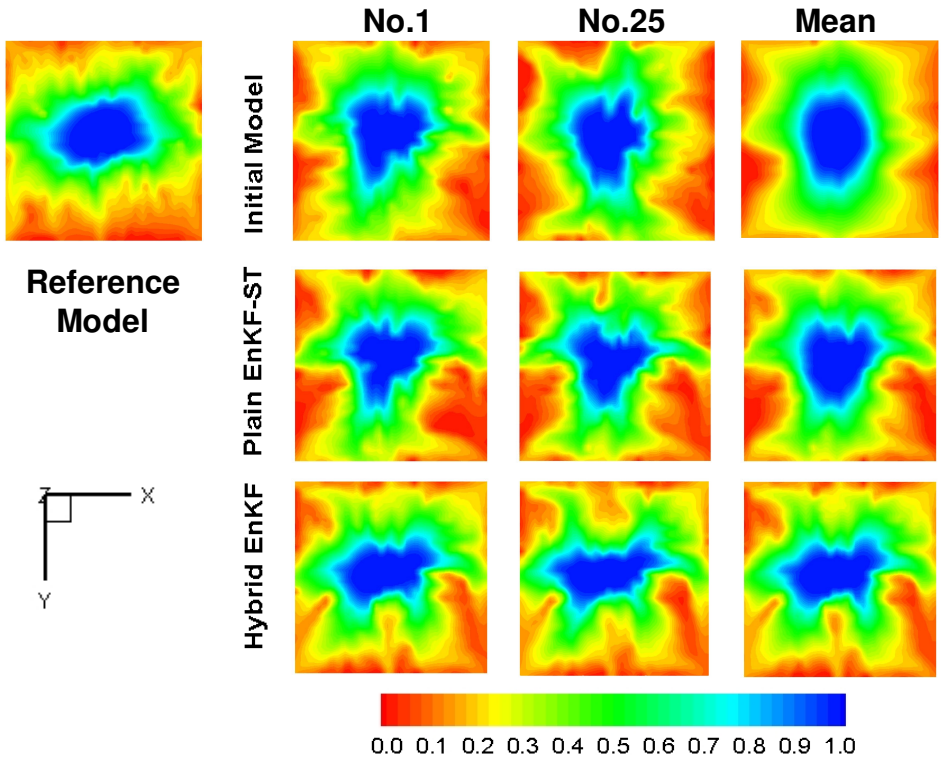


Figure 4.38 Water saturation map at 4000 days comparisons between the initial models, Plain EnKF updated models and Hybrid EnKF updated models; in the first row from left reference model, initial model No. 1, and No. 25, the initial mean model, in the second row, Plain EnKF updated model No. 1, No. 25, and the updated mean ensemble model, in the third row, from left inversion solution model, Hybrid EnKF updated model No. 1, and No. 25, the updated mean ensemble model.

4.5.2 Multi-dimensional Scaling of the Swept Volume Changes

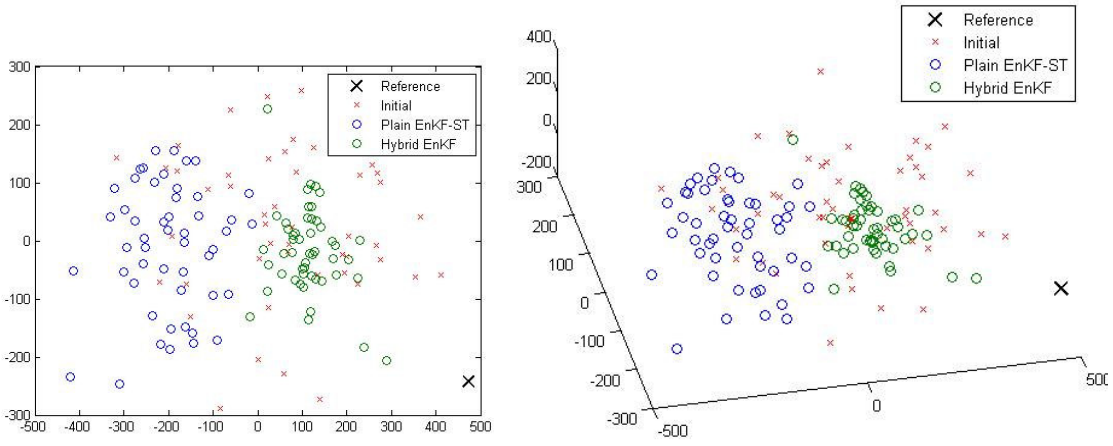
An uncertainty analysis can be conducted using the multi-dimensional scaling (MDS) to visualize the updated ensemble members and their separation. Scheidt and Caers (2007) proposed taking into account the dynamic responses of the models for ranking an ensemble of models. They used fast streamline simulators for flow simulations and then selected members based on the dissimilarities in the response using the kernel principal component analysis (KPCA) and the k-mean clustering methods. We adopt a similar approach here except that the dissimilarities are based on the evolution of the reservoir swept volume with time computed using various thresholding of the streamline time-of-flight. Thus, the dynamics of the flow field is not characterized using a single composite quantity such as ultimate swept volume or recovery but by examining how the swept volume or recovery evolves as a function of time. Also, the dynamic response for the model is approximated only by tracing of streamlines and computing the time of flight without making full flow simulations. These streamline trajectories are generated using the fluid-flux information from a finite-difference simulator. The evolution of swept pore volumes for various time-of-flight threshold values is considered as the dissimilarity measure shown in Eq. (4.1)

$$\Delta SPV(t=\tau_1-\tau_2)=\left|SPV\left(t=\tau_1\right)-SPV\left(t=\tau_2\right)\right|, \tag{4.1}$$

and the connectivity distance between the swept volume changes of two individual realization i and j is defined by

$$\delta_{ij} = \sum_{t=1}^{N_t} \left[\Delta SPV_i(t) - \Delta SPV_j(t) \right]^2 . \tag{4.2}$$

As in Scheidt and Caers (2007), we conduct a principal component analysis with the connectivity matrix $\mathbf{D} = -1/2\,\delta_{ij}^2$ after centering and take the first few principal components to visualize the model separation in two dimensional or three dimensional Euclidean spaces. The algorithm of MDS is described more in detail in the APENDIX B. Figure 4.39 shows the model separations using the first two and three principal components.



(a) 2 dimensions plot **(b)3 dimensions plot**

Figure 4.39 Multi-dimensional scaling of the initial ensemble models and updated ensemble models from Plain EnKF and Hybrid-Corse EnKF; (a) 2 dimensions plot, (b) 3 dimensions plot.

These separations are shown with respect to the ensemble members themselves and also the true model. Clearly, the ensemble members from the hybrid approach are clustered closer to the reference model in terms of their flow response. Also, there is sufficient separation between the ensemble members in the hybrid approach, indicating that there is no evidence of ensemble collapse.

4.5.3 Quantitative Comparison and Validation of Hybrid EnKF

Zupanski (2004) employed four validation measures to examine the performance of the EnKF. In this section, we apply the measures described in that paper to compare the hybrid EnKF with the conventional EnKF implementation. The first and most commonly used validation measure is the root mean square error (RMS) which can be computed as deviations from the reference model given by,

$$rms = \sqrt{\frac{1}{Ne}\sum_{j}^{Ne}\frac{1}{N}\sum_{i}^{N}(m_{i,j} - m_{true})^{2}} \tag{4.3}$$

Figure. 4.40 shows the results for the conventional and hybrid EnKF approaches at different times during the assimilation. The RMS error shows that the conventional EnKF exhibits divergent behavior with increasing RMS error. This is mitigated through localization and the RMS error now monotonically decreases through the assimilation. However, the hybrid EnKF shows a large reduction in the RMS error, especially at assimilation step 7 (1400 days) and 11 (2200 days) when we incorporate the inversion

results and impose the coarse scale permeability constraints. Also, at the end of the assimilation, the hybrid EnKF RMS error is smaller than that of the conventional EnKF with localization.

Another common measure of performance is the error covariance estimate from the ensemble. The posterior error covariance estimate can be obtained from the EnKF after updating at each assimilation step by,

$$\mathbf{C}_M^u = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} \left[m_i^u - \overline{m}^u \right]^T \left[m_i^u - \overline{m}^u \right] \tag{4.4}$$

The diagonal elements of this matrix indicate the deviations of the individual grid block permeabilities from the mean. A plot of the square-root of the diagonal elements of the covariance matrix is shown on a 50x50 grid in Figure. 4.41. The reduction in the deviations from the mean is considerable at 2000 days for the hybrid formulation and continues till the end of the assimilation period. This behavior reinforces the results from the RMS error shown in Figure. 4.40.

The $\chi 2$ validation diagnostics evaluates the correctness of the innovation (observation minus forecast) covariance matrix that employs a predefined observation error covariance and forecast error covariance in EnKF. The $\chi 2$ is defined in the observation space, normalized by the number of observation, N_{obs}.

$$\chi_k^2 = \frac{1}{N_{obs}} [\mathbf{D}_{obs,k} - \mathbf{H}\boldsymbol{\Psi}_k^p]^{\mathbf{T}} \left(\mathbf{H}\mathbf{C}_{\Psi}^p \mathbf{H}^{\mathbf{T}} + \mathbf{C}_{\mathbf{D}} \right)^{-1} [\mathbf{D}_{obs,k} - \mathbf{H}\boldsymbol{\Psi}_k^p] \tag{4.5}$$

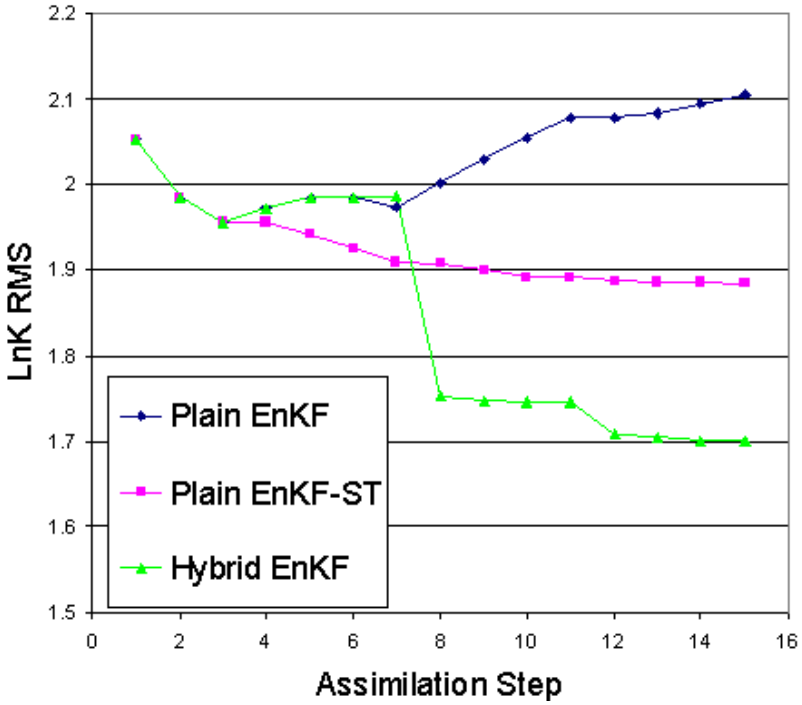


Figure 4.40 RMSE comparisons; Plain EnKF is Plain EnKF without localization, Plain EnKF SL is the plain EnKF with localization and Hybrid EnKF is Hybrid EnKF without localization.

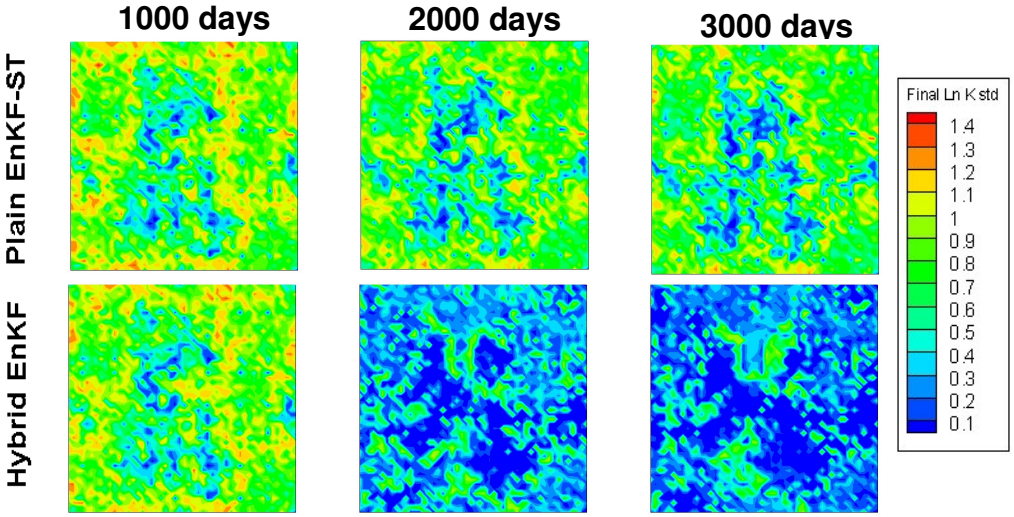


Figure 4.41 Error covariance estimation evolutions; Plain EnKF SL is the plain EnKF with localization and Hybrid EnKF is Hybrid EnKF without localization.

A characteristic of χ^2 is that for Gaussian distribution of innovations and linear observation operator **H**, χ^2 should be equal to 1 if the assimilation is optimal. However, for nonlinear models and a statistically small sample, one can expect only values of χ^2 to be close to 1 and not necessarily equal to 1. Figure 4.42 shows that the hybrid EnKF assimilation quality is better than that of the conventional EnKF and χ^2 for the hybrid EnKF is relatively stable around the value of 1.5.

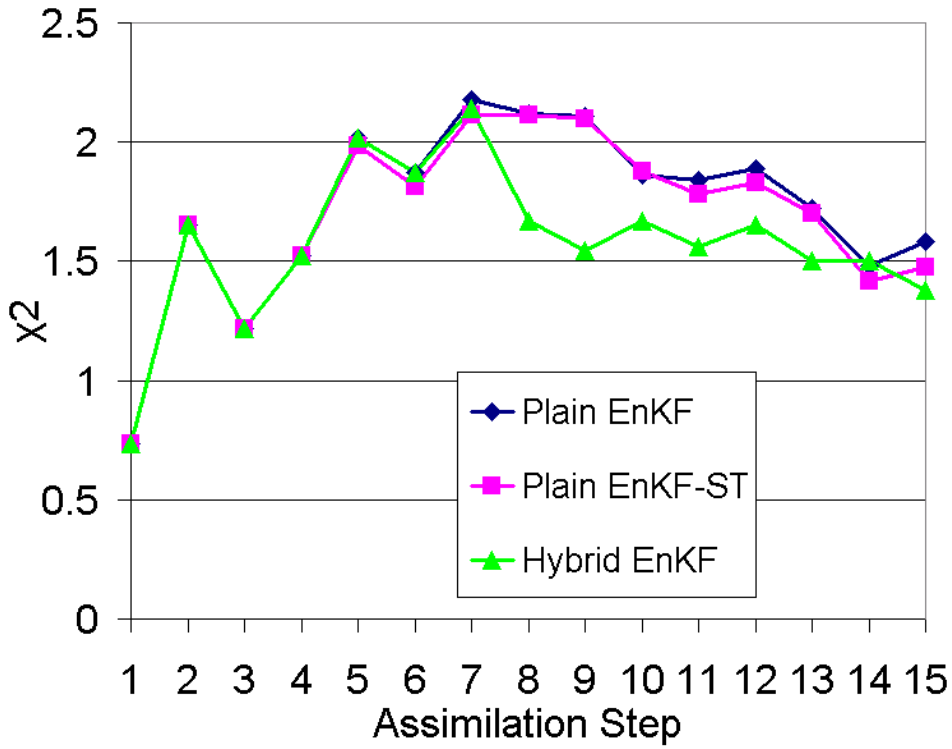
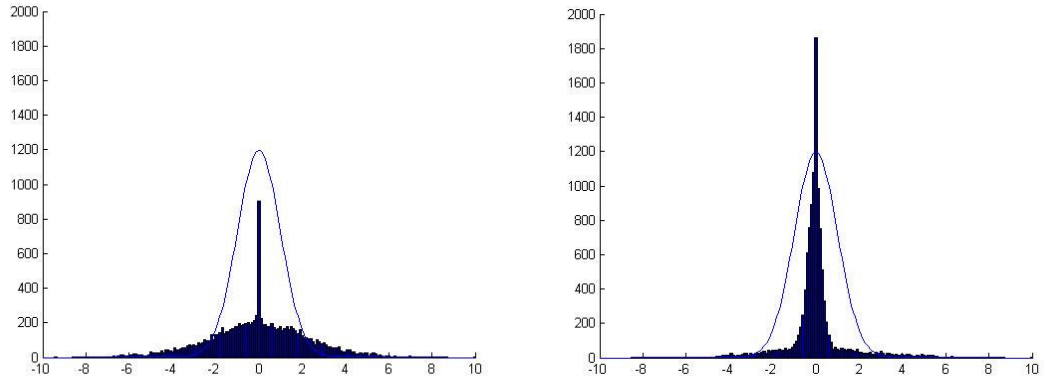


Figure 4.42 χ^2 validation test comparison; the conventional EnKF without streamline trajectory localization, the conventional EnKF with streamline trajectory localization, and the hybrid EnKF

Another approach for the statistical verification of an ensemble-based data assimilation algorithm is the probability density function (PDF) of the innovations. By taking the square root of χ^2 is defined in Eq. (4.5), we have:

$$Innov = \frac{1}{\sqrt{N_{obs}}}\left(\mathbf{H}\mathbf{C}_{\Psi}^{p}\mathbf{H}^{\mathrm{T}} + \mathbf{C}_{\mathbf{D}}\right)^{-1/2}[\mathbf{D}_{obs} - \mathbf{H}\boldsymbol{\Psi}^{p}] \tag{4.6}$$

For a linear dynamic system and observation operators, the PDF of the innovations is expected to be normally distributed with mean 0 and variance 1. However because of the nonlinearity of our problem and the relatively small ensemble size, only an approximate normal distribution can be expected. Figure 4.43 shows that the hybrid EnKF innovation PDF is closer to a standard normal distribution compared to the conventional EnKF with localization.



(a) PlainEnKF-ST **(b) Hybrid EnKF**

Figure 4.43 Innovation distribution comparison; (a) Plain EnKF with streamline trajectory localization, (b) Hybrid EnKF without localization

4.6 Gold Smith Field Application

We now demonstrate the applicability of the proposed hybrid EnKF for the same field scale example as Chapter III sec.3.4. Here is the list of the parameter for this study.

EnKF Parameters

- 9 producers for WWCT history mach
- 58×53×10 grids
- The number of the ensemble is 50
- Assimilation WWCT observation Data from 0 days to 3840 days for 11 times
- WWCT measurement error is 10%
- After assimilation, run all the ensemble from time 0 to 7800 days
- State variables { lnK, P, Sw, WWCT}for 1 st assimilation step
- State variables { lnK, P, Sw, lnK} for 2 nd assimilation step
- Normal Scare transformation for permeability
- No Localization for Hybrid EnKF
- Upscaling from 58 x 53 x 10 to 14 x 13x 5
- Coarse scale data error std lnk=1
- Coarse scale permeability constraint assimilation for 3 times.

The 50 initial model WWCT responses are shown in Figure 4.44. Because of the multimodal nature of the histograms, we use the normal score transforms of the permeability as a part of the EnKF state vector instead of the grid block permeability. A total of 11 assimilation steps over a period of 3840 days are used to integrate WWCT data and calibrate the underlying permeability fields. The measurement error is assumed to be 10% of the WWCT values. We used streamline-based covariance localization for the conventional EnKF and no covariance localization is applied for the hybrid EnKF.

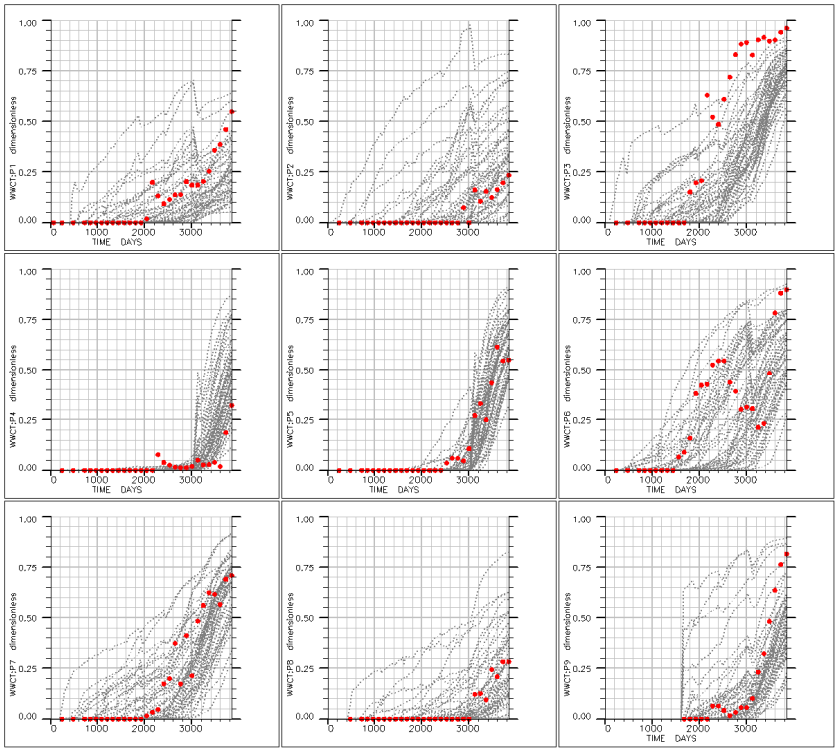
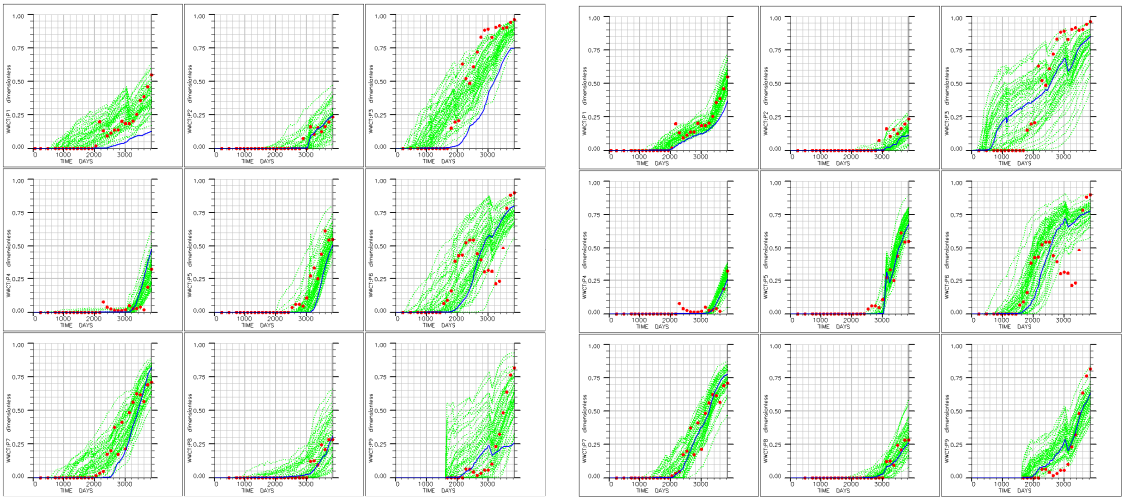


Figure 4.44 Initial 50 ensemble model WWCT responses; the observation data in red points and the initial ensemble model responses are in grey lines; in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5,and P6, and in the third row, from left, P7, P8 and P9.

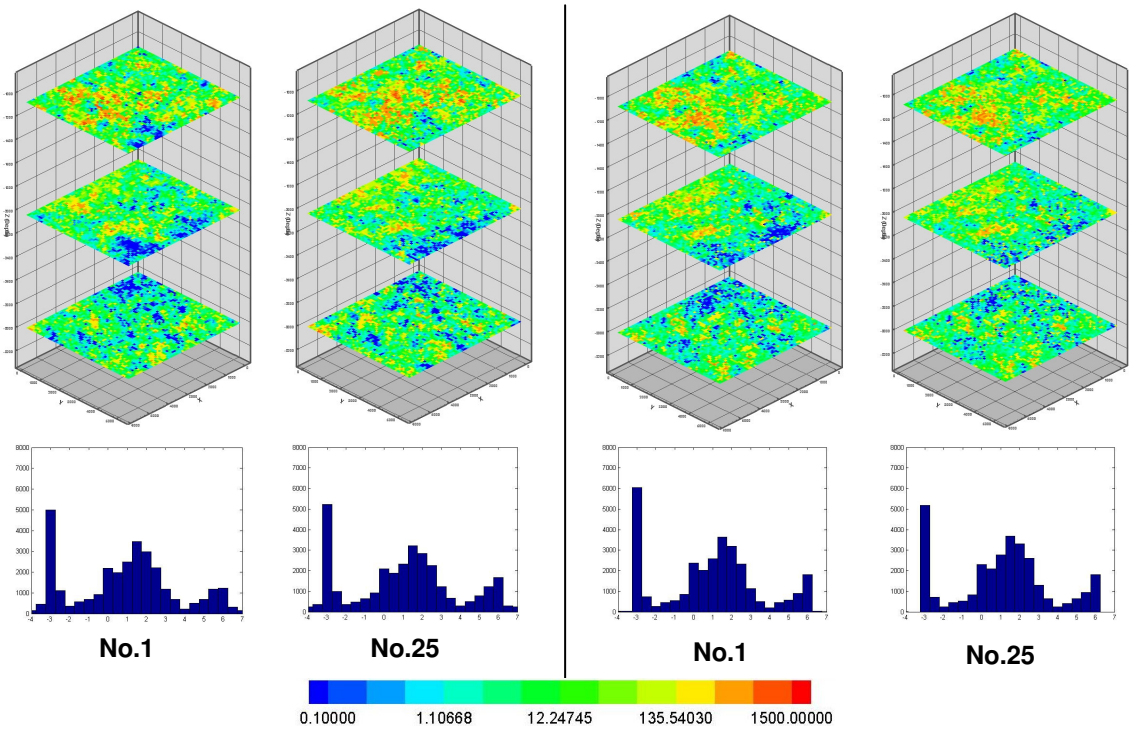
For the hybrid EnKF approach, the simulation grid is upscaled from 58x53x10 to a 14x13x5 coarse-scale model by flow based upscaling method. The standard deviation of the logarithm of the coarse-scale permeability is set to be 1. The coarse-scale constraint derived from the non-linear inversion is applied at 2400, 3030 and 3840 days for the hybrid EnKF. The history matching results are shown in Figure 4.45. The hybrid EnKF clearly outperforms the conventional EnKF in terms of the quality of the match to the WWCT data.



(a) Plain-ST EnKF **(b) Hybrid EnKF**

Figure 4.45 EnKF final 50 updated ensemble model WWCT responses comparisons between (a) Plain ENKF with localization and (b) Hybrid EnKF; the observation data in red points and the ensemble model responses in light green lines and the response from the updated mean ensemble model in blue line, in the first row from the left, P1, P2 and P3, in the second row from left, P4, P5,and P6, and in the third row, from left, P7, P8 and P9. Vertical line shows the last assimilation time step.

Updated permeability field comparisons are shown in Figure 4.46. Although streamline-based covariance localization minimizes spurious correlations and thus, reduces overshoots in the updated permeability field for the conventional EnKF, we still do see some areas of localized patches of low and high permeabilities. The hybrid EnKF not only results in a better match to the data but also preserves geologic continuity in the updated models.



(a) Plain EnKF-ST **(b) Hybrid EnKF**

Figure 4.46 Updated ensemble permeability fields comparison of (a) Plain EnKF with localization, from the left, ensemble No. 1, No. 25 and (b) Hybrid EnKF, from the left, ensemble No. 1, No. 25 at the depth of 1080 ft, 2042 ft and 3004 ft from the top.

4.7 Summary of Chapter IV

Hybrid EnKF applications are presented, and its results are investigated. The quality of the performance in various synthetic case is examined and uncertainty quantification analysis is conducted by several validation methods. The summary of this study is as follows.

- Swapping mean and mode estimation hybrid approach can improve the hybrid EnKF estimation.

- The upscaling factor and the coarse scale data error variance are sensitive to coarse scale permeability constraint. It needs to be adjusted based on the problem.

- Inversion results are crucial to achieve the benefit of the Hybrid EnKF approach, especially WWCT and WBHP simultaneous inversion is necessary to generate the plausible coarse permeability data to impose on the EnKF updating.

- The need for localization is significantly reduced using the hybrid approach.
- Swept volume changes can capture the spatial model responses appropriately. The multidimensional scaling method can be applied to visualize the variability of the updated ensemble model in space

- A synthetic example is used to shows the advantages of the Hybrid EnKF where the initial ensemble members do not span the solution. Whereas the conventional EnKF is unable to reproduce the spatial continuity based on the production data, the hybrid EnKF performs much better in terms of reproducing the permeability distribution.

- Four validation tests were conducted for examining the performance of the Hybrid EnKF. The results show the parameter was resolved better than the Plain EnKF in terms of rms error and error covariance estimation. Also, the χ^2 validation test and innovation PDF statistics show the model performance is more stable for the Hybrid EnKF.

- The practical feasibility of the hybrid EnKF is illustrated by using a field example from the Goldsmith field in West Texas. Compared to the conventional EnKF, the hybrid EnKF not only results in a better match to the data but also preserves geologic continuity in the updated models.

CHAPTER V

CONCLUDING REMARKS AND RECOMMENDATIONS

5.1 Conclusions

A hybrid Ensemble Kalman Filter (EnKF) formulation and its applications for reservoir characterization are proposed, and its performance in history matching is investigated. The applicability of the approach is demonstrated through synthetic examples and a field scale model. Based on the results from this study, the following conclusions can be made.

Characteristics of EnKF

- Limiting the updating values constrains the parameter estimation solutions in an ill-posed inverse problem.

- Values above 10 % of measurement error variance maintain the relative contribution of the data to the prior model information to avoid the overconfidence on the data and prevent ensemble model collapse after the assimilation.

- Streamline-based localization assists updating model parameters in terms of mitigating overshoot and undershoot problems and maintains minimal model change and localizes it in the influential zone of the flow dynamics.

- Normal Score transformation assists preserving the prior non-Gaussian model parameter distribution through the updating.

Characteristics of Hybrid EnKF

- We have proposed a hybrid EnKF that couples the conventional EnKF with non-linear inversion to account for non-linearity in the multiphase history matching problems and also the non-Gaussian property distributions in the geologic models. Specifically, we update the ensemble mean in a conventional EnKF through a non-linear inversion at selected time intervals and replace the ensemble mean with the 'posterior mode' from the inversion. This explicitly recognizes the fact that for non-Gaussian distributions, the posterior mode is a better representation of the central tendency compared to the ensemble mean.

- Our approach ensures that the ensemble members in the conventional EnKF follow the trajectory of the non-linear inversion within a specified degree of tolerance. This is accomplished by imposing the inversion results on each ensemble member via a coarse-scale constraint using a sequential second stage updating in the conventional

EnKF and a flow-based upscaling. This not only allows us to account for non-linearity in the model updates but also prevents filter divergence arising from the use of limited ensemble size.

- The upscaling factor and the coarse scale data error variance is sensitive to our approach and needs to be adjusted based on the problem specification

- We have illustrated the advantages of the Hybrid EnKF using a synthetic example where the initial ensemble members do not span the solution. Whereas the conventional EnKF is unable to reproduce the spatial continuity based on the production data, the hybrid EnKF performs much better in terms of reproducing the permeability distribution and also the underlying saturation front movements. Visualization of the ensemble members using multidimensional scaling and the first few principal components also shows that the hybrid approach is better able to reproduce flow field of the reference model.

- Inversion results is crucial to achieve the benefit of the Hybrid EnKF approach, especially WWCT and WBHP simultaneous inversion is necessary to generate the plausible coarse permeability data to impose on the EnKF updating.

- The need for localization is significantly reduced using the hybrid approach.

- Because the hybrid EnKF requires non-linear inversion only on the ensemble mean, it is well-suited for large-scale field applications compared to other forms of iterative EnKF that require non-linear inversion for each ensemble member. We have illustrated the practical feasibility of the hybrid EnKF using a field example from the Goldsmith field in West Texas. Compared to the conventional EnKF, the hybrid EnKF not only results in a better match to the data but also preserves geologic continuity in the updated models.

Uncertainty quantification

- The multidimensional scaling method visualizes the variability of the updated ensemble model in a space by using swept volume changes as the spatial model response.

- Four validation tests were conducted for examining the performance of EnKF. These tests quantify the performance of the sequential model updating and detect the transition of the updating behavior.

- We have presented a variety of diagnostics to compare the performance of the conventional and the hybrid EnKF. In all of the cases, the hybrid EnKF shows improved performance compared to the conventional EnKF.

5.2 Recommendations

For EnKF work flow

- Initial model selection sensitivity for the quality of the EnKF updating needs to be addressed.

- Sensitivity of the assimilation quality needs to be clarified with respect to the frequency of the assimilation step and the combination of the different type of the observation data.

- Parallel processing for EnKF facilitates the speed of processing and reduces the total computation in the entire work scheme.

- A criteria to define the observation error needs to be developed especially for field applications where the reservoir development activity is frequently changing.

- Implementation of the algorithm needs to be optimized in terms of the calculating Kalman Gain matrix such as direct solver for the inverting matrix calculation

- Localization technique needs to be tested in terms of different quantity of dynamic responses such as pressure responses in addition to the WWCT responses to localize the state and parameter variables effectively in a more rigorous way.

- For non-Gaussian parameters prior distributions, we need to come up with a way to decompose it into combination of Gaussian distributions and apply EnKF for the individual Gaussian distribution.

- Fundamentals of the combined parameter and state estimation problem need to be investigated in terms of the material valance error of both parameter and state updating in the reservoir problem.

For Hybrid EnKF work flow

- Inversion scheme is possible to be conducted in the coarse scale model rather than in the fine scale model to save computation time. However, this will require reparameterization techniques such as Discrete Cosine Transformation (DCT) to reduce the number of parameters for reducing the ill-poseness of the inverse problem

- Inversion algorithm can be converted from the deterministic formulation to the Bayesian formulation such as iterative Gauss-Newton method to be consistent with the Kalman filter equation.

- For a field application, in coarse scale permeability constraint approach, the upscaling method needs to be more general in terms of the geometry of the discretized domain such as corner point geometry.

- The combination of the proposed hybrid approach and streamline based localization needs to be investigated. For example, instead of using the influential zone from the every ensemble model by the streamline trajectory, we can use the ensemble mode model from the inversion to define the influential zone to localize the cross covariance estimation.

REFERENCES

Akella, S., Efendiev, Y. and Datta-Gupta, A. 2008. A Coarse-Scale Constrained Ensemble Kalman Filter for Subsurface Characterization. Paper submitted to *Advances in Water Resources* for review.

Anderson, J.L. and Anderson, S.L. 1999. A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. *Monthly Weather Review* **127**:2741-2758.

Cheng, H., He, Z. and Datta-Gupta, A. 2005. A Comparison of Travel-Time and Amplitude Matching for Field-Scale Production Data Integration: Sensitivity, Nonlinearity and Practical Implications Filter. *SPE Journal* **10** (1): 75-90.

Datta-Gupta, A. and King, M. 2007. *Streamline Simulation: Theory and Practice, Society of Petroleum Engineers Textbook Series vol.11*. SPE, Richardson, TX.

Deutsch, C.V. and Journel, A., 1992. *GSLIB Geostatistical Software Library and User's Guide*, Oxford University Press, New York.

Devegowda, D., Arroyo, E., Datta-Gupta, A. and Douma, S.G. 2007. Efficient and Robust Reservoir Model Updating Using Ensemble Kalman Filter with Sensitivity Based Covariance Localization. Paper SPE 106144 presented at the SPE Reservoir Simulation Symposium, The Woodlands, Texas, 24-26 February.

Devegowda, D., Akella, S., Datta-Gupta, A. and Efendiev, Y. 2009 Interpretation of Partitioning Interwell Tracer Test Using EnKF with Coarse-Scale Constraints. Paper SPE 119125 presented at the SPE Reservoir Simulation Symposium, The Woodlands, Texas, 2-4 February.

Efendiev, Y., Datta-Gupta, A., Osako, I. and Mallick, B. 2005. Multi-scale Data Integration Using Coarse-Scale Models. *Advances in Water Resources* **28**:303-314.

Elkin-Arroyo, Deepak Devegowda, and Akhil Datta-Gupta. 2006. Streamline Assisted Ensemble Kalman Filter for Rapid and Continuous Reservoir Model Updating. Paper SPE 104255 presented at the 2006 SPE International Oil and Gas Conference, Beijing, China, 5-7 December.

Evensen, G. 1994. Sequential Data Assimilation with a Nonlinear Quasi-geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics. *Journal of Geophysical Research* **99** (C5): 10143-10162.

Evensen, G. 2003. The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation. *Ocean Dynamics*, **53**: 343-367.

Evensen G. 2004. Sampling strategies and square root analysis schemes for EnKF. *Ocean Dynamics*, **54**: 539-560.

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

Faruk O. A. ,Mark D. B, and Caers, J. 2008.A flow-based pattern recognition algorithm for rapid quantification of geologic uncertainty. Computational Geosciences Manuscript, accepted for publication

Gaspari, G. and Cohn, S.E. 1999. Construction of Correlation Functions in Two and Three Dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**: 723-757.

Gu, Y. and Oliver, D. S. 2004. History Matching of the PUNQ-S3 Reservoir Model Using the Ensemble Kalman Filter. Paper SPE 89942 presented at the 2004 Annual Technical Conference and Exhibition, Houston 26-29 September.

Gu, Y. and Oliver, D. S. 2006. The Ensemble Kalman Filter for Continuous Updating of Reservoir Simulation Models. *Journal of Energy Resources Technology* **128**: 79-87.

Gu, Y. and Oliver, D. S. 2007. An Iterative Ensemble Kalman Filter for Multiphase Fluid Flow Data Assimilation, *SPE Journal*, **12**(4): 438-446.

Hamill, T.M. and Whitaker, J.S. 2001. Distance-Dependent Filtering of Background Error Covariance Estimate in an Ensemble Kalman Filter. *Monthly Weather Review* **129**: 2776-2790.

He, Z. and Datta-Gupta, A. 2002. Streamline-based Production Data Integration with Gravity and Changing Field Conditions. *SPE Journal* **7** (4): 423-436.

Houtekamer, P.L. and Mitchell, H.L. 2001. A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. J. *Monthly Weather Review* **129**: 123-137.

Lee S.H., Malallah A, Datta-Gupta A, and Higdon D. 2002. Multiscale Data Integration Using Markov Random Fields. *SPE Reservoir Evaluation Engineering* **5** (1): 68-78.

Li, G. and Reynolds, A.C. 2007. An Iterative Ensemble Kalman Filter for Data Assimilation. Paper SPE 109808 presented at the SPE Annual Technical Conference and Exhibition, Anaheim, California, 11-14 November.

Nævdal, G., Johnsen, L.M., Aanonsen, S.I. and Vefring, E.H. 2003. Reservoir Monitoring and Continuous Model Updating Using Ensemble Kalman Filter. Paper SPE 84372 presented at the 2003 Annual Technical Conference and Exhibition, Denver, 8-5 October.

Skjervheim, J.A., Evensen, G., Aanonsens, S.I., Ruud, B.O. and Johnsen, T.A. 2005. Incorporating 4D Seismic Data in Reservoir Simulation Model Using Ensemble Kalman Filter. Paper 95789 presented at the 2005 Annual Technical Conference and Exhibition, Dallas, 9-12 October.

Tarantola, A. 2005. *Inverse Problem Theory and Method for Model Parameter Estimation*, SIAM, Philadelphia.

Vasco, D.W., Yoon, S., and Datta-Gupta, A. 1999. Integrating Dynamic Data into High-Resolution Reservoir Models Using Streamline-Based Analytic Sensitivity Coefficients. *SPE Journal* **4** (4): 389-399.

Vasco, D.W. and Karasaki, K. 2006. Interpretation and Inversion of Low-frequency Head Observations. *Water Resour.*, **42**: W05408.1- W05408.18.

Watanabe S., Datta-Gupta, A., Efendiev, E. and Devegowda, D. 2009. A Hybrid Ensemble Kalman Filter with Coarse Scale Constraint for Nonlinear Dynamics. Paper SPE-124826-MS presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, 4-7 October.

Wen, X.H. and Chen W.H., 2005. Real-Time Reservoir Model Updating Using Ensemble Kalman Filter. Paper SPE 92991 presented at the Reservoir Simulation Symposium. Houston, 31 January – 2 February.

Zafari, M. and Reynolds, A.C. 2005. Assessing the Uncertainty in Reservoir Description and Performance Prediction with the Ensemble Kalman Filter. Paper SPE 95750 presented at the 2005 Annual Technical Conference and Exhibition, Dallas, 9-12 October.

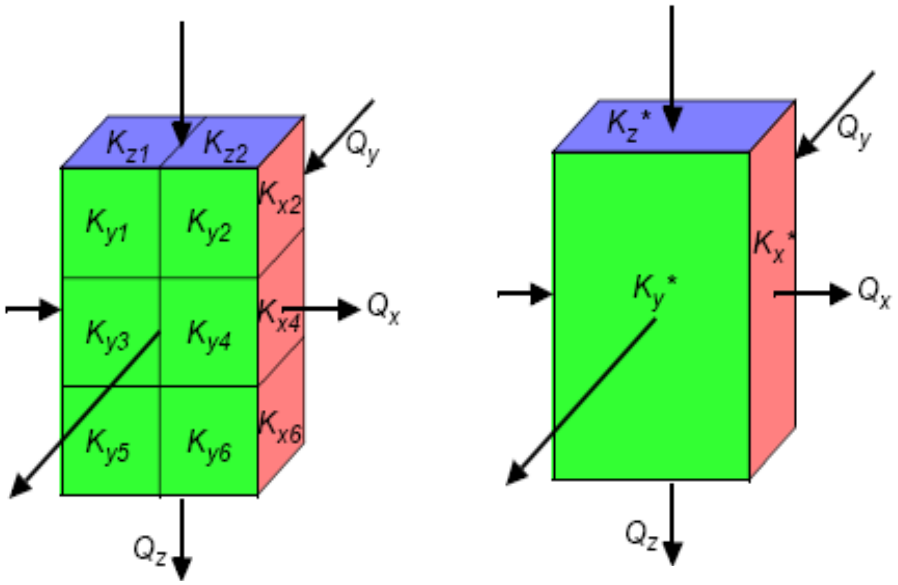
Zhang, F., Zhang, M. and Hansen, J.A. 2009. Coupling Ensemble Kalman Filter with Four Dimensional Variational Data Assimilation. *Advances in Atmospheric Sciences* **26**: 1-8.

Zupanski, M. 2005. Maximum Likelihood Ensemble Filter Theoretical Aspects, *Mon. Wea. Rev* **133**: 1710-1726

APPENDIX A

SINGLE PHASE UPSCALING

We will briefly describe single-phase flow upscaling procedure used in the coarse scale permeability constraint hybrid EnKF. This type of upscaling is discussed by many authors (Durlofsky et al. 1996; Efendiev et al. 2000). The main idea of this approach is to upscale the absolute permeability field k on the fine scale-grid (see Figure A.1), and then solve the original system on the coarse-grid with upscaled permeability field. Below, we will discuss briefly the upscaling of absolute permeability used in our simulations.



(a) Fine scale cells **(b) Coarse scale cell**

Figure A. 1 —Fine scale cells and coarse scale cell.

Consider the fine-scale permeability that is defined in the domain with underlying fine grid as shown in Figure A.1. On the same graph we illustrate a coarse-scale partition of the domain. To calculate the upscaled permeability field at the coarse-level, we use the solutions of local pressure equations. The main idea of the calculation of the coarse-scale permeability is that it delivers the same average fluxes as that of the underlying fine-scale problem locally. For each coarse domain *D*, we solve the local problems

$$div(k(x)\nabla\phi_j) = 0, \tag{A.1}$$

with some coarse-scale boundary conditions. Here $k(x)$ denotes the fine-scale permeability field. We will use the boundary conditions which are given by $\phi_j = 1$ and $\phi_j = 0$ on the opposite sides along the direction e_j and no flow boundary conditions on all other sides. For these boundary conditions, the coarse-scale permeability tensor is given by

$$(k^*(x)e_j, e_l) = \frac{1}{|D|}\int_D \left(k(x)\nabla\phi_j(x), e_l\right)dx \tag{A.2}$$

where ϕ_j is the solution of Eq. A.1 with prescribed boundary conditions. Various boundary conditions can have some influence on the accuracy of the calculations, including periodic, Dirichlet and etc.

APPENDIX B

MULTI-DIMENSIONAL SCALING (MDS)

OF THE SWEPT VOLUME CHANGES

An uncertainty analysis can be conducted by using Multi-Dimensional Scaling. We apply the swept volume differences through time as a dissimilarity measurement and visualize the each model responses as a point in the space to see the variability of the updated ensemble model. Define the swept volume change by threshold time of flight values of the streamline simulation for the updated ensemble model,

$$\Delta SPV(t=\tau_1-\tau_2)=\left|SPV\left(t=\tau_1\right)-SPV\left(t=\tau_2\right)\right|, \tag{B.1}$$

and the connectivity distance between the swept volume changes of two individual realization i and j is defined by

$$\delta_{ij}=\left(\sum_{t=1}^{N}\left[\Delta SPV_i(t)-\Delta SPV_j(t)\right]^2\right) \tag{B.2}$$

where N is the total number of time step that we want to compute the swept volume changes. Then construct a matrix containing connectivity distances $\mathbf{D}$,

$$\mathbf{D}=-\frac{1}{2}\delta_{ij}^2 \tag{B.3}$$

We center the above matrix $\mathbf{D}$ by

$$\mathbf{C} = \mathbf{JDJ} \tag{B.4}$$

with $\mathbf{J} = \mathbf{I} - \frac{1}{N_e}\mathbf{1}^T\mathbf{1}$ where $\mathbf{1} = [11111.....1]_{1\times N_e}$ where $\mathbf{I}$ is the identity matrix of dimension N_e which is the ensemble size. Eigenvalue decomposition of $\mathbf{C}$ yields

$$\mathbf{C} = \mathbf{V}\ \mathbf{\Lambda}\ \mathbf{V}^T, \tag{B.5}$$

where $\mathbf{V}$ is the matrix of eigenvectors and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues. One can construct a vector $\mathbf{X}$ in any dimension from a minimum of one dimension up to a maximum of N_e dimension which meets

$$\mathbf{C} = \mathbf{X}\mathbf{X}^T = \mathbf{V}\ \mathbf{\Lambda}\ \mathbf{V}^T \Rightarrow \mathbf{X} = \mathbf{V}\mathbf{\Lambda}^{1/2}\ . \tag{B.6}$$

If we work with q largest eigenvalues, we can construct a lower dimensional subspace of a geometric space $\boldsymbol{E}$ with

$$\mathbf{X}_q = \mathbf{V}_q\mathbf{\Lambda}_q^{1/2} \tag{B.7}$$

$\mathbf{V}_q$ is the matrix containing the eigenvectors that belong to the q largest eigenvalues in the diagonal matrix $\mathbf{\Lambda}_q$.

VITA

Name:	Shingo Watanabe
Address:	Dept. of Petroleum Engineering c/o Dr. Datta-Gupta Texas A&M University College Station, TX 77843-3116
Email Address:	shingo.watanabe@pe.tamu.edu
Education:	B.E, Environmental Resources Engineering, Waseda University,2007 Tokyo, Japan M.S, Petroleum Engineering, Texas A&M University,2009