Copyright

by

Tiancong Hong

2008

The Dissertation Committee for Tiancong Hong Certifies that this is the approved version of the following dissertation:

MCMC algorithm, integrated 4D seismic reservoir characterization and uncertainty analysis in a Bayesian framework

Committee:

Mrinal K. Sen, Supervisor

Paul L. Stoffa

Robert H. Tatham

Stephen P. Grand

Charles S. Jackson

MCMC algorithm, integrated 4D seismic reservoir characterization and uncertainty analysis in a Bayesian framework

by

Tiancong Hong, B.S.; B.S.; M.S.; M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas at Austin

August 2008

To my parents

Chengyun Hong and Zhenying Lu

To my wife

Jianxin Liu

To my daughter

Emma (Feilan) Hong

Acknowledgements

First and foremost, I would like to express my gratitude to my advisor, Dr. Mrinal K. Sen, for his guidance and wisdom throughout my Ph.D. study and research. Mrinal brought me into a brand new research field and helped me work out an exciting Ph.D. research subject. He has always been around for encouragement, assistance and inspiration in many aspects during the past three years. Without his constant support, this dissertation would never have shaped. I would also like to thank Dr. Paul Stoffa, Dr. Robert Tatham, Dr. Steve Grand and Dr. Charles Jackson for serving on my committee. Their insightful comments provided valuable inputs to my dissertation.

I sincerely thank Dr. Long Jin for valuable discussions that benefited my research. I extend my warmest gratitude to my friends and colleagues at UT Austin: Jonas De Basabe, Seif Roustam, Yi Deng, Chaoshu Hu, Chunlei Chu, Russell Young, Sanjay Sood, Rishi Bansal, Chandan Kumar, Samik Sil, Will Burnett, Erick Leuro, Shaoping Lu, Jitao Ma, Abdulaziz Almukaidib and Achintya Pal.

I also sincerely thank Mark Wiederspahn, Kevin Johnson and John Gerboc for their technical support; Phillip Guerrero, Susan Beaubien, Don Yarbro, Judy Sansom, Jan Everett, Lisa Gahagan and Patricia E Ganey-Curry for their administrative assistance.

Many thanks also go to the donors of multiple financial support sources without which my Ph.D. study and research would have been impossible. They are: National Science Foundation grants of OCE 0415738 and CNS 0427005, UT Jackson School of

Geosciences, EDGER forum, SEG Scholarship, Chevron Scholarship, BP Fellowship and ConocoPhillips Fellowship.

Finally, I would like to thank my family for standing behind me all the time: My parents, Chengyun Hong and Zhenying Lu. I wouldn’t have come to this stage of my life without their constant love and support. My wife, Jianxin Liu, for her endless love and tolerance. My daughter, Emma (Feilan) Hong, for making my life different.

MCMC algorithm, integrated 4D seismic reservoir characterization and uncertainty analysis in a Bayesian framework

Publication No._____

Tiancong Hong, Ph.D.

The University of Texas at Austin, 2008

Supervisor: Mrinal K. Sen

One of the important goals in petroleum exploration and production is to make quantitative estimates of a reservoir's properties from all available but indirectly related surface data, which constitutes an inverse problem. Due to the inherent non-uniqueness of most inverse procedures, a deterministic solution may be impossible, and it makes more sense to formulate the inverse problem in a statistical Bayesian framework and to fully solve it by constructing the Posterior Probability Density (PPD) function using Markov Chain Monte Carlo (MCMC) algorithms. The derived PPD is the complete solution of an inverse problem and describes all the consistent models for the given data. Therefore, the estimated PPD not only leads to the most likely model or solution but also provides a theoretically correct way to quantify corresponding uncertainty. However, for many realistic applications, MCMC can be computationally expensive due to the strong nonlinearity and high dimensionality of the problem. In this research, to address the fundamental issues of efficiency and accuracy in parameter estimation and uncertainty

quantification, I have incorporated some new developments and designed a new multi-scale MCMC algorithm. The new algorithm is justified using an analytical example, and its performance is evaluated using a nonlinear pre-stack seismic waveform inversion application. I also find that the new technique of multi-scaling is particularly attractive in addressing model parameterization issues especially for the seismic waveform inversion.

To derive an accurate reservoir model and therefore to obtain a reliable reservoir performance prediction with as little uncertainty as possible, I propose a workflow to integrate 4D seismic and well production data in a Bayesian framework. This challenging 4D seismic history matching problem is solved using the new multi-scale MCMC algorithm for reasonably accurate reservoir characterization and uncertainty analysis within an acceptable time period. To take advantage of the benefits from both the fine scale and the coarse scale, a 3D reservoir model is parameterized into two different scales. It is demonstrated that the coarse-scale model works like a regularization operator to make the derived fine-scale reservoir model smooth and more realistic. The derived best-fitting static petrophysical model is further used to image the evolution of a reservoir's dynamic features such as pore pressure and fluid saturation, which provide a direct indication of the internal dynamic fluid flow.

Table of Contents

Acknowledgements v

Abstract vii

List of Tables xii

List of Figures xiii

Chapter 1. Introduction 1

1.1 Inverse problem in a Bayesian framework 1
1.2 Markov Chain Monte Carlo algorithms 5
1.3 Integrated 4D seismic reservoir characterization.....10
1.4 Uncertainty quantification 15
1.5 Objectives and dissertation organization 18

Chapter 2. Multi-scale real-coded hybrid Genetic Algorithm 21

2.1 Basic Genetic Algorithm.....24
2.2 Multi-scale real-coded hybrid Genetic Algorithm.....27
2.2.1 Combining elements of SA into a GA 27
2.2.2 Real-coded GA 29
2.2.3 Multi-scale real-coded hybrid GA 33
2.3 Example: pre-stack seismic waveform inversion 37

Chapter 3. Multi-scale GA based MCMC algorithm 55

3.1 Multi-scale GA based MCMC.....57
3.1.1 Multi-scale coupled MCMC 57

3.1.2 Crossover swapping 59
3.2 Analytical example 62
3.3 Example: Seismic parameter estimation and uncertainty analysis 65

Chapter 4. Integrated 4D seismic history matching 80

4.1 4D seismic.....82
4.1.1 Feasibility study 84
4.1.2 Repeatability study.....88
4.2 Post-stack seismic modeling and flow simulation.....90
4.2.1 Post-stack seismic modeling 91
4.2.2 Flow simulation 94
4.3 Seismic rock physics.....98
4.4 Integrated 4D seismic history matching workflow.....104
4.4.1 History matching 105
4.4.2 Seismic history matching 108
4.5 2D example 110

Chapter 5. Integrated 4D seismic reservoir characterization and uncertainty analysis in a Bayesian framework 123

5.1 Formulation in a Bayesian framework.....125
5.1.1 Likelihood functions 126
5.1.2 Prior model.....128
5.1.3 Posterior model 129
5.2 MCMC sampling from the PPD 131
5.3 Example: Bayesian inference of reservoir properties and uncertainty analysis.....133
5.3.1 Reference model and model parameterization 135
5.3.2 Prior model and likelihood functions 138
5.3.3 Static petrophysical model and uncertainty quantification 140
5.3.4 Dynamic pore pressure and fluid saturation 145

Chapter 6. Conclusions and future directions 172

6.1 Conclusions....172

6.1.1 MCMC algorithm 172

6.1.2 Integrated seismic history matching in a Bayesian framework 173

6.2 Future directions 175

6.2.1 MCMC algorithm 175

6.2.2 Integrated seismic history matching in a Bayesian framework ... .176

Appendix A. Metropolis-Hastings algorithm 179

Bibliography 181

Vita 191

List of Tables

Table 4.1: Critical factors in the physical feasibility of time-lapse seismic reservoir monitoring (Wang, 1997) 86
Table 4.2: Cases of possible high contrast in pore fluid compressibility (Wang, 1997) 87
Table 4.3: Good candidate reservoirs for time-lapse seismic reservoir monitoring (Wang, 1997) 87
Table 4.4: Ideal parameters of time-lapse seismic operations (Lumley et al., 1997) 88
Table 4.5: Rock and flow parameter values for the 2D seismic history matching test 112
Table 5.1: Rock and flow parameter values that are assumed to be constant throughout the production history 136

List of Figures

Figure 1.1: An inverse problem formulated in a Bayesian framework, which is solved by stochastically constructing the Posterior Probability Density (PPD) function 2

Figure 1.2: Schematic representation of the relationship among prior, likelihood function and posterior 4

Figure 2.1: Possible area of the range of the offspring variable for the given parent model variables. 44

Figure 2.2: Illustration of the mutation on a variable, which is equivalent to adding a white noise on the original variable. 44

Figure 2.3: Pseudo code of the multi-scale real-coded hybrid GA 45

Figure 2.4: 1D earth model and its parameterizations in 4 different scales. 46

Figure 2.5: Inverted optimal *Vp* structures using multi-scale GA for the 4 scales, which are compared against well logs. 47

Figure 2.6: Inverted optimal *Vs* structures using multi-scale GA for the 4 scales, which are compared against well logs. 48

Figure 2.7: Inverted optimal density structures using multi-scale GA for the 4 scales, which are compared against well logs. 49

Figure 2.8: Comparison between true and inverted seismograms for the 4 scales, which correspond to the final optimal structures. 50

Figure 2.9: Residues between the observed gather and the synthetic gathers for the four scales 51

Figure 2.10: Fitness histories produced by running the multi-scale GA. 52

Figure 2.11: Inverted optimal *Vp* structures by running the conventional single-scale GA separately on each of the 4 scales, which are compared against well logs. 53

Figure 2.12: Fitness histories produced by running the conventional single-scale GA separately on each of the 4 scales. 54

Figure 3.1: The joint PPD surface with two free parameters X_1 and X_2. 72

Figure 3.2: (a) The true joint posterior distribution projected on the (X_1, X_2) plane. (b) The contours of the true PPD on the (X_1, X_2) plane. (c) Marginal PPD for parameter X_1 with a bimodal distribution. (d) Marginal PPD for parameter X_2 with a univariate Gaussian distribution 73

Figure 3.3: 9000 samples drawn by the multi-scale GA based MCMC algorithm, which overlie the contours of the true PPD. 74

Figure 3.4: Marginal PPDs are estimated with the relative frequency distributions of the 9000 sample points, which are compared against the true marginal PPDs for (a) X_1 and (b) X_2 75

Figure 3.5: Samples of parameter V_p of the 29^{th} layer, which are drawn by the 7 MCMC chains at scale #3. 76

Figure 3.6: Inferred marginal PPDs and MAP models of *Vp* by the multi-scale GA-type MCMC for 4 scales. 77

Figure 3.7: CDFs and marginal PPDs of *Vp* of the grayed layers in Figure 3.6 for the 4 scales, which indicates both the most probable (highest frequency) values and the corresponding uncertainty bounds. 78
Figure 3.8: Parts of the posterior correlation matrices for scale #2 ranging from 7th to 10 the layer, and for scale #3 ranging from 13th to 20th layer.....79
Figure 4.1: Graphic description of the reflection and transmission phenomenon of an incident plane P-wave at a planar interface (*Z*=0) which separates two different elastic media.....117
Figure 4.2: The proposed integrated workflow to simultaneously history match time-lapse seismic data, well production data and sensor information 117
Figure 4.3: (a) True permeability model, (b) Inverted mean reservoir permeability model, and (c) Corresponding standard deviation. The well control strategy is also shown 118
Figure 4.4: Normalized error history through 35 iterations. 119
Figure 4.5: Comparison of observed and synthetic data fits of well cumulative production (STB) and oil-water ratio measurements.....119
Figure 4.6: Comparison of (a) observed and (b) synthetic time-lapse data fits for the reservoir cross-section at three different legacy times.....120
Figure 4.7: Distributions of fluid pressure (*kPa*) at three different monitor times… 121
Figure 4.8: Distributions of oil concentration at three different monitor times.121
Figure 4.9: Distributions of the resulting P-impedance.....122
Figure 5.1: Integrated workflow used to derive reservoir model parameters by simultaneously integrating 4D seismic and well production data. .148
Figure 5.2: 3D reservoir model used in the study, which is parameterized into two different scales of 20x20x5 and 10x10x5. 149
Figure 5.3: Reference porosity model that is assumed to be the actual reservoir model that is taken as the target to estimate. 150
Figure 5.4: Reference porosity distributions for three different layers that are extracted from the 3D reference model in Figure 5.3. 151
Figure 5.5: Observed (red) and synthetic (blue) well production data, that is, water (left) and oil (right) rates. 152
Figure 5.6: Observed time-lapse seismic data for a particular cross-section (*X*=10) after different time steps of 30, 300 and 700 *days*. 153
Figure 5.7: Synthetic time-lapse seismic data for a particular cross-section (*X*=10) after 30, 300 and 700 *days* respectively. 154
Figure 5.8: The distributions of P-impedance (acoustic) for the same cross-section and after the same different time periods as in Figure 5.7. 155
Figure 5.9: Fitness histories for both the fine-scale and the coarse-scale models… 156
Figure 5.10: The most likely fine-scale reservoir model that is derived based on the multi-scale MCMC samples defined on the scale of 20x20x5. 157
Figure 5.11: The most likely coarse-scale reservoir model that is derived based on the multi-scale MCMC samples defined on the scale of 10x10x5. 158

Figure 5.12: The derived porosity distributions using the multi-scale MCMC and for three different layers that are extracted from the derived fine-scale 3D reservoir porosity model in Figure 5.10.....159
Figure 5.13: The derived porosity distributions using the single-scale MCMC and for three of the five layers in the fine-scale model. 160
Figure 5.14: The derived porosity distributions using the multi-scale MCMC for three of the five layers but with well production data in use only. 161
Figure 5.15: The derived porosity distributions using the multi-scale MCMC for three of the five layers but with 4D seismic data in use only. 162
Figure 5.16: Calculated histograms of four particular grid blocks based on the multi-scale MCMC samples on the fine scale, which are used as the estimated marginal PPDs to determine the most likely porosity values and to quantify corresponding uncertainties.....163
Figure 5.17: Derived relative porosity deviation model of scale 20x20x5 and with respect to the most likely porosities, which is represented in a 3D cube and slices as well.....164
Figure 5.18: Derived relative porosity deviation for three different layers extracted from the 3D deviation cube in Figure 5.17. 165
Figure 5.19: Distributions of the dynamic pore pressure for a particular layer (Z=3) after 30, 300 and 700 *days*. 166
Figure 5.20: Distributions of the dynamic pore pressure for a particular cross-section (X=10) after 30, 300 and 700 *days*.....167
Figure 5.21: Differential decreases in pore pressure for a particular layer (Z=3) and a particular cross-section (X=10) respectively from (a) 100 to 200 *days*, and (b) 500 to 600 *days*.....168
Figure 5.22: Distributions of the dynamic water saturation for a particular layer (Z=3) after 30, 300 and 700 *days* 169
Figure 5.23: Distributions of the dynamic water saturation for a particular cross-section (X=10) after 30, 300 and 700 *days* 170
Figure 5.24: Differential increases in water saturation for a particular layer (Z=3) and a particular cross-section (X=10) respectively from (a) 100 to 200 *days*, and (b) 500 to 600 *days*.....171

Chapter 1

Introduction

1.1 Inverse problems in a Bayesian framework

A fundamental role of exploration and production geophysics is to make quantitative inference of a reservoir's interior physical properties from surface-recorded but indirectly related seismic data and other measurements. This task constitutes an inverse problem, or an inductive reasoning process in the sense of logic. Due to the inherent non-uniqueness, a deterministic solution of inverse problems may be impossible. In statistical terms, we are unable to apply deductive reasoning to prove or disprove models or values of model parameters.

In order to resolve the non-uniqueness issue, different strategies have been employed to exploit the model space for the optimal or best-fitting model (e.g., Sen and Stoffa, 1991; Stoffa and Sen, 1992; Sen and Stoffa, 1996; Sen and Roy, 2003). However, it is better to cast the inverse problem in a statistical Bayesian framework and to fully solve it by stochastically exploring the model space to construct the Posterior Probability Density (PPD) surface as illustrated in Figure 1.1 (Tarantola, 1987). In a Bayesian inference framework, the PPD describes all the consistent models together with uncertainties in light of the current state of knowledge, which allows direct calculation of the probability of any particular model or the value of any particular parameter. "Parameter estimation" and "model selection" are the two major Bayesian inference

problems. The former deals with problems in which a parameterized model is assumed to be true and the values of model parameters and their uncertainties are estimated; the latter compares competing models of different parameterizations by automatically penalizing complicated models using the built-in "Occam's razor" and hence determines the best model parameterization for the given data.

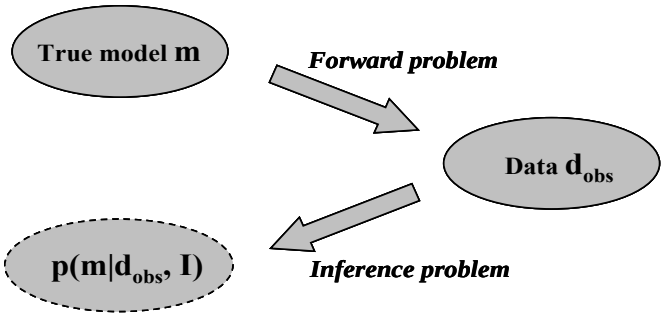

Figure 1.1: An inverse problem formulated in a Bayesian framework, which is solved by stochastically constructing the Posterior Probability Density (PPD) function.

The way to mathematically formulate an inverse problem in a Bayesian framework is based on the Bayes theorem. According to the Bayes theorem, the state of our knowledge about the target model is updated by combining the independent prior information with the available data to construct the PPD. The PPD describes all the consistent models as well as the associated likelihoods. For a parameter estimation problem, the PPD of the model vector can be expressed as

$$p(\mathbf{m} \mid \mathbf{d}_{\mathbf{obs}}, \mathbf{I}) = \frac{p(\mathbf{m} \mid \mathbf{I}) \cdot p(\mathbf{d}_{\mathbf{obs}} \mid \mathbf{m}, \mathbf{I})}{p(\mathbf{d}_{\mathbf{obs}} \mid \mathbf{I})} \quad , \qquad (1.1)$$

where $\mathbf{m}$, $\mathbf{d}_{\mathbf{obs}}$ and $\mathbf{I}$ represent model, data vectors and independent prior information respectively; the conditional probability density function (*pdf*) $p(\mathbf{m} \mid \mathbf{d}_{\mathbf{obs}}, \mathbf{I})$ is the target PPD representing the complete solution of an inverse problem. Prior knowledge about the model, $p(\mathbf{m} \mid \mathbf{I})$, summarizes all available information from sources independent of the current data, which is necessary to initiate a Bayesian inference. Likelihood function $p(\mathbf{d}_{\mathbf{obs}} \mid \mathbf{m}, \mathbf{I})$, often expressed as $L(\mathbf{d}_{\mathbf{obs}} \mid \mathbf{m}, \mathbf{I})$, relates to the forward process and represents the probability of obtaining the observations for a given model and the prior information. Denominator $p(\mathbf{d}_{\mathbf{obs}} \mid \mathbf{I})$, usually called marginal likelihood or evidence, is a normalization factor to guarantee that the sum of the PPD is unity and actually is an integral of the prior times the likelihood over the entire model space, as shown in equation (1.2). This integration poses a daunting challenge for high-dimensional inverse problems. In other words, it is impractical to compute this integration analytically due to the high-dimensionality, and as a result the PPD cannot be derived deterministically. This is why Markov Chain Monte Carlo (MCMC) sampling methods have been used extensively for PPD estimation while circumventing direct calculation of the denominator integration (Sen and Stoffa, 1996; Ulrych *et al.,* 2001). Figure 1.2 schematically shows the relationship between prior, likelihood and posterior. Basically, a broad, rough prior is refined by observed data to generate a more accurate posterior and to narrow the knowledge about the target model.

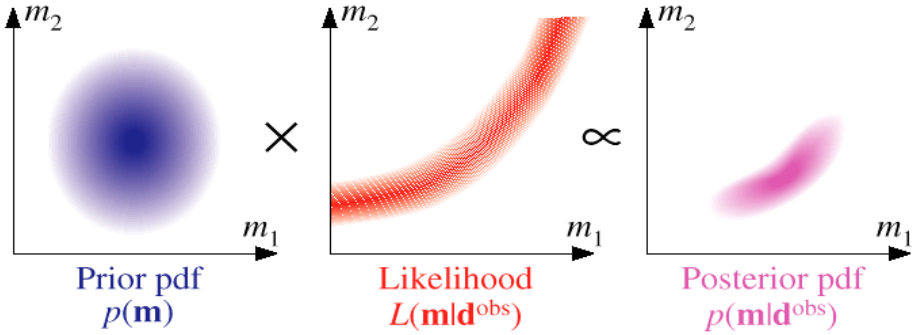

Figure 1.2: Schematic representation of the relationship among prior, likelihood function and posterior.

$$p(\mathbf{d}_{obs} \mid \mathbf{I}) = \int p(\mathbf{m} \mid \mathbf{I}) p(\mathbf{d}_{obs} \mid \mathbf{m}, \mathbf{I}) d\mathbf{m} \tag{1.2}$$

How to translate the vague prior information into a probability distribution is a controversial subject (Mosegaard and Tarantola, 1995; Kass and Wasserman, 1996; Curtis and Lomax, 2001). Unless a certain form of prior probability distribution is indicated by specific prior information, uniform distribution may be assumed in case more biased constraints are attached. This is especially common when Bayesian inversion is performed using stochastic sampling methods because pseudo-random, uniform sampling is easy to perform (Curtis and Lomax, 2001). The likelihood function $p(\mathbf{d}_{obs} \mid \mathbf{m}, \mathbf{I})$ is related to forward simulation and is therefore determined by the data error or noise, which is very difficult to characterize in terms of a single probability distribution. However, based on the Central Limit Theorem, it is safe to assume the most conservative Gaussian distribution for data error in realistic problems. The Central Limit Theorem is of great practical value in data analysis and provides a deep understanding of

why data uncertainties frequently have a Gaussian distribution. The distribution results is because the measured quantity is often the result of a large number of effects (Gregory, 2005). In practical applications, uniform prior and Gaussian data errors result in a multivariate Gaussian posterior surface with a simple topology for linear problems. In contrast, a multimodal posterior surface with a complicated topology of many hills and valleys for nonlinear problems.

1.2 Markov Chain Monte Carlo algorithms

As mentioned above, for the nonlinear, high-dimensional problems that usually result in a very complicated PPD topology of multiple hills and valleys, Markov Chain Monte Carlo (MCMC) algorithms may be the only possible tool to derive the target PPD while circumventing the direct evaluation of the computational demanding integration in equation (1.2) over the entire model space (Sen and Stoffa, 1996; Brooks, 1998; Gregory 2005).

In probability theory, a stochastic process is said to have the Markov property if the conditional probability distribution of future states of the process depends only upon the current state but not on any past states. This process is usually called the Markov process. Accordingly, a Markov chain is a discrete-time stochastic process with the Markov property, namely, given the present state, the future states are independent of the past states. Formally,

$$P_r\left(X_{n+1}=x \mid X_n=x_n,\ldots,X_1=x_1,X_0=x_0\right)=P_r\left(X_{n+1}=x \mid X_n=x_n\right). \quad (1.3)$$

The probabilities of going from state *i* to state *j* in single-step and *n* time steps are defined respectively as:

$$p_{ij} = P_r\left(X_1 = j \mid X_0 = i\right); \qquad p_{ij}^n = P_r\left(X_n = j \mid X_0 = i\right). \tag{1.4}$$

A Markov chain has a stationary distribution (independent of the initial state) *if and only if* it has the following two properties:

1) *Irreducibility*, that is each state in a Markov chain can access every other state; a set of states in which all members of the set are reachable by other members is called an *ergodic* class. Irreducibility of a Markov chain implies that every state can be reached from every other state in a finite number of transitions with a finite probability.

2) *Aperiodicity*, that is a state can recur at each next time step; this state has a period of 1 and is called aperiodic state. This property helps stop the chain from oscillating between different states in a regular periodic movement.

MCMC methods (which include random Monte Carlo walks) are a class of algorithms for sampling from target probability distributions (the PPD) based on constructing a Markov chain that has the target distribution as its stationary distribution. The state of the chain after a large number of steps is then used as a sample from the desired distribution. The most common application of these algorithms is in numerical calculation of multi-dimensional integrals. In these methods, an ensemble of "walkers" moves around randomly. At each point where the walker steps, the integrand value at that

point is counted towards the integral. The walker then may make a number of tentative steps around the area, looking for a place with a reasonably high contribution to the integral to move into the next. A Markov chain is constructed in such a way as to have the integrand as its equilibrium distribution. Surprisingly, this is often easy to do. Multi-dimensional integrals often arise in Bayesian statistics and computational physics, so MCMC methods are widely used in those fields.

The commonly used random walk MCMC methods are based on the Metropolis-Hastings algorithm (Metropolis and Ulam, 1949; Metropolis *et al*., 1953; Hastings, 1970), which generates a random walk using a proposal density and a method for rejecting proposed moves, and the Gibbs sampling, which requires that all the conditional distributions of the target distribution can be sampled exactly (see Appendix A for details). However, these methods move around the equilibrium distribution in relatively small steps, with no tendency for the steps to proceed in the same direction. Therefore, it unfortunately takes a long time for the walker to explore the entire space. More sophisticated algorithms, such as Simultaneous Over-relaxing, and Hybrid Monte Carlo, etc., that prevent the walker from doubling back have been developed for faster convergence, but they are hard to implement (Geman and Geman, 1984). Genetic Algorithm (GA) and Simulated Annealing (SA) have also been adapted to make MCMC analysis, but efficiency still remains a critical issue.

Davis (1991) showed that GA can be viewed as a Markov chain in that the conditional dependence of each population on its predecessor is completely described by its dependence on the parent (immediate predecessor) population, which is strictly

consistent with the definition of equation (1.3). Further, Davis and Principe (1993) demonstrated that, when the mutation operator is used, the Markov chain of GA is irreducible and consequently has a unique stationary distribution, and the mutation is a control parameter analogous to temperature in SA. Based on Davis's work, Suzuki (1998) made some extension and gave a more rigorous theoretical basis to the Markov chain of GA by additionally considering diminishing genetic operators. Suzuki (1998) proved that GA converges to a stationary distribution focusing on the uniform population with the optimal solutions, which forms a sound basis to use GA as a safe optimization method or sampling tool.

Although MCMC methods have been favorably used to explore the PPD in high-dimensional problems (Mosegaard and Tarantola, 1995; Sen and Stoffa, 1996; Brooks, 1998; Floris *et al.,* 1999), uncertainty quantification based on the PPD can still be computationally daunting, especially in problems with a time-consuming nonlinear forward simulator. Starting from the classical Metropolis-Hastings method and Gibbs' sampler (Metropolis and Ulam, 1949; Hastings, 1970; Geman and Geman, 1984; Gelfand and Smith, 1990; Gelfand *et al.,* 1992), some developments in MCMC algorithms have been made for better performance. The idea of a temperature ladder was introduced to design new MCMC methods such as Parallel Tempering, in which a sequence of distributions is simulated along a temperature ladder (Marinari and Parisi, 1992; Earl and Deem, 2005). Further, the multi-scaling technique was proposed, and the corresponding multi-scale coupled MCMC methods were designed. In these methods, a coarser scale is incorporated, although the model at the fine scale is of primary interest, and multiple

chains of different scales are run simultaneously. By swapping information between chains, the coarser scale helps speed up the forward simulation as well as facilitates better exploration of the posterior at the fine scale (Bouman and Liu, 1991; Yoon *et. al.,* 1999; Higdon *et al.,* 2002). In addition, Genetic Algorithms have also been proved to converge to a stationary distribution (Davis, 1991; Davis and Principe, 1993; Suzuki, 1998) and adapted as sampling tools (Holmes and Mallick, 1998; Liang and Wong, 2000, 2001). Liang and Wong (2000) proposed a new algorithm, the Evolutionary Monte Carlo, by incorporating attractive features of Genetic Algorithms and Simulated Annealing into the framework of the MCMC. A similar approach was reported by Stoffa and Sen (1991). The improvement of this new MCMC algorithm in both optimization and PPD simulation was demonstrated. Nevertheless, despite the developments mentioned above, computational efficiency and estimation accuracy still remain a critical factor for practical applications such as pre-stack seismic waveform inversion.

Model parameterization focusing on the number of free parameters (unknowns) is another very important issue in solving inverse problems. It is difficult to define an appropriate parameterization before the data are inverted. Usually, over-parameterization results in non-uniqueness and substantially increases computational cost while under-parameterization may be fast but can often lead to an inadequate data fit (Sen and Stoffa, 1991). Researchers recently proposed to directly treat the number of unknowns itself as an unknown, which leads to a trans-dimensional inverse problem, that is, one where the dimension of the parameter space is a variable to be determined by the data (Malinverno, 2000, 2002; Sambridge *et al.,* 2006). In a Bayesian framework, this essentially is a

"model selection" problem and is usually solved based on the estimated PPDs of different parameterizations (Carlin and Chib, 1993). A particular type of MCMC algorithm and a quantity termed "marginal likelihood" or "evidence" were highlighted in the work of Sambridge *et al.,* (2006) for sampling from the variable dimension spaces. Applications to both linear/linearizable problems with many unknowns and fully nonlinear problems with few unknowns were successfully demonstrated by Sambridge *et al.,* (2006). A similar algorithm was also applied to invert the walkaway VSP (vertical seismic profile) data for prediction of elastic properties in a layered earth model as well as for quantifying the posterior uncertainty (Malinverno and Leaney, 2005), which results in a good agreement between independent well logs and predicted elastic properties with smaller uncertainty by using a far-offset walkaway VSP. Appendix B provides more descriptions about MCMC algorithms.

Nevertheless, besides certain issues (e.g., model parameterizatio), accuracy of PPD estimation and computational efficiency still remain critical factors for practical applications. These problems lead to a central objective of this dissertation, which is to introduce a more powerful MCMC algorithm that has not yet become prominent in geophysics.

1.3 Integrated 4D seismic reservoir characterization

A reasonably accurate description of the reservoir model plays a very important role in multiple stages throughout a reservoir's life cycle, from the early exploration stage

to the final field-abandonment stage. In addition, reservoir surveillance during production is a key means to meeting goals of reduced operating costs and maximized recovery. However, a reservoir is usually buried hundreds to thousands of meters in the subsurface, and the only place where geoscientists are able to directly access the reservoir is at the well sites. Besides direct measurements of the production volumes of oil, water, and gas and pore pressure and fluid saturation at well locations, reservoir flow simulation has also been commonly used to understand and predict important geological, geophysical and engineering reservoir properties. Differences between actual and predicted observations are typically used to update the geological model of the reservoir and to revise the production strategy.

Better reservoir description calls for the integration of various data sources and expertise from different disciplines. Different data types usually provide different resolution; they correspond to different physics as well. To derive an accurate reservoir model and therefore obtain a reliable performance prediction with as little uncertainty as possible, the idea of dynamic data integration has evolved. History matching was introduced to capture a reservoir's static and dynamic features by integrating dynamic production data and static geologic scenarios. Thus, the resulting petrophysical model of the reservoir is found so that it honors the production measurements at well locations throughout the entire production time. Static data summarize all available measurements and interpretations about the target reservoir, such as:

- Core data: porosity, permeability, etc.
- Well logs: lithology and fluid types around well bores

- Geologic outcrop analog data
- Interpretation of sedimentation and stratigraphy
- Interpretation of horizons and faults
- Rock physics and PVT data

This history matching scheme is expected to reduce uncertainty in reservoir characterization by incorporating more constraints from various data sources. However, as described above, evaluation of data fit is only performed at a few sparse well locations and therefore lacks lateral resolution of the reservoir heterogeneity. It certainly would be beneficial to incorporate some type of spatially dense data in this history matching scheme in order to assist in constraining the dynamic reservoir characterization between wells. From this viewpoint, seismic data is the most promising candidate because of the extensive spatial coverage and dense lateral sampling it entails.

Time-lapse, or 4D, seismic monitoring is a reservoir exploitation technique based on successive 3D seismic surveys. In a 4D seismic study, it is assumed that the reservoir's geology remains constant through time, and differences over time in seismic attributes are due to changes in fluid flow variables such as fluid saturation and pore pressure, etc. during production/injection. Studies using 4D seismic can be traced back to late 1980's and early 1990's (Wayland and Lee, 1986; Greaves and Fulp, 1987; Dunlop et al., 1991; Lumley et al., 1999; Behrens et al., 2002). According to Lumley and Viejo (2004), total time-lapse seismic expenditures in 2002-2003 were about $500 million. 4D seismic technique involves repeating the seismic surveys to construct and compare

seismic images to monitor time-varying dynamic fluid flow properties in the subsurface during production/injection. The difference between the seismic surveys can then be interpreted in terms of the production-related changes in reservoir properties. Using 4D seismic data h increases reserves and recovery by

- Locating bypassed and undrained reserves
- Optimizing infill well locations and flood patterns
- Identifying reservoir compartmentalization and permeability pathways

Utilizing 4D seismic data can also help decrease operating costs by

- Reducing initial development well counts
- Optimizing phased developments using early field-wide surveillance data
- Reducing reservoir model uncertainty
- Reducing dry holes and targeting optimal completions

As a result, the time-lapse, or 4D, technique has recently been developed to derive the static petrophysics of a reservoir as well as to image the inside dynamic fluid flow over production time (Landro, 2001; Lumley *et al*., 2003), as the changes in static petrophysical properties and dynamic fluid movement can be decoupled. Accordingly, schemes of seismic history matching have been proposed for more accurate reservoir characterization (Chapin *et al*., 2002; Pickering *et al.,* 2005), in which a reservoir model is found such that, after both flow simulation and seismic modeling, the synthetic seismic and production data match the recorded seismic data and production rates at the test wells. If the match is poor, the reservoir model will be stochastically perturbed to improve the data fit. Nevertheless, most existing applications using 4D seismic data have tended to be

qualitative rather than quantitative, and they are time-consuming iterative processes posing several challenges in practical applications. First, in the formulation part, how to integrate all available data sources corresponding to different physics and with different resolutions is not trivial. Second, in the algorithm part, it is difficult to design a powerful algorithm to solve this integrated inverse problem for relatively accurate model estimation within an acceptable time period. These two aspects are addressed in this research.

Seismic history matching is a process for deriving reservoir model parameters from surface-recorded 4D seismic and well production data, which constitutes a joint inverse problem. Considering the inherent non-uniqueness of the inverse problem and the unique feature of Bayesian inference in data integration and uncertainty analysis, it is superior to formulate the inverse problem in a statistical Bayesian framework and to fully solve it by stochastically reconstructing the PPD function using MCMC algorithms (Sen and Stoffa, 1996; Tarantola, 2005; Buland and Ouair, 2006). This PPD is the complete solution of the integrated seismic history matching problem, providing all consistent reservoir models.

In this research, I have proposed a seismic history matching scheme to simultaneously integrate 4D seismic and well production data for quantitative reservoir characterization. This integrated scheme is formulated in a Bayesian framework and solved by PPD sampling using a new MCMC method. To integrate the time-lapse seismic and production data that are assumed to be conditionally independent of each other, the joint PPD can be described as:

$$p(\mathbf{m}|\mathbf{S}^{obs},\mathbf{P}^{obs}) = \frac{p(\mathbf{S}^{obs}|\mathbf{m})\cdot p(\mathbf{P}^{obs}|\mathbf{m})\cdot p(\mathbf{m})}{p(\mathbf{S}^{obs})\cdot p(\mathbf{P}^{obs})}, \tag{1.5}$$

where **m** is the reservoir model vector. The term $p(\mathbf{m}|\mathbf{S}^{obs},\mathbf{P}^{obs})$ is the joint PPD of reservoir properties for the given time-lapse seismic and production data, which provides a complete solution of the joint inverse problem as well as a theoretically correct way to quantify corresponding uncertainties. Likelihood function $p(\mathbf{S}^{obs}|\mathbf{m})$ corresponds to the forward process of seismic modeling, which relates to modeled reservoir properties **m** through a rock physics model. Likelihood function $p(\mathbf{P}^{obs}|\mathbf{m})$ relates to the fluid flow simulation that is performed directly on the model vector **m**. Prior knowledge $p(\mathbf{m})$ may come from other independent sources such as well logs and laboratory measurements. Denominator term of equation (1.5) is a normalization factor that is the integration of the numerator over the entire model space. Again, as mentioned in section 1.3, reasonably accurate and efficient estimation of the joint PPD in equation (1.5) calls for powerful a MCMC algorithm.

1.4 Uncertainty quantification

Quantitative characterization of physical properties of the earth's interior from surface-recorded data and other measurements is essentially a mapping from data space to model space. However, this mapping is dubious because it is not a one-to-one mapping due to the inherent non-uniqueness, which is caused by incompleteness and inaccuracy of the observed data. Non-uniqueness usually leads to a range of models equally consistent

with the data and may preclude a deterministic solution. In statistical terminology, our state of knowledge about an earth model is inaccurate, incomplete and therefore necessarily probabilistic. Bayesian inference provides a means to handle situations in which information is incomplete or inaccurate and a deterministic reasoning cannot be made, as it constructs the probability distribution for assessing the plausibility of competing models and the corresponding uncertainties (Tarantola, 1987; Gregory, 2005). This framework is superior since it provides not only a complete range of solutions for given information but also a convenient way to quantify uncertainties arising from the inherent incompleteness and inaccuracy. Simply speaking, the most likely model parameter values as well as the corresponding uncertainty bound can be determined based on the estimated PPD coming from MCMC samples. However, it is very hard to display the PPD with a complicated multi-modal topology in a high-dimensional space. As a result, several measures of dispersion (mean, covariance, etc.) and marginal PPDs are often used to describe the most probable solution and to quantify uncertainty. The posterior mean model and posterior model covariance matrix are calculated according to

$$\langle \mathbf{m} \rangle = \int \mathbf{m} p(\mathbf{m} \mid \mathbf{d}_{\mathbf{obs}}, \mathbf{I}) d\mathbf{m}, \tag{1.6}$$

$$\mathbf{C}_{\mathbf{M}} = \int (\mathbf{m} - \langle \mathbf{m} \rangle)(\mathbf{m} - \langle \mathbf{m} \rangle)^T p(\mathbf{m} \mid \mathbf{d}_{\mathbf{obs}}, \mathbf{I}) d\mathbf{m}, \tag{1.7}$$

$$p(m_i \mid \mathbf{d}_{\mathbf{obs}}, \mathbf{I}) = \int dm_1 \int dm_2 \cdots \int dm_{i-1} \int dm_{i+1} \cdots \int p(\mathbf{m} \mid \mathbf{d}_{\mathbf{obs}}, \mathbf{I}) dm_M \ . \tag{1.8}$$

The posterior correlation matrix can also be derived from the covariance matrix to describe the inter-dependence between different parameters. The marginal PPD is a

special kind of projection of the joint PPD to a particular dimension axis, and it can be approximated by the frequency distribution of that particular parameter. All these integrals fall into a general form as shown in equation (1.9), which is the integration of the derived PPD over the entire model space. It is not tractable in realistic high-dimensional applications. This means that it is still not easy to quantify uncertainties even though the PPD was derived. Therefore, a way to derive these dispersion measurements while avoiding the direct integration computation is still necessary.

$$\Omega = \int f(\mathbf{m}) p(\mathbf{m} \mid \mathbf{d}_{\mathbf{obs}}, \mathbf{I}) d\mathbf{m} \quad . \qquad (1.9)$$

Uncertainty quantification is important and therefore is also specifically addressed in this research for two major reasons. First, in terms of an inverse problem itself, it is logical to derive not only the most likely model but also the corresponding uncertainty in order to account for the inherent non-uniqueness. By doing so, both a complete solution of an inverse problem and an appropriate model parameterization in some cases can be obtained. Second, in petroleum exploration and production, decisions of reservoir development and management, especially for complex deep-water reservoirs, are always related to risks due to uncertainties present in the process (Costa and Schiozer, 2003; Schiozer *et al*., 2004). One of the most common uncertainties comes from the geological models of the reservoir. Therefore, a precise risk assessment requires not only an accurate production prediction and model estimation but also an accurate uncertainty quantification. Based on the sampled PPD of reservoir properties, measures of dispersion in model space will be correctly quantified for uncertainty analysis, as illustrated in

equations (1.6), (1.7) and (1.8), which provide the necessary ingredients to facilitate an engineer's risk assessment and a manager's decision making. In this research, all these uncertainty measurements are estimated based on the PPD samples drawn by a new MCMC algorithm.

1.5 Objectives and dissertation organization

The central objective of this dissertation is to report on the development of new and powerful global optimization and MCMC sampling algorithms that make nonlinear high-dimensional inverse problems manageable. Specifically, better performance of the new algorithms is demonstrated by testing on a nonlinear pre-stack seismic waveform inversion problem as well as an integrated 4D seismic history matching problem for quantitative reservoir characterization in a Bayesian framework. I will show that the proposed integrated workflow in the Bayesian framework leads to an accurate static reservoir petrophysical model estimation and a dynamic imaging of fluid flow within the reservoir. Using the newly developed MCMC algorithm, the corresponding uncertainty in quantitative reservoir characterization can be correctly quantified, which facilitates an accurate risk assessment associated with reservoir decision making and management.

In chapter 2, the conventional binary Genetic Algorithm is reviewed. I incorporate some new developments in the conventional GA to develop a new GA with better performance in terms of estimation accuracy and computational efficiency. Specifically, the traditional binary coding is replaced by real coding to promote better and more efficient exploitation of the model space for estimation of the best-fitting model. Some

concepts of Simulated Annealing (SA) are also incorporated to speed up convergence while preventing premature stagnation. This real-coded GA is further developed using a multi-scaling technique to take advantage of benefits from both the fine scale and the coarse scale. A comparison is shown between the conventional and the new GAs. The new real-coded multi-scale GA is also applied to a pre-stack seismic waveform inversion problem to derive a set of best-fitting elastic parameters.

In chapter 3, the new real-coded multi-scale GA is further developed as a powerful MCMC algorithm, a multi-scale GA-based MCMC, for better exploration of the model space aiming to construct the target PPD surface. The applicability and usefulness of this new MCMC algorithm are investigated. This new multi-scale coupled MCMC is first justified using an analytical example, and then further applied to estimate seismic elastic parameters of a 1D earth model and to quantify corresponding uncertainties based on pre-stack seismic waveforms. It is found that multi-scaling is particularly useful in addressing the model parameterization issue.

Discussions of 4D seismic, integrated seismic history matching, flow simulation, seismic modeling and rock physics models are presented in chapter 4. A seismic history matching workflow is proposed to integrate as many data sources as available for quantitative reservoir petrophysics and inside fluid flow imaging using a global optimization algorithm, which is tested on a 2D numerical example. In chapter 5, the integrated history matching of 4D seismic and well production data, which essentially is a joint inverse problem, is formulated in a Bayesian framework and is stochastically solved by constructing the PPD surface using the new multi-scale MCMC algorithm. The

estimated PPD enables the determination of the most likely reservoir model for the given datasets. In addition, based on the samples drawn by the MCMC method, the uncertainty associated with a reservoir model estimation is quantified. A synthetic application in a 3D reservoir is used to demonstrate how the multi-scale MCMC is applied to quantitatively and accurately derive the static and dynamic properties of a reservoir as well as to quantify uncertainties.

Chapter 6 summarizes the research covered by this dissertation and discusses some potential directions for future research.

Chapter 2

Multi-scale real-coded hybrid Genetic Algorithm

Inverse problems, the processes for deriving subsurface properties from observed data and measurements, are common in geophysics. Geophysical inverse methods have been a subject of active research for decades and can be broadly classified into two major categories: (1) direct inversion or operator-based inversion methods and (2) model-based inversion methods.

Direct inversion methods are formulated based on the physics of the forward problem by designing a mathematical inverse operator to the observed data in order to directly derive a model of interest. The best known direct inversion method in seismology is the so-called layer-stripping method (Singh *et al.*, 1989; Ziolkowski *et al.*, 1989).

As a comparison, no attempt is made to reverse the forward operator in a model-based inversion method. Instead, synthetic data are usually generated for an assumed model and compared against the observed data to evaluate the match between them. If the match is satisfactory, the derived model is accepted as the solution. Otherwise, the model is updated, and the synthetics are re-computed and re-compared against the observations, which is repeated until an acceptable data match is obtained. Thus, in this approach, the inversion is iterative and can be viewed as an optimization process. Depending on the search scheme applied to find the optimal solutions, model-based inversion methods can be classified into the following categories (Sen and Stoffa, 1995):

- *Linear/linearized methods*: These methods assume that the data are a linear function of the model parameters. Therefore, it is possible to design a matrix as the forward operator and directly apply it to the model. In this way, the inverse problem is formulated into a well-known linear algebra problem.
- *Gradient-based methods*: These methods use gradient (derivative) information to update the current model, which is repeated until the updates become negligible.
- *Enumerative or grid search method*: This method involves searching each point in the entire model space. Computation of synthetic data for a large model space in many applications is demanding and is usually not practical.
- *Monte Carlo methods*: These methods involve the random sampling of model space with the goal of obtaining good solutions in a finite number of trials. This is a completely blind search, and therefore, it may be computationally very expensive.
- *Directed Monte Carlo methods*: The global optimization methods Simulated Annealing (SA) and Genetic Algorithm (GA) belong to this category. These methods perform random sampling with some directivity to guide the searching.

Compared to local optimization methods such as the gradient-based methods, global optimization methods have several benefits. Firstly, even for the objective function

with a complicated topology of multiple optima, the global optimization process theoretically converges to the global optimum no matter where it starts in the model space. Secondly, the global optimization methods do not require that a gradient and sometimes a curvature (second-order derivative) matrix be calculated at each point through iterations. These gradient calculations may be a computationally formidable task in many cases. In addition, they do not require calculating the inverse of a large matrix, which is not a trivial task.

Since the work of Kirkpatrick *et al.* (1983), Simulated Annealing (SA) has been applied to a wide variety of problems in geophysics. Most SA algorithms are statistically guaranteed to attain equilibrium distribution and possibly to reach the global optimum, and therefore they are suitable for best-fitting model estimation. In practice, however, these algorithms are sometimes problematic due to inappropriate parameter selection. For example, it is important to choose the starting temperature and cooling schedule properly.

Unlike SA, which is based on an analogy with a physical annealing process, Genetic Algorithm (GA) (Holland, 1975; Goldberg, 1989) is based on analogy with the process of natural biological evolution. The basic GA is quite robust and not particularly sensitive to the randomly selected starting models as long as a sufficient number of models are employed. The primary advantage of a basic GA is that it always converges toward models with higher fitness values. There are, however, several problems with using the basic GA that need to be addressed. First, there is no guarantee that the optimum solution will be found, although the finally derived solution may be good enough in many cases. The convergence of a GA can be premature if a small number of

models are employed. Rapid convergence to minimize computational cost may be undesirable since model space will not be sufficiently exploited and the population may become homogeneous around a local fitness maximum that is not near the global maximum. In contrast, convergence toward the global maximum may be slow because some model parameters have only minor impact on the fitness, so that extensive sampling of model space often results in minor improvements at significant computational cost.

As mentioned in chapter 1, it is better to formulate an inverse problem in a statistical framework and fully solve it by constructing the PPD surface. However, in many situations, rather than the PPD surface, users may be only interested in the best-fitting or most likely model that is the maximization of the posterior distribution. Therefore, the main objective of this chapter is to develop a new global optimization method, a multi-scale real-coded hybrid GA (Hong and Sen, 2006; Hong and Sen, 2007). This new GA is expected to have better performance in terms of model estimation accuracy and computational efficiency. Specifically, starting from a conventional binary-coded GA, a new optimization algorithm is developed by integrating new features of real-coding and multi-scaling techniques as well as some concepts from Simulated Annealing.

2.1 Basic Genetic Algorithm

Genetic Algorithm (GA) is an intelligent maximization technique for functions defined on high-dimensional spaces, which simulates the biological evolutionary processes of *selection*, *crossover* and *mutation* to increase the fitness toward better solutions. The unique feature of a GA is that it is able to work on a population of models

simultaneously. A typical GA requires two conditions to be defined: (1) a genetic representation of solutions or models, and (2) a fitness function to evaluate the goodness of those solutions or models. Users of GAs have had great success in search and optimization problems. One reason for much of their success is the ability to exploit the information accumulated about an initially unknown search space in order to bias subsequent searches into useful subspaces. Chatterjee *et al*. (1996) provides a good overview of the concepts in GA as related to statistics.

Before introducing the new real-coded multi-scale GA that I developed, the conventional GA is reviewed as follows. Suppose the fitness function to be maximized is $F(\mathbf{m})$, where $\mathbf{m}$ is the model vector $\mathbf{m} = [m_1, m_2, ..., m_L]^T$ with fixed length of L (number of model parameters), and each generation consists of M solution models: $\mathbf{m}^1, \mathbf{m}^2, ..., \mathbf{m}^M$. The initial step in a conventional GA is to design a binary coding scheme that represents these solution models and creates an initial population of models to a given problem. This way, each bit corresponds to a *gene*, and each individual model in the population is described by its bit string or *chromosome*. Conventionally, chromosomes in a GA have fixed scale or length equal to the number of model parameters L, but M, the number of models in each generation, may be adjusted. GA optimization is initiated by randomly choosing a population of models from the model space based on the calculated values of fitness. For example, an initial population of models $\mathbf{m}^1_{(0)}, \mathbf{m}^2_{(0)}, ..., \mathbf{m}^M_{(0)}$ could be created by randomly selecting M vectors of length L based on certain distribution from the model space. Here, the subscript in parentheses denotes the iteration of generation number within the algorithm. These starting models

are then binary coded and progressively optimized for better estimation by repeatedly employing the three genetic operators (*selection*, *mutation* and *crossover*).

Selection.In this step, the population of models is altered to allow better models to remain whereas poorer models may be removed. Selection proceeds by choosing M model vectors with replacement from the current population to form a new generation based on probabilities proportional to the corresponding fitness $f(\mathbf{m}^i)$. Therefore, for one model $\mathbf{m}_{(t)}^i$, the probability of being selected is $f(\mathbf{m}_{(t)}^i)/\sum_{j=1}^M f(\mathbf{m}_{(t)}^j)$. In this way, those models with higher fitness values are more likely to be selected for further operation whereas poorer models are potentially discarded.

Crossover. To perform a crossover, models chosen from the selection step are paired, and some of their values are traded between each individual pair. It has been recommended that the probability of performing a trade should be between 0.6 and 0.95 (see Back (1993) for more discussion). Several crossover mechanisms may be performed on a single pair to generate a new pair of models. In a one-point crossover, a single element is randomly picked between 1 and L and all elements after this chosen element are traded. In a k-point crossover, k ($<L$) elements are chosen, and segments between those chosen elements alternate between trading and not trading. In a uniform crossover, every element has some finite probability of being traded. In addition, more complicated forms of crossover have been proposed, such as the snooker crossover discussed in Liang and Wong (2001).

Mutation. In this step, each element of each model vector in the current population is perturbed with a probability. Mutation helps to make the models diverse to prevent premature convergence, and helps to jump out of local optima. The mutation probability has to be very small, otherwise current good models are likely to be destroyed. For instance, if the mutation probability is 0.01, approximately 0.01xMxL elements are mutated in each optimization cycle.

At the end of the mutation step, the next generation of the algorithm begins. These three steps are repeated until some measure of convergence is satisfied. Several stopping criteria could be used to terminate the GA optimization process. Usually, convergence is considered satisfied if the algorithm does not improve the fitness of the most model vectors for many iterations, that is, the ratio of average fitness to maximum fitness exceeds some limit, or the absolute fitness of the most of model vectors exceeds some preset threshold. The user can also simply specify the maximum number of generations as the stopping criterion.

2.2 Multi-scale real-coded hybrid Genetic Algorithm

2.2.1 Combining elements of SA into GA

To address the issues of premature convergence in a conventional GA, that is a rapid stagnation of search caused by the lack of diversity in the population, and slow convergence to the global maximum because some model parameters have minor impact on the fitness, different schemes have been introduced (Baker, 1987; Goldberg, 1989;

Whitley, 1989; Stoffa and Sen, 1991; Sen and Stoffa, 1992). In this section, we develop a hybrid GA by adopting the scheme of Stoffa and Sen (1991) to incorporate some concepts of Simulated Annealing (SA), such as temperature and update probability, into a basic GA.

Recognizing that the stretching of the fitness employed in the GA plays the same role as temperature in an SA, all of the ideas related to temperature and temperature schedules can readily be applied to a GA. Therefore, the selection probability as described in section 2.1 is replaced as follows:

$$p_s(\mathbf{m}_{(t)}^i) = \frac{\exp\left(\frac{f(\mathbf{m}_{(t)}^i)}{T_{(t)}}\right)}{\sum_{j=1}^{M}\exp\left(\frac{f(\mathbf{m}_{(t)}^j)}{T_{(t)}}\right)} \quad , \qquad (2.1)$$

where subscript (t) is the index of current generation and $T_{(t)}$ is the temperature value at the current generation. Again, $f(\mathbf{m}^i)$ is the fitness function on the model vector $\mathbf{m}_{(t)}^i$. After the adaptation mentioned above, in the selection step, model $\mathbf{m}_{(t)}^i$ is selected based on a probability p_s defined in equation (2.1). The temperature may follow a pre-defined cooling schedule.

The genetic operators of selection, crossover and mutation can be grouped together and considered as a replacement for the pure random walk in an SA. By simply keeping track of the previous generation's models and their fitness values, the algorithm is able to form a temperature-dependent acceptance probability for all the models individually. Therefore, an update probability is incorporated into a GA to decide

whether a model from a previous generation should be used in place of a current model. Once a new model's fitness has been evaluated, it is compared to the fitness of the corresponding model from the previous generation. If the current model's fitness is greater, the current model is always kept. If it is less, the previous generation's model replaces the current model with the specified update probability p_u, which now plays the role of the acceptability as used in the classical SA.

As a result, the hybrid GA has the ad-hoc fitness stretching replaced by a temperature-dependent probability. The hybrid GA includes the update information from comparing a model's current fitness to its fitness in the previous generation.

2.2.2 Real-coded GA

Fixed-length and binary-coded bit strings for the representation of model vectors have dominated GA research because they are shown to be the most appropriate ones and amenable to simple implementation (Goldberg, 1991). However, the GA's good properties do not stem from the use of bit strings (Antonisse, 1989; Radcliffe, 1992). In addition, the binary representation encounters certain difficulties when dealing with continuous search spaces with large dimensions and when great numerical precision is required (Herrera *et al.,* 1998). For these reasons, alternative non-binary representation has drawn much attention for certain application problems. One of the most important non-binary representations is the so-called real number representation, which seems particularly natural for optimization problems with variables in continuous search spaces (Davis, 1991; Wright, 1991; Herrera *et al.,* 1995). In real representation, a chromosome is

a vector of floating point numbers with a size the same as the length of the model vector, which is the candidate solution to the problem, and its precision is restricted to that of the computer by which the algorithm is carried out. In this way, each gene represents a variable of the problem. The values of the genes are forced to remain in the interval established by the variables that they represent.

The use of real representation makes it possible to use large domains for the variables, which is difficult to achieve in binary implementations where increasing the domain would mean sacrificing precision. Another advantage of using real coding is the capacity to exploit the graduality of the functions with continuous variables. Here, the concept of graduality refers to the fact that slight changes in variables correspond to slight changes in the function (Herrera *et al.,* 1998). A comparative study conducted by Janikow and Michalewicz (1991) concluded that the real-coded GAs outperformed the binary-coded ones in many optimization problems, especially the high-dimensional nonlinear problems. Accordingly, different mechanisms of crossover and mutation operations are also tuned for using the real-floating point numbers instead of the long strings of zeros and ones (Wright, 1991; Ono and Kobayashi, 1997; Herrera and Lozano, 2000; Liang and Wong, 2001).

The hybrid GA introduced in section 2.2.1 is further developed by replacing the conventional binary representation to real representation in which model vectors are simply kept in decimal forms. Using real representation is very close to the natural formulation of many problems, and therefore, the coding and decoding processes in a

binary GA are avoided, which increases the GA's implementation speed. The real coded crossover and mutation operations in this new real coded hybrid GA are described below.

Different crossover mechanisms may be used to trade information between a pair of model vectors that are chosen through the selection operator, such as single-point crossover, multi-point crossover and uniform crossover. The uniform crossover mechanism is used in this research, which means that every parameter pair of the selected model vectors will potentially be traded, based on a probability. Once a parameter pair is determined for trading information, a scheme called intermediate recombination is performed to stochastically combine the two parent variables to generate corresponding two offspring variables, following the rule shown below:

$$offspring = parent1 + alpha \cdot (parent2 - parent1)$$

where *alpha* is a scaling factor chosen uniformly at random over an interval [-*d*, 1+*d*]. A recommended choice is *d*=0.25. In this way, each variable in the offspring is the result of combining the variables according to the above expression with a new *alpha* chosen for each variable. Figure 2.1 on page 44 illustrates the area of the range of offspring variable, which is defined by the range of parent variable with *d*=0.25.

In crossover, if a single-point crossover is used, a crossover point is first randomly picked (e.g., position of I). All elements of the selected and paired model vectors after this point are traded according to the mechanism described above, which is also illustrated in the following diagram. If a uniform crossover is used, every element of the paired model vectors will be traded individually based on an identical probability. In

the diagram below, the starred elements after the selected single crossover point denote that their values are affected by the crossover step.

$$\mathbf{m}_{(t)}^{1} = \left\{ m_1^1, m_2^1 \ldots, m_I^1, m_{I+1}^1, \ldots, m_L^1 \right\} \quad \xrightarrow{\text{Crossover}} \quad \left\{ m_1^1, m_2^1 \ldots, m_I^1, m_{I+1}^{1*}, \ldots, m_L^{1*} \right\}$$

$$\mathbf{m}_{(t)}^{2} = \left\{ m_1^2, m_2^2 \ldots, m_I^2, m_{I+1}^2, \ldots, m_L^2 \right\} \quad \xrightarrow{\text{Crossover}} \quad \left\{ m_1^2, m_2^2 \ldots, m_I^2, m_{I+1}^{2*}, \ldots, m_L^{2*} \right\}$$

After crossover, offspring model vectors undergo mutation. Offspring parameters are mutated by adding white noise based on a low probability. Many papers (Back, 1993; Muhlenbein and Schlierkamp-Voosen, 1993) reported on the results using the optimal mutation probability. Muhlenbein and Schlierkamp-Voosen (1993) concluded that a mutation rate of *1/n* produced good results for a broad class of test functions, where *n* represents the number of parameters in the model vector. However, the mutation rate is independent of the population size. Similar results are also reported in Back (1993). Supposing that a variable has a lower and an upper bound of 8 and 28, respectively, Figure 2.2 shows how the mutation works on a variable with a value of 12 following the rule:

$$mutated\ \ parameter = parameter + delta \cdot (rand - 0.5) \cdot range,$$

where *rand* stands for a random number between 0 and 1; *range* is determined by the variable's minimum and maximum values, that is 20 in this example; *delta* is the size of the mutation step, which is usually difficult to choose. Small steps are often successful, although sometimes bigger steps are quicker. The optimal step size depends on the

problem considered and may vary during the optimization process. In this example, if *delta* is set to 0.1, the resulting mutated variable will be in the range of [11, 13], which is equivalent to adding a small white noise to the current value of 12.

After mutation, the generated new population of real-coded model vectors will be evaluated against the previous population for an update based on the probability in equation (3.1), which is immediately followed by a new genetic cycle without decoding and coding as needed in a binary GA.

2.2.3 Multi-scale real-coded hybrid GA

The study of geophysical inverse problems involves properly discretizing the continuous earth and then estimating the free parameters based on the observed data. Whether we have multi-scale data, or our primary interest is only in the fine-scale earth model, information from different scales are useful to facilitate exploration of the model space on the fine scale. Typically, a model at a fine scale may be of primary interest as it provides detailed information about the earth's interior. However, when modeling with a fine parameterization, more free parameters will be induced into the model and therefore computational cost will be increased, especially when a time-consuming forward simulator is involved. On the other hand, modeling with a coarser parameterization requires fewer free parameters, which helps to speed up the convergence. However, a coarser scale usually leads to rough information about the model and yields inadequate data fit as well. In order to take advantage of the benefits of both the fine scale and the coarse scale, multi-scaling techniques have been performed in several fields to estimate

parameters simultaneously modeled at multiple resolutions in high-dimensional spaces (Bouman and Liu, 1991; Yoon *et al*., 1999).

A GA that is able to accommodate variable model length was proposed by Goldberg *et al*. (1989). Their GA overcomes the limitation of fixed resolution and facilitates maximization. In this study, we also employ multi-scaling technique to extend the new hybrid GA described in section 2.2.2 in order to take advantage of the benefits from different scales to efficiently and accurately exploit the model space for the best-fitting solution at fine scale. To accommodate multiple resolutions (*e.g*., total number R of resolution levels), the likelihood function has to be defined on different scales. For scale i, it may be described as $L^{(i)}\left(\mathbf{m}^{(i)} \mid \mathbf{d}_{\mathbf{obs}}{}^{(i)}, \mathbf{I}^{(i)}\right)$, where $\mathbf{m}^{(i)}$, $\mathbf{d}_{\mathbf{obs}}{}^{(i)}$ and $\mathbf{I}^{(i)}$ stand for the model vector, data vector and prior at scale i, respectively. If only one data vector exists, then $\mathbf{d}_{\mathbf{obs}}{}^{(1)} = \mathbf{d}_{\mathbf{obs}}{}^{(2)} = ... = \mathbf{d}_{\mathbf{obs}}{}^{(R)} = \mathbf{d}_{\mathbf{obs}}$. This leads to the following overall joint likelihood function that is to be maximized

$$L\left(\mathbf{m} \mid \mathbf{d}_{\mathbf{obs}}, \mathbf{I}\right) = \prod_{i=1}^{R} L^{(i)}\left(\mathbf{m}^{(i)} \mid \mathbf{d}_{\mathbf{obs}}{}^{(i)}, \mathbf{I}^{(i)}\right) \tag{2.2}$$

In this framework, R conditionally independent likelihood functions need to be exploited. The multi-scale GA starts with a randomly selected population of models defined on R scales. Because these models are of different lengths, modifications have to be made in order to perform the genetic operations appropriately. Before performing selection, models are grouped according to the scale, that is, all models of the same scale are collected into one subgroup. The conditionally independent portion of the joint likelihood, that is $L^{(i)}\left(\mathbf{m}^{(i)} \mid \mathbf{d}_{\mathbf{obs}}{}^{(i)}, \mathbf{I}^{(i)}\right)$ and $i \in \{1, 2, ..., R\}$, is used as a measure of fitness of proposed

solution vectors at different scales, and the corresponding selection probability is defined by the following equation.

$$p_s(\mathbf{m}_{(t)}^{i}) = \frac{L^{(i)}\left(\mathbf{m}_{(t)}^{i} \mid \mathbf{d}_{\mathbf{obs}}{}^{(i)}, \mathbf{I}^{(i)}\right)}{\sum_{j=1}^{R} L^{(j)}\left(\mathbf{m}_{(t)}^{j} \mid \mathbf{d}_{\mathbf{obs}}{}^{(j)}, \mathbf{I}^{(j)}\right)} \tag{2.3}$$

Based on this probability, selection is then performed separately within each subgroup exactly as is done in a traditional GA. There should be at least two model vectors for each scale, and experimentation has shown that the algorithm works better with several solution vectors at each scale. At this point, the algorithm is ready to go through crossover and mutation for updates.

Crossover of multi-scale GA differs dramatically from the traditional GA because it cannot be directly performed on models of varying lengths even though the entire population can be paired as before. To overcome this difficulty, models have to be decomposed appropriately so that they have one part that is of common length and another part that contains additional information for completely specifying the model parameters. The decomposed model vector can be expressed as $\mathbf{m}^{(i)} = f^{(i)}\left(\boldsymbol{\varphi}^{(i)}, \boldsymbol{\lambda}^{(i)}\right)$, where $f^{(i)}$ is a up-scaling or down-scaling operator decomposing a vector of higher dimension to a vector of lower dimension or composing a vector of lower dimension to a vector of higher dimension. It can simply be a linear operation such as averaging or summing over fine pixels to yield coarse pixels. Different averaging techniques can be performed for the linear operation such as the physics-based Backus averaging (Backus,

1962), which involves averaging the model parameters of a stack of thin layers into effective average properties similar to those of a single thick layer. For simplicity, the standard average is used in this paper as the linear operator. Vector $\boldsymbol{\varphi}^{(i)}$ has the same length and physical meaning regardless of the scale i and therefore crossover is performed only on $\boldsymbol{\varphi}$, and the $\boldsymbol{\lambda}$ vector is not involved. The following diagram schematically illustrates a single-point crossover between two solution vectors of different scales. First, a crossover point is randomly picked (position of I), and all elements after this point and in gray areas are traded according to the mechanism described in section 2.2.2. As shown, information is exchanged only between $\boldsymbol{\varphi}$ parts that have a common length for all the scales. Starred parameters denote that their values are affected by the crossover operation. It is through these multi-scale crossovers that information is shared across scales or resolutions. In this dissertation, a uniform crossover mechanism is used to trade information, which is supposed to have better performance. As mentioned before, with a uniform crossover, all the paired model parameters are stochastically combined or not based on an identical probability.

$$\begin{array}{l} \mathbf{m}_{(t)}^{1} = \left\{ \varphi_1^1, \ldots, \varphi_I^1, \varphi_{I+1}^1, \ldots, \varphi_L^1, \lambda^1 \right\} \\ \mathbf{m}_{(t)}^{2} = \left\{ \varphi_1^2, \ldots, \varphi_I^2, \varphi_{I+1}^2, \ldots, \varphi_L^2, \lambda^2 \right\} \end{array} \xrightarrow{\text{Crossover}} \begin{array}{l} \left\{ \varphi_1^1, \ldots, \varphi_I^1, \varphi_{I+1}^{1*}, \ldots, \varphi_L^{1*}, \lambda^1 \right\} \\ \left\{ \varphi_1^2, \ldots, \varphi_I^2, \varphi_{I+1}^{2*}, \ldots, \varphi_L^{2*}, \lambda^2 \right\} \end{array}$$

Mutation of this multi-scale GA remains the same as that in a conventional single-scale GA. In this step, values of some of the parameters are perturbed following the scheme introduced in section 2.2.2. With real coded model vectors, random noise is

generally added to each component with a probability. This mutation probability has to be small in case that some good model vectors are destroyed. The elitist strategy has been proposed to save the best model vector on the scale of interest from being destroyed through crossover or mutation steps, which may increase the convergence speed (DeJong, 1975). The pseudo code of this multi-scale real-coded hybrid GA is provided in Figure 2.3. Several stopping criteria can be used to terminate the optimization process. The process can be stopped automatically when the maximum fitness or the ratio of average fitness to maximum fitness exceeds some limit. Users can also simply specify the maximum number of generations as the stopping criterion, as was done in our examples. The superior performance of this new algorithm is demonstrated in section 2.3 by using a 1D pre-stack seismic waveform inversion example.

2.3 Example: pre-stack seismic waveform inversion

One approach to solving inverse problems may be termed “exploitation of model space”, in which the inverse problem is cast into an optimization framework. This involves searching a multi-dimensional model space for an optimal set of model parameters that best explains the observation. As a developed global optimization algorithm, the multi-scale real-coded hybrid GA is applied to a pre-stack seismic waveform inverse problem for the best-fitting 1D earth models at different resolutions. Performance of this new algorithm in terms of computational efficiency, and estimation accuracy, as well as the capability of multi-scaling in addressing the model

parameterization issue, is evaluated by comparing the inversion results against those from a conventional single-scale GA.

The purpose of this application is to estimate 1D earth model parameters (P-wave velocity, S-wave velocity, density) from the pre-stack seismic gathers in (x,t) domain. It is noteworthy here that the layer thickness or two-way travel time (twt) of the 1D model is no longer a free parameter as in the case of using multi-scaling technique. Real well logs of V_p, V_s and density ρ from the Gulf of Mexico are assumed to represent the actual earth model and are used to calculate seismograms that are taken as observed data for inversion. A zero-phase Ricker wavelet with a peak frequency of 35Hz is used to approximate the source wavelet. Numerical simulation of pre-stack seismic data was performed using a reflectivity method (Kennett, 1983). This method generally computes the full-wave response of a stack of horizontal layers including all converted waves and propagation modes. More specifically, a simplified version of the reflectivity method was used to efficiently compute synthetic seismograms for P-primaries only in the offset-time domain. Considering the incidence angle or offset dependence of the reflection coefficients, this forward modeling problem is a nonlinear computationally intensive process. The measure of goodness of fit between the synthetic and observed data is evaluated using the following cross-correlation function

$$C(\mathbf{m}) = \frac{\mathbf{d}_{\mathbf{obs}} \otimes \mathrm{g}(\mathbf{m})}{\left(\mathbf{d}_{\mathbf{obs}} \otimes \mathbf{d}_{\mathbf{obs}}\right)^{1/2}\left(\mathrm{g}(\mathbf{m}) \otimes \mathrm{g}(\mathbf{m})\right)^{1/2}}, \qquad (2.4)$$

where $\otimes$ represents correlation; $\mathbf{d}_{\mathbf{obs}}$ represents observed data; g represents forward operator. In seismic waveform inversion, the low-frequency trends of well logs may be used as prior information to construct the starting models and define search bounds to narrow the model space. Therefore, besides the fit between synthetic and observed data, three normalized terms can also be incorporated into the fitness/likelihood function to represent additional constraints from well logs on V_p, V_s and ρ respectively as given in equation (3.4)

$$fit(\mathbf{m}) = W_1 C(\mathbf{m}) + W_2\left(1 - \sum_{i=1}^{N} \frac{\left|\mathbf{P}^i_{\text{mod}} - \mathbf{P}^i_{pri}\right|}{\Delta P^i}\right) + W_3\left(1 - \sum_{i=1}^{N} \frac{\left|\mathbf{S}^i_{\text{mod}} - \mathbf{S}^i_{pri}\right|}{\Delta S^i}\right) + W_4\left(1 - \sum_{i=1}^{N} \frac{\left|\mathbf{D}^i_{\text{mod}} - \mathbf{D}^i_{pri}\right|}{\Delta D^i}\right), \quad (2.5)$$

where $\Delta P^i, \Delta S^i$, and ΔD^i are search intervals for V_p, V_s, and ρ respectively; $\mathbf{X}^i_{\text{mod}}$ and $\mathbf{X}^i_{pri}$ are respectively the current model value and the prior low-frequency trend for each parameter. Choice of weights is a more or less subjective task. Different weights are assigned to the four terms to reflect different levels of confidence in them. Of course, the last three weights can be set to zeros if no additional smoothing terms are incorporated. In this study, preliminary tests with different combinations of weights were run to find the best set of weights. The pre-determined set of weights is then used in equation (2.5) for this example.

To utilize multi-scaling technique, the 1D model is parameterized into 4 different scales of 10, 20, 40 and 80 layers, respectively (Figure 2.4). The finest scale has 80 layers and corresponds to the full-scale parameterization relative to the true model represented by the well logs, i.e., it is also of 80 sample points and the layer thickness is fixed as 4*ms*.

Low frequency trends of well logs are used as prior information to define the parameter search bounds. V_p, V_s, and ρ are limited within $\pm 0.5 km/s$, $\pm 0.4 km/s$ and $\pm 0.3 g/cc$ around the corresponding low frequency trends. The starting models are randomly chosen within these bounds. The multi-scale hybrid GA as outlined in Figure 2.3 can be easily implemented on the identical (*x,t*) pre-stack seismic gathers to derive the best-fitting 1D earth models of different scales. Each implementation of the new algorithm consists of 5 independent runs starting with different random seeds, and each run goes through 1500 generations of updates on a total of 28 models at 4 different scales. Thus, each scale consists of seven models in a population. Model parameters are mutated with a probability of 0.2, and the uniform crossover is performed on the paired models with a probability of 0.7 through the runs and generations. Of course, crossover is performed only on the φ parts that are of common length. For this 1D model, multi-scaling can be implemented in a straightforward way. For example, if the paired two models are at different scales and have 20 and 40 layers respectively, then each layer on the coarse scale corresponds to two layers on the fine scale. When a crossover step is reached, the fine-scale model proposes a coarse-scale realization by taking the standard average of each of the two fine layers that correspond to a single coarse layer in its current realization. In the meantime, associated remainders between the original current average realizations are stored in λ part of each of the two fine layers. Therefore, the model realization can be composed back by directly summing the average and the corresponding remainder afterward.

The inverted optimal V_p, V_s and density structures of the four different scales are respectively shown in Figure 2.5, Figure 2.6 and Figure 2.7 based on the observed gathers in Figure 2.8, which are compared against the true structures represented by the well logs. The observed data have a lateral coverage of 600 m and consist of 31 offsets with a 20 m interval (Figure 2.8). It is very difficult to choose an appropriate number of layers in the model before actually inverting the observed data. Therefore, instead of assuming a certain number of layers, we use multi-scaling to avoid layer definition. This optimization process was implemented on a PC, and the total CPU time taken through the 1500 generations for this multi-scale case is about 23138 seconds. As shown, all the parameterized models of different layers basically converge to the actual structure and the fine scales lead to better fits. Compared to results of V_p and density, the result of V_s is relatively poor, which indicates that using Poisson's ratio as an alternative model variable to V_s may be a better choice. Figure 2.8 also shows the computed seismograms corresponding to the derived optimal models at the four scales. Due to the rough parameterization, data fits between the observed and inverted gathers for scales 1 and 2 are not necessarily good. In contrast, the inverted seismograms for scales 3 and 4 fit the observed gathers very well, fits that are confirmed by the residual seismograms shown in Figure 2.9. This is also consistent with the observations in the derived best fitting model structures. Figure 2.10 shows the histories of fitness values through 1500 generations and 23138 seconds for all four scales. As expected, after only about 800 generations, the coarse scales (#1, #2) converge rapidly to relatively low fitness values of about 0.72 and

0.87, respectively. In contrast, the fine scales (#3, #4) both converge relatively slower but to a better value as high as about 0.99.

For comparison, the conventional single-scale GA was also implemented four times on a PC for the four scales of 10, 20, 40 and 80 layers individually, and 1500 generations of updates are also run on a total of 28 single-scale models for each of the four scales. Every implementation of the single-scale GA also consists of five different runs, which start with different random seeds. The total CPU times taken through the 1500 generations are 6051, 11823, 24352 and 49113 seconds for the four scales, respectively. As an example, Figure 2.11 summarizes the optimization results of V_p from the four individual implementations of the conventional single-scale GA. This is separately done on the four different parameterization cases. Similarly, all the parameterized models finally converge to the actual model, and the finer-scale (#3, #4) models lead to better fits while they take longer time periods. The corresponding fitness histories of the four single-scale runs are illustrated in Figure 2.12. Similarly, the coarse scales (#1, #2) converge to relatively lower fitness values of about 0.73 and 0.85, respectively, while the fine scales (#3, #4) arrive at better values of about 0.98 after long time updates. Compared to the results from the multi-scale GA, all four single-scale fitness values in Figure 2.12 converge much more slowly until after as long as about 1200 generations, or even more. Additionally, in terms of the CPU cost, the finest #4 scale in Figure 2.12 needs at least twice the total time of the multi-scale case (Figure 2.10) in order to converge to the similarly good fitness value of 0.99.

This comparison shows that, by using multi-scaling and exchanging information between different scales, convergence of the fine scales is accelerated to an excellent fitness value of 0.99 that a conventional single-scale GA can only obtains after a much longer time period. Therefore, the new multi-scale hybrid GA is demonstrated to have better performance in terms of efficiency and accuracy compared to a single-scale stand-alone GA, in that the additionally incorporated coarse scales of faster mixing property facilitate better exploitation of the model space on the fine scale, and this leads to an accurate parameter estimation.

In addition, this example also shows promise of using multi-scaling to overcome the model parameterization problem by circumventing layer definition before the data are used. In other words, we do not have to identify layer boundaries before we actually invert seismic data. Instead, the best layer definition can be determined based on the multiple inverted models at different scales. For example, in this application, the final optimal fitness values for scale #3 and #4 are both about 0.99 (Figure 2.10), but according to “Occam’s razor” paraphrased as "when you have two competing theories which make exactly the same predictions, the one that is simpler is the better.", the inverted model and the corresponding parameterization for scale #3 may be preferred over scale #4 because scale #3 has fewer layers. In this way, both the best model parameterization and the corresponding best-fitting model can be evaluated and derived simultaneously. Therefore, “parameter estimation” sometimes may do the job of “model selection” by incorporating the multi-scaling technique, at least for this nonlinear 1D pre-stack seismic waveform inversion with 200-300 free parameters.

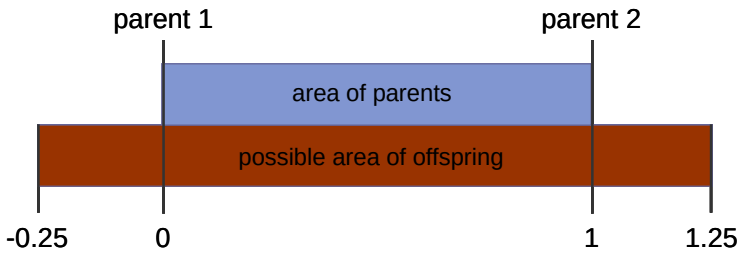

Figure 2.1: Possible area of the range of the offspring variable for the given parent model variables. *d* is set up as 0.25 for example.

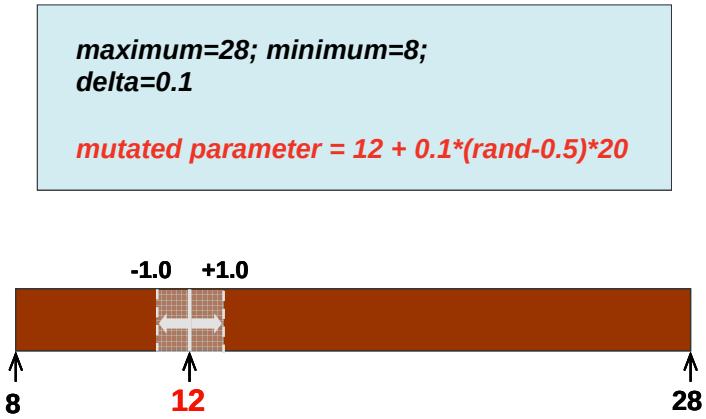

Figure 2.2: Illustration of the mutation on a variable, which is equivalent to adding a white noise on the original variable. The possible area of the mutated variable is indicated in gray.

Multi-scale hybrid GA

1) For generation 1 (t=1), randomly select an ensemble of N models for total R scales, so each scale has N/R models. For scale i

   $\mathbf{m}_{t,j}^{(i)}$, j=1, … , N/R; i=1, … , R; t=1, … , NG.

2) Track temperature T_t along a pre-defined cooling schedule.

3) Evaluate the fitness function $f_{t,j}^{(i)}$ for each model of each scale:

   if t=NG maximum generations → STOP

   let $F_t^{(i)} = \max\left(f_{t,j}^{(i)}\right)$ and $\overline{F}^{(i)} = \sum_{k=1}^{N/R} f_{t,k}^{(i)} / (N/R)$ i=1, …, R; j=1, …, N/R;

   if $F_t^{(i)} >$ some limit → STOP

   if $\overline{F}^{(i)} / F_t^{(i)} >$ some limit → STOP

4) For t>1 (all but the first generation), compare current fitness function $f_{t,j}^{(i)}$ to the last generation $f_{t-1,j}^{(i)}$ for each model and each scale

   if $f_{t,j}^{(i)} > f_{t-1,j}^{(i)}$ → Accept $\mathbf{m}_{t,j}^{(i)}$

   if $f_{t,j}^{(i)} < f_{t-1,j}^{(i)}$, then evaluate $p_{up}^{(i)} = \exp\left(\dfrac{f_{t,j}^{(i)} - f_{t-1,j}^{(i)}}{T_t}\right)$

   if $p_{up}^{(i)} > rand[0,1]$ then

   $\mathbf{m}_{t,j}^{(i)} = \mathbf{m}_{t-1,j}^{(i)}$, and $f_{t,j}^{(i)} = f_{t-1,j}^{(i)}$

   else

   Accept $\mathbf{m}_{t,j}^{(i)}$ and $f_{t,j}^{(i)}$

   end if

5) Select N/R models for each scale for reproduction based on probability:

$$p_{t,j}^{(i)} = \frac{\exp(f_{t,j}^{(i)} / T_t)}{\sum_{k=1}^{N/R} \exp(f_{t,k}^{(i)} / T_t)}$$

6) Mixing and pairing all the selected total N models.

7) Decompose the models of each pair into two parts:

$$\mathbf{m}^{(i)} = f^{(i)}(\boldsymbol{\varphi}^{(i)}, \boldsymbol{\lambda}^{(i)})$$

8) For each pair, perform crossover on the $\boldsymbol{\varphi}$ part based on a pre-specified probability.

9) Compose the $\boldsymbol{\varphi}$ and $\boldsymbol{\lambda}$ parts into $\mathbf{m}_{t,j}^{(i)}$.

10) Perform mutation on all the N models based on a pre-specified probability.

Figure 2.3: Pseudo code of the multi-scale real-coded hybrid GA.

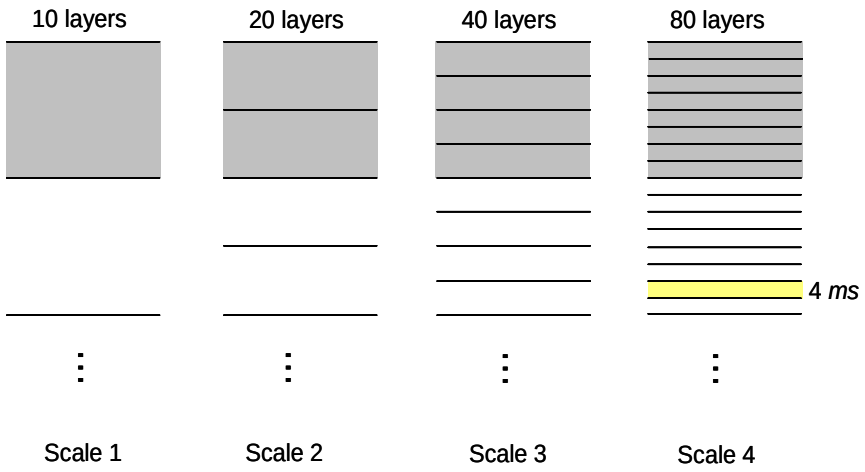

Figure 2.4: 1D earth model and its parameterizations in 4 different scales. The finest scale has 80 layers and the layer thickness is fixed as 4*ms*, which corresponds to the case of full-scale parameterization.

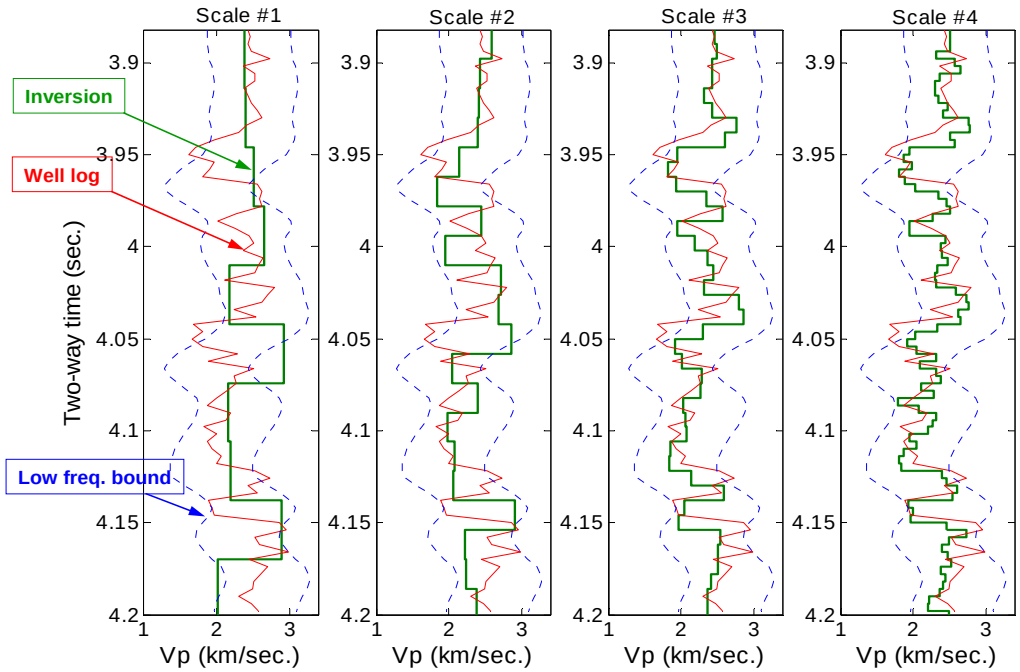

Figure 2.5: Inverted optimal *Vp* structures using multi-scale GA for the 4 scales, which are compared against well logs. As shown, over-parameterized models converge to the actual model and finer cases lead to better fits. This also shows the promise of multi-scaling to overcome parameterization issue in seismic waveform inversion.

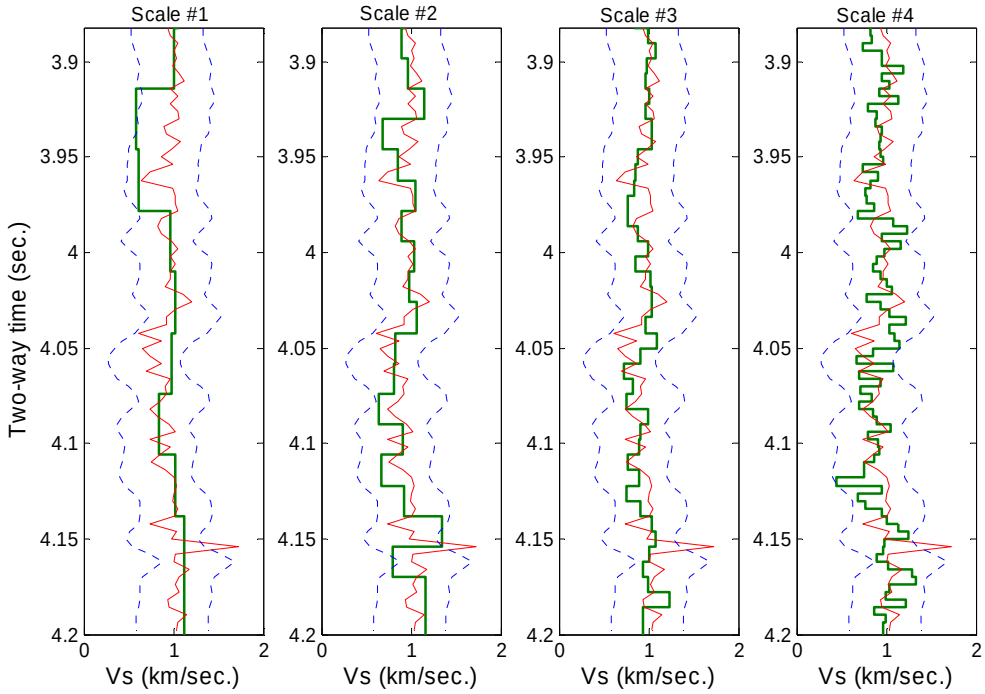

Figure 2.6: Inverted optimal *Vs* structures using multi-scale GA for the 4 scales, which are compared against well logs. As shown, the derived structures have relatively bigger uncertainties and do not fit the actual structure very well. This indicates that using Poisson's ratio as an alternative unknown to *Vs* may be a better choice.

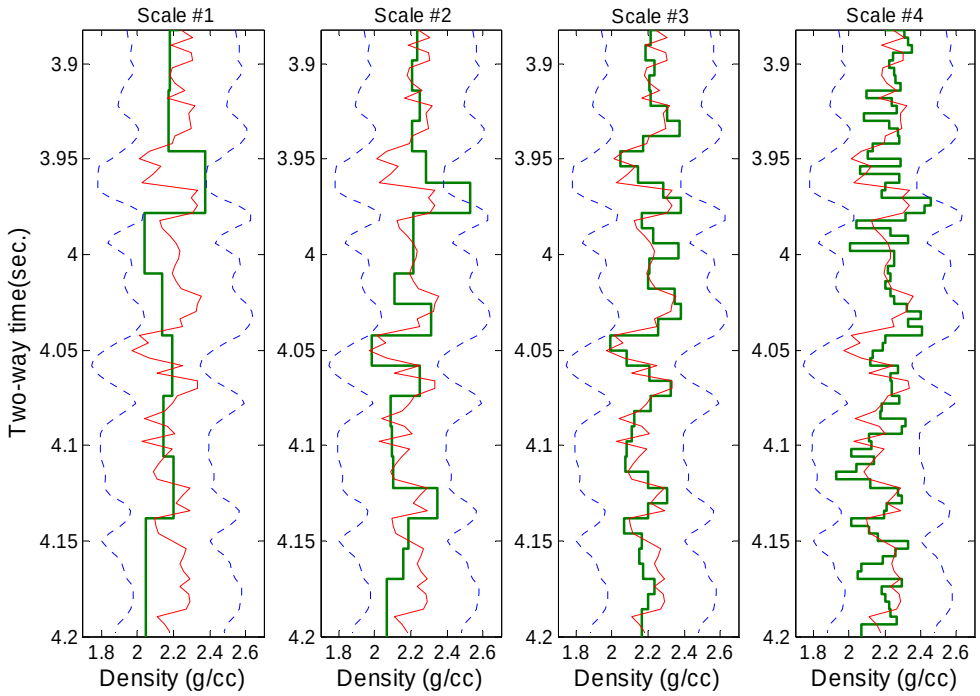

Figure 2.7: Inverted optimal density structures using multi-scale GA for the 4 scales, which are compared against well logs. As shown, the scale #3 leads to the best fit.

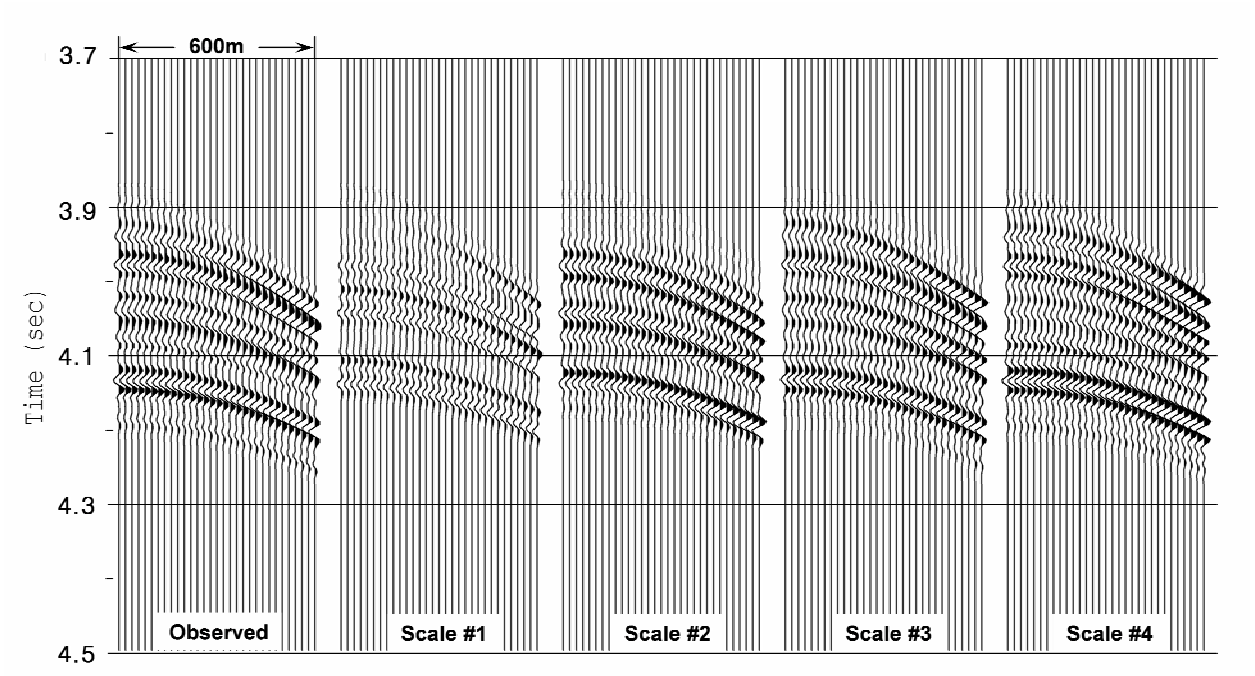

Figure 2.8: Comparison between true and inverted seismograms for the 4 scales, which correspond to the final optimal structures. Coarse scales result in bad data fits but speed up the convergence while fine scales lead to good data fits but take a very long time to converge.

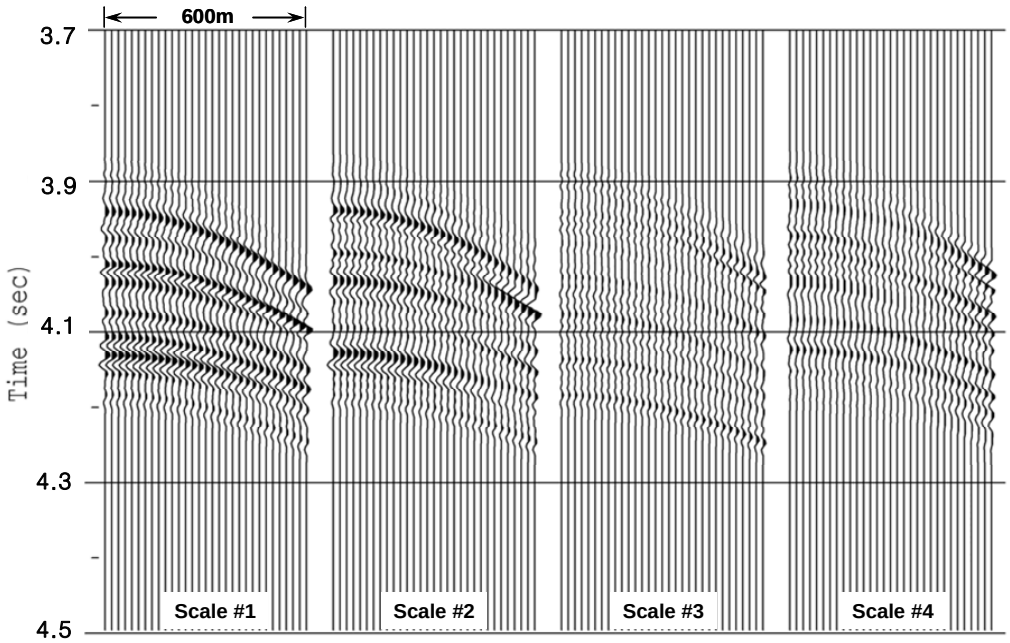

Figure 2.9: Residues between the observed gather and the synthetic gathers for the four scales..

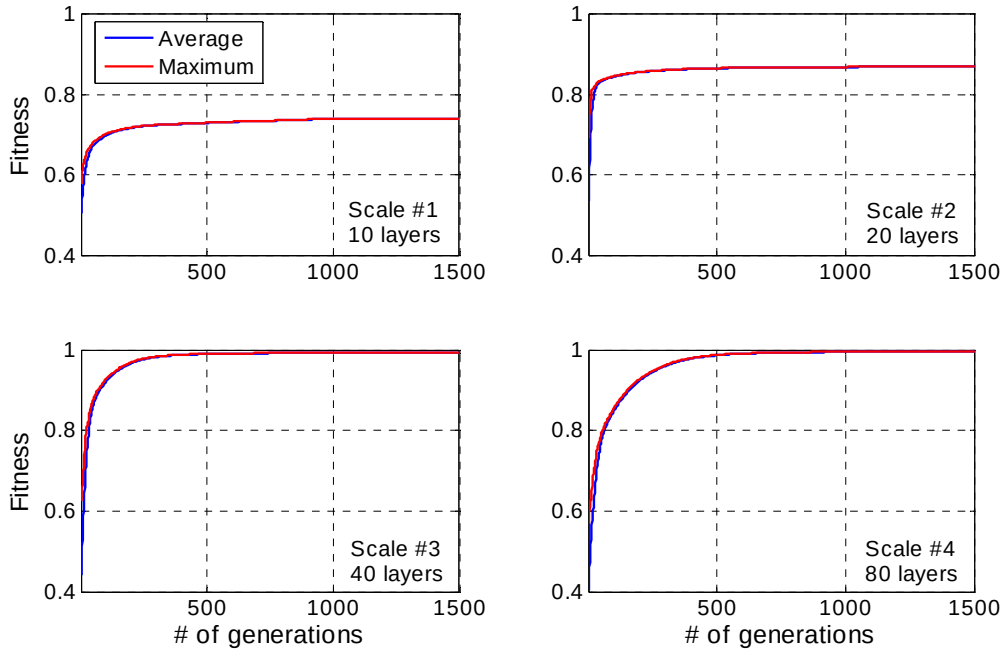

Figure 2.10: Fitness histories produced by running the multi-scale GA. As shown, all of them converge rapidly after about 500 generations; coarse scales (#1, #2) reach fitness values of about 0.72 and 0.87 respectively; fine scales (#3, #4) converge to a value of about 0.99.

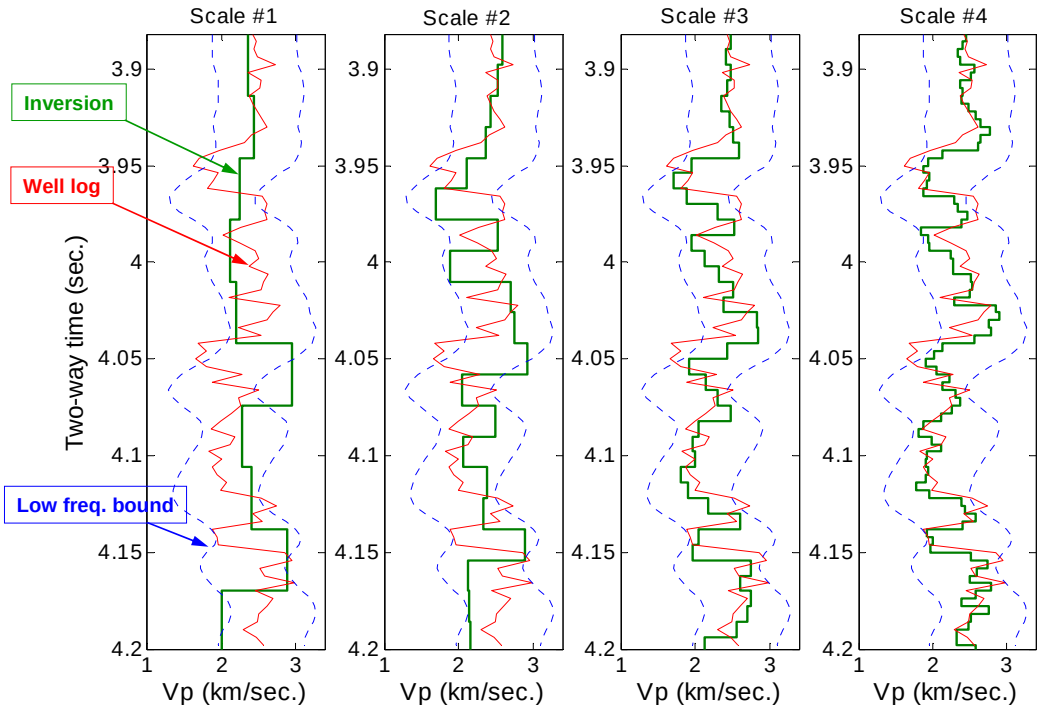

Figure 2.11: Inverted optimal *Vp* structures by running the conventional single-scale GA separately on each of the 4 scales, which are compared against well logs. As shown, over-parameterized models converge to the actual model and finer cases lead to better fits after long time updates.

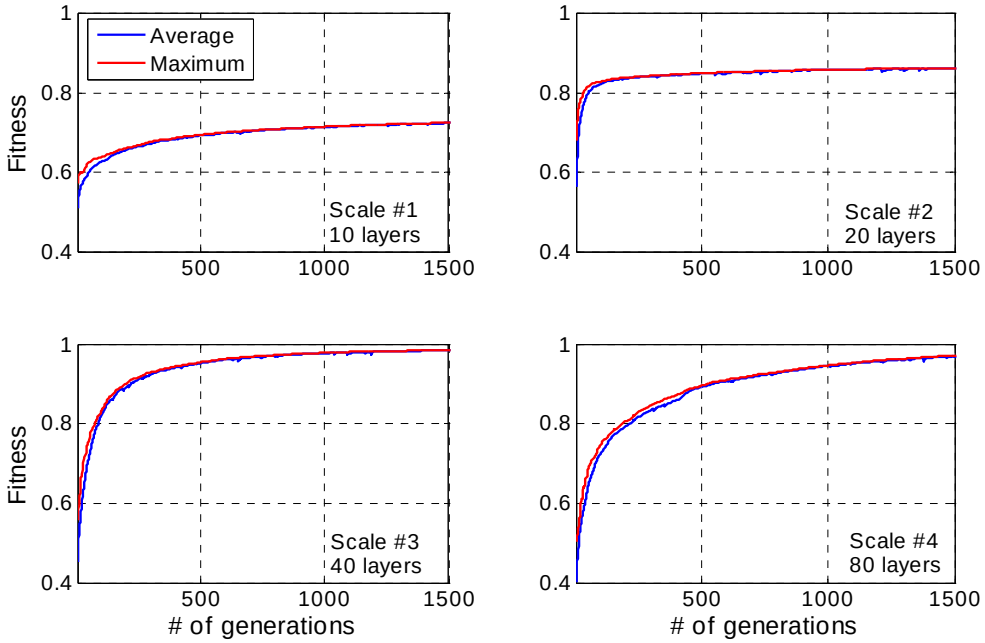

Figure 2.12: Fitness histories produced by running the conventional single-scale GA separately on each of the 4 scales. As shown, they converge slowly until after about 1200 generations; coarse scales (#1, #2) reach fitness values of about 0.72 and 0.87 respectively; fine scales (#3, #4) arrive at a value of about 0.99 after long time updates.

Chapter 3

Multi-scale GA based MCMC algorithm

Besides the "exploitation of the model space," another approach to inverse problems may be termed "exploration of the model space", which essentially casts the inverse problem in a Bayesian framework as introduced in chapter 1. This involves applying a MCMC algorithm to reconstruct the PPD surface that describes all the likely models consistent with the observations as well as the corresponding likelihoods of those models. It therefore provides a theoretically correct way to quantify uncertainty associated with parameter estimation. The use of MCMC methods for exploring the posterior distribution has gained more popularity compared to alternative approaches (Floris *et al.*, 1999). In practice, however, MCMC can be computationally expensive, particularly in complicated inverse problems. In addition, the use of posterior estimation to quantify uncertainty can still be a computationally demanding task, especially for problems with a time-consuming forward simulator.

Typically, solving inverse problems requires numerous runs of a time-consuming forward simulation. In many realistic applications with extensive computational demands, this can be an unfavorable limitation that restricts the number of forward simulations, which makes an MCMC-based approach difficult or even impossible. One way around this difficulty is to run the simulator on a coarser input model that has fewer free parameters, which speeds up the forward simulation but yields less accurate estimation.

In this chapter, the developed multi-scale hybrid GA is further adapted to facilitate MCMC analysis, in which multiple Markov chains of different scales are run simultaneously in parallel, with each chain acting analogously to a single model vector in a GA. At each generation, information from different chains of different parameterizations is exchanged according to the genetic operators by using up-scaling or down-scaling as appropriate to propose intelligent realizations. This exchange facilitates the exploration of the model space of interest (Hong and Sen, 2007; Hong and Sen, 2008a). For every single chain, the operator of mutation is used as a within-chain update exactly the same way as in a basic GA. Selection and crossover are performed across a pair of chains to trade information and develop intelligent proposals. These proposals are accepted or rejected based on the Metropolis-Hastings rule.

This new multi-scale GA-based MCMC algorithm is similar to the approaches proposed by other researchers (Holmes and Mallick, 1998; Liang and Wong, 2001, Holloman, 2002; Higdon *et al.*, 2002). In this chapter, the feasibility of this new MCMC approach is first justified using an analytical example. Its performance in PPD estimation and uncertainty analysis is further evaluated by applying it to a nonlinear pre-stack seismic waveform inversion problem. In this application, multi-scaling is particularly attractive in addressing the model parameterization issue, especially for the seismic waveform inverse problem.

3.1 Multi-scale GA based MCMC

3.1.1 Multi-scale coupled MCMC

For given data at a single scale or multiple scales, the PPDs can be defined at different resolutions corresponding to different model parameterization schemes. The posteriors at different scales are related but conditionally independent so that they can be sampled individually using separate MCMC chains. However, separate sampling does not allow information exchange across different resolutions for intelligent proposals.

In order to take advantage of both the faster mixing speed of the coarse-scale chains and the greater detail of the fine-scale chains, information has to be traded between the scales. Geyer (1991) proposed a Metropolis-coupled MCMC with the idea of swapping information between realizations of conditionally independent posteriors to improve mixing at a single scale. Higdon *et al.* (2002) further presented a methodology of multi-scale coupled MCMC by additionally incorporating a coarsened version of the problem. Realizations from two different scales are exchanged via the swapping step. As in chapter 2, the model vector can be parameterized in different scales. To build multiple MCMC chains running at different scales, a model has to be discretized into multiple scales accordingly. Thus, each chain explores a portion of the overall joint posterior because the posteriors at different scales are conditionally independent. And as before, the model vector at a certain scale can be decomposed into two parts, $\mathbf{m}^{(i)} = f^{(i)}\left(\boldsymbol{\varphi}^{(i)}, \boldsymbol{\lambda}^{(i)}\right)$, where $\boldsymbol{\varphi}$ is of common length and physical interpretation across scales. In the Bayesian setting, the relationship between $\mathbf{m}$, $\boldsymbol{\varphi}$ and $\boldsymbol{\lambda}$ is not necessarily

deterministic, and it could be probabilistic. For example, PPDs at two different scales are illustrated as follows where superscripts in parentheses represent different scales.

fine: $$p^{(1)}(\mathbf{m}^{(1)} \mid \mathbf{d}_{\mathbf{obs}}{}^{(1)}, \mathbf{I}^{(1)}) \propto L^{(1)}(\mathbf{d}_{\mathbf{obs}}{}^{(1)} \mid \mathbf{m}^{(1)}, \mathbf{I}^{(1)}) p^{(1)}(\mathbf{m}^{(1)} \mid \mathbf{I}^{(1)}) \quad (3.1)$$

coarse: $$p^{(2)}(\mathbf{m}^{(2)} \mid \mathbf{d}_{\mathbf{obs}}{}^{(2)}, \mathbf{I}^{(2)}) \propto L^{(2)}(\mathbf{d}_{\mathbf{obs}}{}^{(2)} \mid \mathbf{m}^{(2)}, \mathbf{I}^{(2)}) p^{(2)}(\mathbf{m}^{(2)} \mid I^{(2)}) \quad (3.2)$$

The resulting PPD for scale *i* is described in equation (3.3), and the overall joint PPD is shown in equation (3.4).

$$p^{(i)}(\mathbf{m}^{(i)}, \boldsymbol{\varphi}^{(i)}, \boldsymbol{\lambda}^{(i)} \mid \mathbf{d}_{\mathbf{obs}}{}^{(i)}, \mathbf{I}^{(i)}) \propto L^{(i)}(\mathbf{d}_{\mathbf{obs}}{}^{(i)} \mid \mathbf{m}^{(i)}, \mathbf{I}^{(i)}) \cdot p^{(i)}(\mathbf{m}^{(i)} \mid \boldsymbol{\varphi}^{(i)}, \boldsymbol{\lambda}^{(i)}, \mathbf{I}^{(i)}) \cdot p^{(i)}(\boldsymbol{\varphi}^{(i)}, \boldsymbol{\lambda}^{(i)} \mid \mathbf{I}^{(i)}) \quad (3.3)$$

$$p(\mathbf{m}, \boldsymbol{\varphi}, \boldsymbol{\lambda} \mid \mathbf{d}_{\mathbf{obs}}, \mathbf{I}) = \prod_{i=1}^{R} p^{(i)}(\mathbf{m}^{(i)}, \boldsymbol{\varphi}^{(i)}, \boldsymbol{\lambda}^{(i)} \mid \mathbf{d}_{\mathbf{obs}}{}^{(i)}, \mathbf{I}^{(i)}) \quad (3.4)$$

Although the fine scale provides more model details and is of primary interest, benefits from the introduction of coarse scale are obvious. The coarse scale yields a more tractable PPD and a faster forward simulation, allowing more MCMC updates per unit time. The basic idea of this multi-scale coupled MCMC is: Instead of running two separate MCMC chains for two separate PPDs, one on the fine space $p^{(1)}(\mathbf{m}^{(1)} \mid \mathbf{d}_{\mathbf{obs}}{}^{(1)}, \mathbf{I}^{(1)})$ and the other on the coarse space $p^{(2)}(\mathbf{m}^{(2)} \mid \mathbf{d}_{\mathbf{obs}}{}^{(2)}, \mathbf{I}^{(2)})$, run a single coupled chain in the product space. This coupled chain has a stationary distribution as in equation (3.5) so that if we take only the fine-scale realizations from this chain, the PPD $p^{(1)}(\mathbf{m}^{(1)} \mid \mathbf{d}_{\mathbf{obs}}{}^{(1)}, \mathbf{I}^{(1)})$ can be estimated accordingly (Higdon *et al.,* 2002).

$$p(\mathbf{m} \mid \mathbf{d}_{\mathbf{obs}}, \mathbf{I}) = p^{(1)}(\mathbf{m}^{(1)} \mid \mathbf{d}_{\mathbf{obs}}{}^{(1)}, \mathbf{I}^{(1)}) \times p^{(2)}(\mathbf{m}^{(2)} \mid \mathbf{d}_{\mathbf{obs}}{}^{(2)}, \mathbf{I}^{(2)}) \quad (3.5)$$

Periodically, the two separate MCMC chains at the two different PPDs attempt to fully swap information between their current realizations. To exchange information between the two different scales, a model vector has to be decomposed into two parts $\mathbf{m}^{(i)} = f^{(i)}\left(\boldsymbol{\varphi}^{(i)}, \boldsymbol{\lambda}^{(i)}\right)$, and the swapping operation is only applied on the $\boldsymbol{\varphi}$ parts that have the common length. The Markov chains consist of two types of updates: one is the ordinary within-chain update, and the other is the swapping update across different chains. Both updates are accepted or rejected according to the Metropolis-Hastings rule. A swapping process for two MCMC chains is schematically shown as

$$\begin{array}{l} Z^{(1)}(\boldsymbol{\varphi}^{1}_{(1)}, \boldsymbol{\lambda}^{1}_{(1)}) \overset{\text{MCMC}}{\rightarrow} Z^{(1)}(\boldsymbol{\varphi}^{1}_{(2)}, \boldsymbol{\lambda}^{1}_{(2)}) \\ Z^{(2)}(\boldsymbol{\varphi}^{2}_{(1)}, \boldsymbol{\lambda}^{2}_{(1)}) \overset{\text{MCMC}}{\rightarrow} Z^{(2)}(\boldsymbol{\varphi}^{2}_{(2)}, \boldsymbol{\lambda}^{2}_{(2)}) \end{array} \xrightarrow{\text{Swap attempt}} \begin{array}{l} Z^{(1)}(\boldsymbol{\varphi}^{1*}_{(3)}, \boldsymbol{\lambda}^{1}_{(3)}) \overset{\text{MCMC}}{\rightarrow} Z^{(1)}(\boldsymbol{\varphi}^{1*}_{(4)}, \boldsymbol{\lambda}^{1*}_{(4)}) \\ Z^{(2)}(\boldsymbol{\varphi}^{2*}_{(3)}, \boldsymbol{\lambda}^{2}_{(3)}) \overset{\text{MCMC}}{\rightarrow} Z^{(2)}(\boldsymbol{\varphi}^{2*}_{(4)}, \boldsymbol{\lambda}^{2*}_{(4)}) \end{array}.$$

At time step 2, the two MCMC chains attempt to fully swap their current values of the $\boldsymbol{\varphi}$ parts. The starred realizations denote that they have different values, depending on the success or failure of the swap attempt. If the swap is accepted, $\boldsymbol{\varphi}^{1*}_{(3)} = \boldsymbol{\varphi}^{2}_{(2)}$ and $\boldsymbol{\varphi}^{2*}_{(3)} = \boldsymbol{\varphi}^{1}_{(2)}$, otherwise $\boldsymbol{\varphi}^{1*}_{(3)} = \boldsymbol{\varphi}^{1}_{(2)}$ and $\boldsymbol{\varphi}^{2*}_{(3)} = \boldsymbol{\varphi}^{2}_{(2)}$.

3.1.2 Crossover swapping

The full swaps trade entire $\boldsymbol{\varphi}$ vectors between chains so that $\boldsymbol{\varphi}$ vectors that were associated with one resolution of **m** are then associated with a different resolution of **m**. Despite the advantages offered by the full swap, parts of the space may still be left

unexplored due to the slow mixing of even the coarsest scale. Recently, it has been shown that GA can be viewed as a Markov chain as the conditional dependence of each population on its predecessor is completely described by its dependence on the parent (immediate predecessor) population (Davis, 1991). Further, Davis and Principe (1993) demonstrated that when the mutation operator is used, the Markov chain of GA is irreducible and consequently has a unique stationary distribution. Thus, the mutation is a control parameter analogous to temperature in SA.

Based on the work of Davis (1991) and Davis and Principe (1993), Suzuki (1998) made some extension and gave a more rigorous theoretical basis on the Markov chain property of GA by additionally considering diminishing genetic operators. Suzuki (1998) also proved that the GA converges to a stationary distribution focusing on the uniform population of optimal solutions, which forms a sound basis to use GA as a safe sampling tool in addition to a maximization method. Elements of GA have been incorporated into the MCMC schemes, often called Evolutionary Monte Carlo, to make intelligent Metropolis-style proposals in high-dimensional spaces (Holmes and Mallick 1998; Liang and Wong 2000, 2001). The multi-scale GA introduced in chapter 2 has also been adapted into a new multi-scale GA-based MCMC.

In this new algorithm, multiple MCMC chains at different scales are run in parallel with each chain acting analogously to a single-scale GA run. In order to gain benefits of both the fine scale and the coarse scale, information must be exchanged between different scales of the joint PPD. Instead of the full swaps of $\boldsymbol{\varphi}$ vectors, another type swap, the crossover swap, is implemented. This is shown in the diagram below in

which three chains of different scales as an example are illustrated, and the first two chains are selected for trading.

$$
\begin{array}{lllll}
Z^{(1)}(\boldsymbol{\varphi}^{1}_{(1)},\boldsymbol{\lambda}^{1}_{(1)}) \overset{\text{MCMC}}{\rightarrow} Z^{(1)}(\boldsymbol{\varphi}^{1}_{(2)},\boldsymbol{\lambda}^{1}_{(2)}) & \overset{\text{Crossover}}{\longrightarrow} & Z^{(1)}(\boldsymbol{\varphi}^{1*}_{(3)},\boldsymbol{\lambda}^{1}_{(3)}) \overset{\text{MCMC}}{\rightarrow} Z^{(1)}(\boldsymbol{\varphi}^{1*}_{(4)},\boldsymbol{\lambda}^{1*}_{(4)}) \\
Z^{(2)}(\boldsymbol{\varphi}^{2}_{(1)},\boldsymbol{\lambda}^{2}_{(1)}) \overset{\text{MCMC}}{\rightarrow} Z^{(2)}(\boldsymbol{\varphi}^{2}_{(2)},\boldsymbol{\lambda}^{2}_{(2)}) & & Z^{(2)}(\boldsymbol{\varphi}^{2*}_{(3)},\boldsymbol{\lambda}^{2}_{(3)}) \overset{\text{MCMC}}{\rightarrow} Z^{(2)}(\boldsymbol{\varphi}^{2*}_{(4)},\boldsymbol{\lambda}^{2*}_{(4)}) \\
Z^{(3)}(\boldsymbol{\varphi}^{3}_{(1)},\boldsymbol{\lambda}^{3}_{(1)}) \overset{\text{MCMC}}{\rightarrow} Z^{(3)}(\boldsymbol{\varphi}^{3}_{(2)},\boldsymbol{\lambda}^{3}_{(2)}) & \overset{\text{MCMC}}{\longrightarrow} & Z^{(3)}(\boldsymbol{\varphi}^{3}_{(3)},\boldsymbol{\lambda}^{3}_{(3)}) \overset{\text{MCMC}}{\rightarrow} Z^{(3)}(\boldsymbol{\varphi}^{3}_{(4)},\boldsymbol{\lambda}^{3}_{(4)})
\end{array}
$$

As before, superscripts in the diagram index the scale, and subscripts in parentheses index the number of the current generation. In this scheme, the three chains advance using the mutation operation as the within-chain updates until reaching the trading point. At that point, a pair of chains are selected without replacement based on a probability $p_s(\mathbf{m}^{(i)}_{(t)})$ that is directly proportional to the posterior density of the current realization and can be defined as follows.

$$
p_s(\mathbf{m}^{(i)}_{(t)}) = \frac{L^{(i)}\left(\mathbf{m}^{(i)}_{(t)},\boldsymbol{\varphi}^{(i)}_{(t)},\boldsymbol{\lambda}^{(i)}_{(t)} \mid \mathbf{d}_{\text{obs}}{}^{(i)},\mathbf{I}^{(i)}\right)\cdot p^{(i)}\left(\mathbf{m}^{(i)}_{(t)} \mid \boldsymbol{\varphi}^{(i)}_{(t)},\boldsymbol{\lambda}^{(i)}_{(t)},\mathbf{I}^{(i)}\right)\cdot p^{(i)}\left(\boldsymbol{\varphi}^{(i)}_{(t)},\boldsymbol{\lambda}^{(i)}_{(t)} \mid \mathbf{I}^{(i)}\right)}{\sum_{j=1}^{3} L^{(j)}\left(\mathbf{m}^{(j)}_{(t)},\boldsymbol{\varphi}^{(j)}_{(t)},\boldsymbol{\lambda}^{(j)}_{(t)} \mid \mathbf{d}_{\text{obs}}{}^{(j)},\mathbf{I}^{(j)}\right)\cdot p^{(j)}\left(\mathbf{m}^{(j)}_{(t)} \mid \boldsymbol{\varphi}^{(j)}_{(t)},\boldsymbol{\lambda}^{(j)}_{(t)},\mathbf{I}^{(j)}\right)\cdot p^{(j)}\left(\boldsymbol{\varphi}^{(j)}_{(t)},\boldsymbol{\lambda}^{(j)}_{(t)} \mid \mathbf{I}^{(j)}\right)} \tag{3.6}
$$

Unlike the diagram in section 3.1.1 where the full swap is attempted, current realizations of the selected chains are not fully swapped but are partially traded using a crossover mechanism such as the uniform crossover. In other words, the proposed model values for different scales are unique combinations of the two realizations. Starred realizations denote that they are affected by the crossover operation. Again, crossover is performed only on the $\boldsymbol{\varphi}$ part. Chains of high posterior density are more likely to be

combined to create intelligent proposals, and these proposals may further explore previously unexplored regions of high density. The updated models are either accepted or rejected according to the Metropolis-Hastings norm at a constant temperature. Applications of this new MCMC to an analytical example and to a 1D pre-stack seismic waveform inverse problem are shown in sections 3.2 and 3.3, respectively.

3.2 Analytical example

In this application, the multi-scale GA-based MCMC is applied to sample from a joint PPD $p(X_1, X_2 | \mathbf{D}, \mathbf{I})$ in the model space of two parameters X_1 and X_2, that has a double peak structure. Using this two-dimensional example instead of a higher dimensional one provides a direct way to graphically illustrate the joint PPD and therefore offers a straightforward way to evaluate the performance of the new MCMC approach. As shown in equation (3.7), this PPD is a sum of two bivariate normal distributions of $PDF_1(X_1, X_2)$ and $PDF_2(X_1, X_2)$. The factor of 0.5 ensures that the resulting posterior is normalized to an area of one.

$$p(X_1, X_2 | \mathbf{D}, \mathbf{I}) = 0.5 \cdot \left[PDF_1(X_1, X_2) + PDF_2(X_1, X_2) \right] \qquad (3.7)$$

For a bivariate normal distribution in (X_1, X_2) with a mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and a covariance matrix $\boldsymbol{\sigma}$ as in equation (3.8), the resulting joint *pdf* of X_1 and X_2 is described in equation (3.9).

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \tag{3.8}$$

$$p(X_1, X_2 \,|\, \mathbf{D}, \mathbf{I}) = \frac{1}{2\pi\sqrt{(1-\rho^2)\sigma_{11}\sigma_{22}}} \exp\left[-\frac{Z}{2(1-\rho^2)} \right] \tag{3.9}$$

where

$$Z = \frac{(X_1 - \mu_1)^2}{\sigma_{11}} - \frac{2\rho(X_1 - \mu_1)(X_2 - \mu_2)}{\sqrt{\sigma_{11}\sigma_{22}}} + \frac{(X_2 - \mu_2)^2}{\sigma_{22}}, \tag{3.10}$$

and ρ is the correlation coefficient of X_1 and X_2 as defined in equation (3.11).

$$\rho = cor(X_1, X_2) = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}. \tag{3.11}$$

In this example, the bivariate normal distribution $PDF_1(X_1, X_2)$ has a mean vector of $\boldsymbol{\mu} = (0, 0)$ and a covariance matrix as in equation (3.12), which indicates that X_1 and X_2 are uncorrelated to each other. That is, the correlation coefficient $\rho = 0$; the bivariate normal distribution $PDF_2(X_1, X_2)$ has a mean vector of $\boldsymbol{\mu} = (4, 0)$ and a covariance matrix as in equation (3.13), which indicates that X_1 and X_2 are correlated to each other and the correlation coefficient $\rho = 0.4$.

$$\boldsymbol{\sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.12}$$

$$\boldsymbol{\sigma} = \begin{bmatrix} 2 & 0.8 \\ 0.8 & 2 \end{bmatrix}. \tag{3.13}$$

By summing the $PDF_1(X_1, X_2)$ and $PDF_2(X_1, X_2)$ as described above following the expression (3.7), the resulting joint PPD $p(X_1, X_2|\mathbf{D,I})$ is a bimodal surface with double peaks as shown in Figure 3.1. One peak is located at (0, 0) with a value of about 0.317 and the other secondary peak is located at (4, 0) with a value of about 0.172, which corresponds to the modes of PDF_1 and PDF_2 , respectively. It is this joint probability distribution surface that will be estimated by using the new multi-scale GA-based MCMC described herein. Figure 3.2(a) is a color map of the probability distribution projected on the (X_1, X_2) plane where the locations of peak values can be easily identified. Panel (b) in Figure 3.2 illustrates the same distribution on the (X_1, X_2) plane, but it is represented in contours. Following equation (1.8), the true marginal PPDs for X_1 and X_2 are calculated analytically and shown on panels (c) and (d), respectively. A marginal PPD is a special kind of projection of the joint PPD to a particular parameter axis by integrating out all the other parameters. Note that the marginal PPD for X_1 is a bimodal distribution, whereas it is a uni-variate normal distribution for X_2.

Samples are picked randomly from the (X_1, X_2) space using the new multi-scale GA-based MCMC. These samples are accepted or rejected based on the Metropolis-Hastings rule and with a probability that is directly proportional to the PPD values. Finally, 9000 samples are drawn and shown in Figure 3.3, superimposing on the true PPD contours as in Figure 3.2(b). Note that the samples within the burn-in period are excluded. Theoretically, the number of samples drawn over a small region is proportional to the PPD value of that region. As shown, more samples are drawn over the peak regions, and the distribution of sample points matches the contours of the true PPD very well. Based

on the 9000 sample points, the relative frequency histograms for X_1 and X_2 are calculated to approximate the marginal PPDs of X_1 and X_2, respectively. The histograms are illustrated in Figure 3.4 and compared against the true marginal PPDs. Note that the fit between the estimated marginal PPDs and the true PPDs is fairly good, which confirms that the number of samples over a small region is directly proportional to the PPD value of that small region. Therefore, through the approximate PPDs derived from the 9000 sample points, we are able to correctly determine the most likely parameter values that correspond to the maximum PPD value. We can also estimate the corresponding uncertainty bounds. The most likely values determined for X_1 and X_2 actually correspond to the two peak locations that have bigger probability values. The very good estimation shown in Figure 3.4 corroborates that the new multi-scale GA-based MCMC algorithm is a reliable sampling tool for PPD estimation and uncertainty analysis; therefore, it is safe to use for practical inverse problems in the Bayesian setting.

3.3 Seismic parameter estimation and uncertainty analysis

Recalling the 1D pre-stack seismic waveform inverse problem introduced in chapter 2, the best-fitting earth models at four different scales were obtained by applying the multi-scale real-coded hybrid GA. Here, we further apply the multi-scale GA based MCMC algorithm to fully solve this problem by estimating the posterior distributions as well as quantifying the uncertainties at different scales. Again, the 1D earth model is parameterized into four scales of 10, 20, 40 and 80 layers, respectively, exactly in the same manner as described in chapter 2. Multiple MCMC chains are run simultaneously to

sample from the overall product space, and all the realizations of a specific scale are archived and used to estimate the PPD of that single scale accordingly. In this new algorithm, GA-type mutation acts as the within-chain update, and GA-type selection and crossover act to trade information across a pair of models at different scales for intelligent proposals that are accepted or rejected based on the Metropolis-Hastings rule and with an SA-style update probability.

The choice of the likelihood function relates to the forward process and depends on the distribution of the noise or error in the data (Box and Tiao, 1973; Cary and Chapman, 1988; Sen and Stoffa, 1995). The error can be due to measurements (*e.g.*, instrument errors) or to the use of inexact theory in the prediction of the data (Tarantola, 1987). How to set the likelihood function requires prior knowledge of the error distribution in the data and is a very important issue because it is often very difficult to obtain an estimate of noise statistics.

Based on the Central Limit Theorem (CLT) in statistics, that states "Any quantity that stems from a large number of sub-processes is expected to have a Gaussian distribution", it is safe to assume the most conservative (*i.e.*, greater uncertainty than one would get from choosing a more restricted distribution) Gaussian error distribution in seismic data. The CLT is both remarkable and of great practical value in data analysis. In frequentist statistics, there is uncertainty about the form of the sampling distribution from which the data are drawn. The equivalent problem in a Bayesian analysis is the choice of likelihood function to use. By working with the averages of data points (frequently as few as five points) and invoking the CLT, researchers can use a Gaussian distribution for the

sampling distribution or likelihood function. The CLT also provides a deep understanding of why measurement uncertainties frequently have a Gaussian distribution. This is because the measured quantity is often the result of a large number of effects, i.e., it is some kind of averaged resultant of these effects (random variables in a frequentist context). Since the distribution of the average of random variables tends to be Gaussian, this is often observed from seismic data, which are often the result of a large number of sub-processes.

On the other hand, the Maximum Entropy Principle says that unless there is some additional prior information that justifies the use of some other sampling distribution, then a Gaussian sampling distribution should be used. It makes the fewest assumptions about the information and will lead to the most conservative Gaussian distribution (Gregory, 2005). Assuming Gaussian distribution, the likelihood function at scale *i* takes the form given in equation (3.14) below. Note that $\mathbf{d}_{\mathbf{obs}}{}^{(1)} = \mathbf{d}_{\mathbf{obs}}{}^{(2)} = ... = \mathbf{d}_{\mathbf{obs}}{}^{(R)} = \mathbf{d}_{\mathbf{obs}}$ in this problem.

$$L^{(i)}(\mathbf{d}_{\mathbf{obs}} \,|\, \mathbf{m}^{(i)}, \mathbf{I}^{(i)}) \propto \exp(-\mathrm{E}^{(i)}(\mathbf{m}^{(i)})) \,, \qquad (3.14)$$

where $\mathrm{E}^{(i)}(\mathbf{m}^{(i)})$ is the error function at scale *i* and given as

$$E^{(i)}(\mathbf{m}^{(i)}) = \frac{1}{2}\left(\mathbf{d}_{\mathbf{obs}} - \mathrm{g}^{(i)}(\mathbf{m}^{(i)})\right)^{T} \mathbf{C}_{\mathbf{D}}^{-1} \left(\mathbf{d}_{\mathbf{obs}} - \mathrm{g}^{(i)}(\mathbf{m}^{(i)})\right), \qquad (3.15)$$

where $\mathrm{g}^{(i)}$ is the forward modeling operator at scale *i* which is usually nonlinear; $\mathbf{C}_{\mathbf{D}}$ is called the data covariance matrix. In different cases, $\mathbf{C}_{\mathbf{D}}$ may have different forms. With

identical, independent Gaussians for the errors, the error function is simply the product of multiple Gaussians, one corresponding to each of the data points. This also leads the $\mathbf{C_D}$ to be a diagonal matrix with all the diagonal elements equal to one. If the data points are independent of each other but have different variances, the resulting $\mathbf{C_D}$ is still a diagonal matrix but the diagonal elements are not identical and have different values. In general, however, the errors can have different variances and could be correlated. Therefore, $\mathbf{C_D}$ is usually not a diagonal matrix and the off-diagonal elements are not all zeros.

To initiate Bayesian inference, the prior distribution of model parameters has to be specified. However, in many practical inverse problems, prior information is vague and deciding how to encode this vague information into a probability distribution is controversial. Recall in chapter 2 where the multi-scale hybrid GA is applied for an optimization example, low frequency trends of well logs are used as prior information to define the search bounds. For example, V_p is limited within $\pm 0.5 km/s$ around its low frequency trend. Such prior information about the search bound is often called the independent linear constraint on each parameter. Different distributions can be used to formulate this prior information; for example, the uniform and Jeffrey priors are discussed by Gregory (2005).

How to translate the vague prior information into a probability distribution is controversial and has been a popular topic (Mosegaard and Tarantola, 1995; Kass and Wasserman, 1996; Curtis and Lomax, 2001). Unless a certain form of probability distribution is justified by some specific prior information, a uniform prior distribution may be chosen. This is a common choice when stochastic sampling methods are used for

Bayesian inference problems because pseudo-random uniform sampling is easy to perform. In addition, the uniform distribution helps avoid more biased constraints to be incorporated (Curtis and Lomax, 2001). As a result, the uniform prior distribution at scale i is given as

$$p^{(i)}(\mathbf{m}^{(i)} \mid \mathbf{I}^{(i)}) = \prod_{j=1}^{L} \frac{1}{m_{\max j}^{(i)} - m_{\min j}^{(i)}} \quad , \tag{3.16}$$

where $m_{\max}^{(i)}$ and $m_{\min}^{(i)}$ are the upper and lower bounds of a specific model parameter at scale i; subscript j represents the jth parameter of total L parameters.

Using the uniform prior and Gaussian likelihood functions as described above, the resulting posterior distribution at a single scale i is

$$p^{(i)}(\mathbf{m}^{(i)} \mid \mathbf{d}_{\mathbf{obs}}{}^{(i)}, \mathbf{I}^{(i)}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{d}_{\mathbf{obs}} - \mathrm{g}^{(i)}(\mathbf{m}^{(i)})\right)^{T} \mathbf{C}_{\mathbf{D}}^{-1} \left(\mathbf{d}_{\mathbf{obs}} - \mathrm{g}^{(i)}(\mathbf{m}^{(i)})\right)\right) \cdot \prod_{j=1}^{L} \frac{1}{m_{\max j}^{(i)} - m_{\min j}^{(i)}} . \tag{3.17}$$

Again, the solution models at different scales are assumed to be conditionally independent of each other for the given data, and the overall PPD is the product of the PPDs for the individual scales. For a linear problem, the uniform prior and the Gaussian likelihood function lead to a multivariate Gaussian posterior with a simple topology. As compared, for a nonlinear problem, the uniform prior and the Gaussian likelihood usually result in a highly multi-modal posterior with a very complicated topology of many hills and valleys, which is common in seismic inverse problems. In order to speed up convergence and enhance exploration of this complicated PPD surface, models from different scales are combined to propose intelligent realizations by using the uniform

crossover scheme. Crossover operations are performed on each parameter based on a given probability of 0.7. The proposals are accepted or rejected according to the Metropolis-Hastings rule.

In this implementation, 5 independent runs of the multi-scale GA based MCMC algorithm were independently and sequentially carried out with different random seeds, and each run went through 1500 generations. Seven chains were run at each of the four scales for a total of 28 chains. All the realizations of different scales are tracked and stored separately through the 1500 generations.

For example, Figure 3.5 illustrates the evolution of the realizations of parameter V_p for the 29th layer, which were drawn by the seven chains at scale #3. Even though starting at different points, all of the 7 chains finally converge to the value of about 1.82 *km/s*. These stored samples are also used to approximate the marginal PPDs and the corresponding model covariance matrices for identification of the most likely parameter values as well as for characterization of the uncertainties at every single scale.

For example, the marginal PPDs of V_p for the four scales are computed based on equation (1.8) and shown in Figure 3.6, in which the most likely models are indicated using the white dashed lines and the probabilities of the different values are represented by colors over the entire model space. The most likely models are the models of maximum probabilities, sometimes called *maximum a posteriori* (MAP), and actually correspond to the best-fitting models inferred by the optimization method as shown in Figure 2.5. Therefore, comparison between the MAP and true parameter values can be read directly from Figure 2.5. The estimated marginal PPDs for particular layers around

the two-way travel time of 4.10 seconds and highlighted in gray are shown in Figure 3.7 along with the computed cumulative distribution function (CDF) for each scale. According to equation (1.7), the posterior model correlation matrices were also computed for all the four scales. As an example for the scales #2 and #3, parts of the correlation matrices that correspond to the *TWT* range of 3.98-4.05 seconds are shown in Figure 3.8 to illustrate the inter-dependence between estimated parameters. For scale #2, the 7-10th layers are shown, and each layer has only three parameters of V_p, V_s, and ρ; for scale #3, corresponding 13-20th layers with a total of 8 layers and 24 parameters are shown. Black represents a strong negative correlation whereas white represents a strong positive correlation between parameters. The middle gray stands for uncorrelation between parameters. For example, the correlation plot for scale #2, layer #9 reveals that for this layer, *Vp* is almost uncorrelated with *Vs* while *Vp* has a strong negative correlation with density. Strong positive correlation means if the two parameters are both either increased or decreased, the resulting seismogram will not change; strong negative correlation means if decreasing one parameter while increasing the other or vice versa, the resulting seismogram will not change.

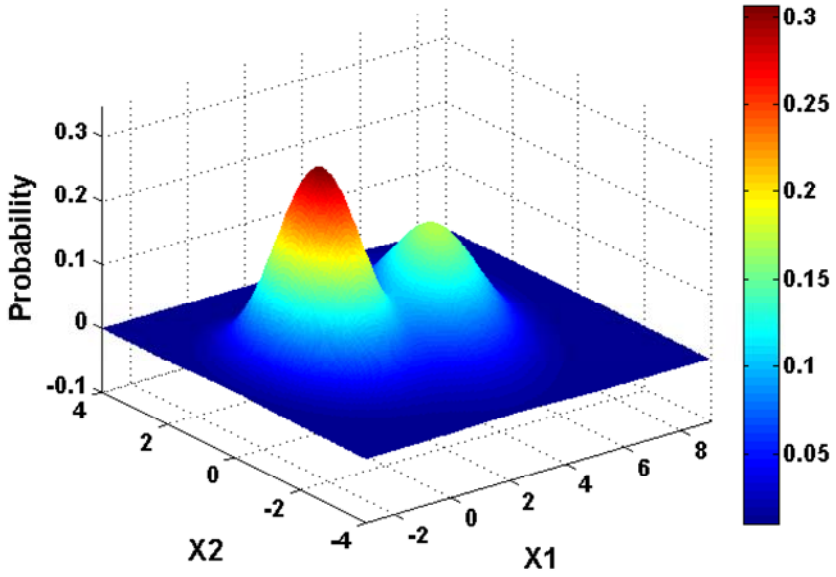

Figure 3.1: The joint PPD surface with two free parameters X_1 and X_2. It has a double-peak structure, one peak location is at (0, 0) and the other is at (4, 0). Color represents the PPD values.

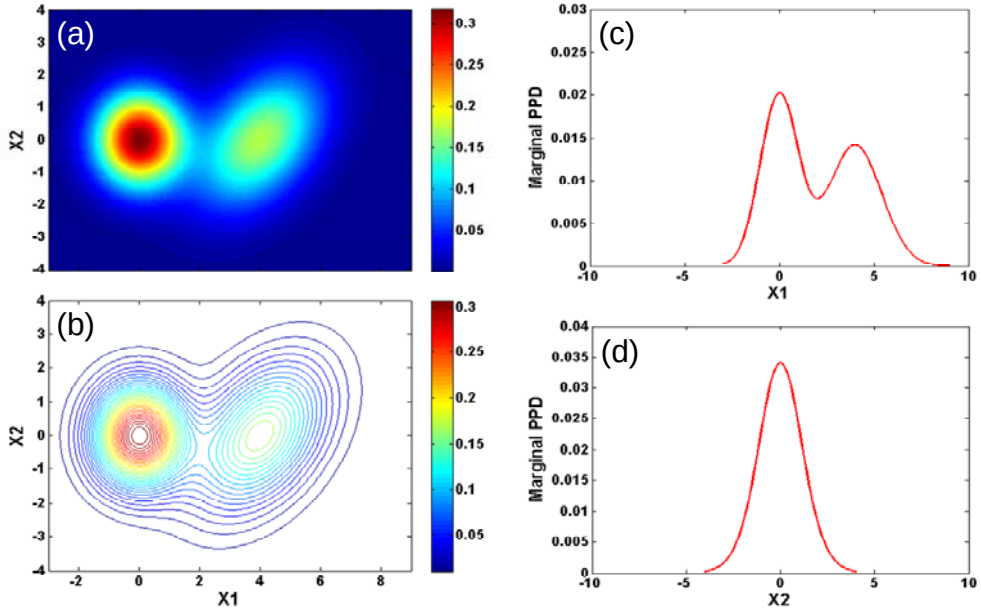

Figure 3.2: (a) The true joint posterior distribution projected on the (X_1, X_2) plane. (b) The contours of the true PPD on the (X_1, X_2) plane. (c) Marginal PPD for parameter X_1 with a bimodal distribution. (d) Marginal PPD for parameter X_2 with a univariate Gaussian distribution.

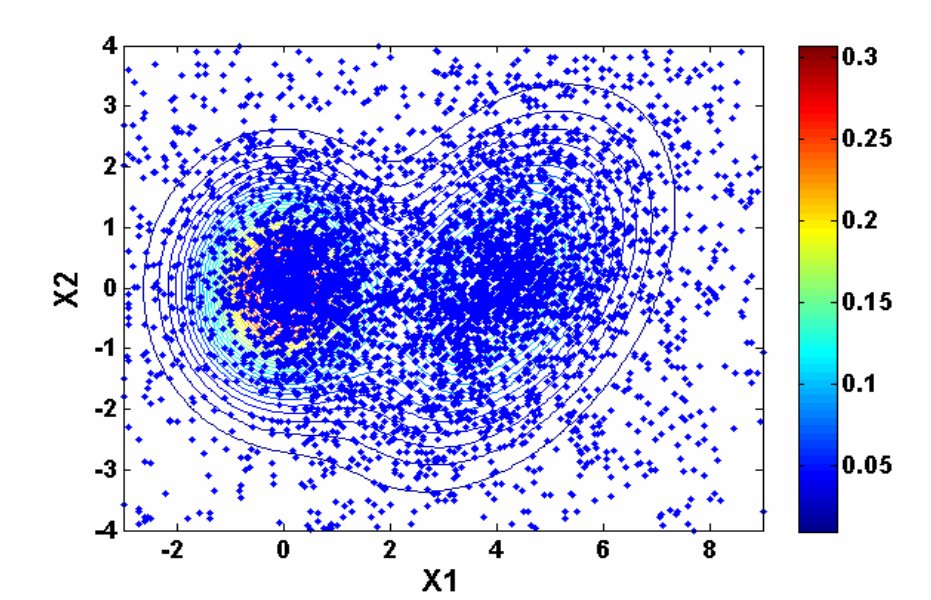

Figure 3.3: 9000 samples drawn by the multi-scale GA based MCMC algorithm, which overlie the contours of the true PPD. More samples are drawn over the peak regions and sample density distribution matches the contours well.

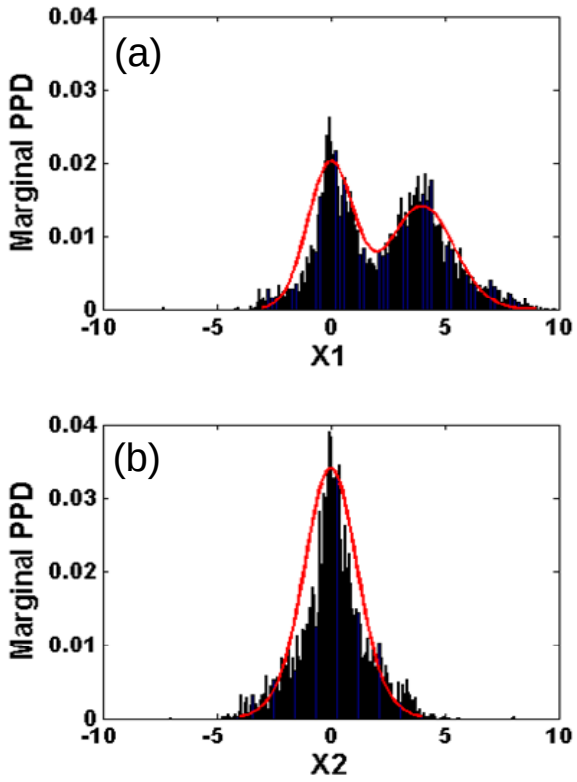

Figure 3.4: Marginal PPDs are estimated with the relative frequency distributions of the 9000 sample points, which are compared against the true marginal PPDs for (a) X_1 and (b) X_2. As shown, accurate estimations are obtained.

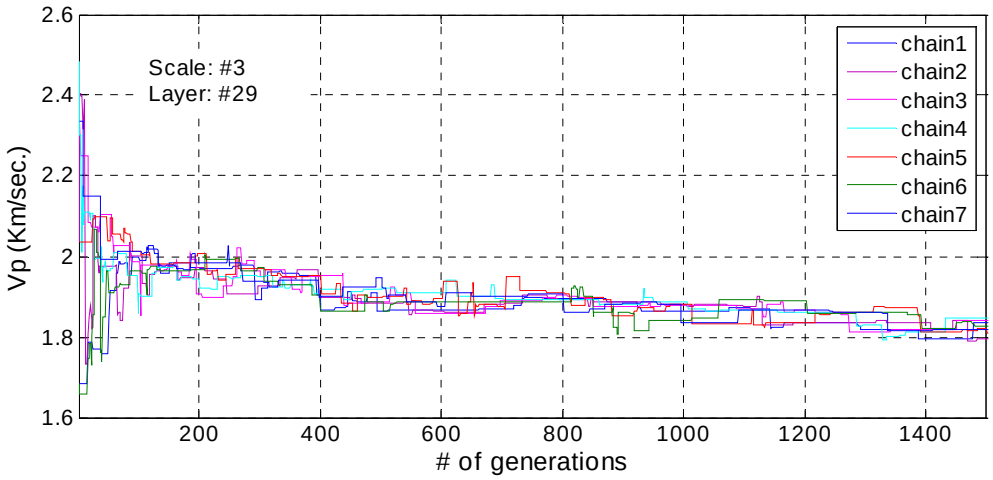

Figure 3.5: Samples of parameter V_p of the 29th layer, which are drawn by the 7 MCMC chains at scale #3. As shown, all the 7 chains converge to the same value of about 1.82 *km/s,* even though they start with different seeds.

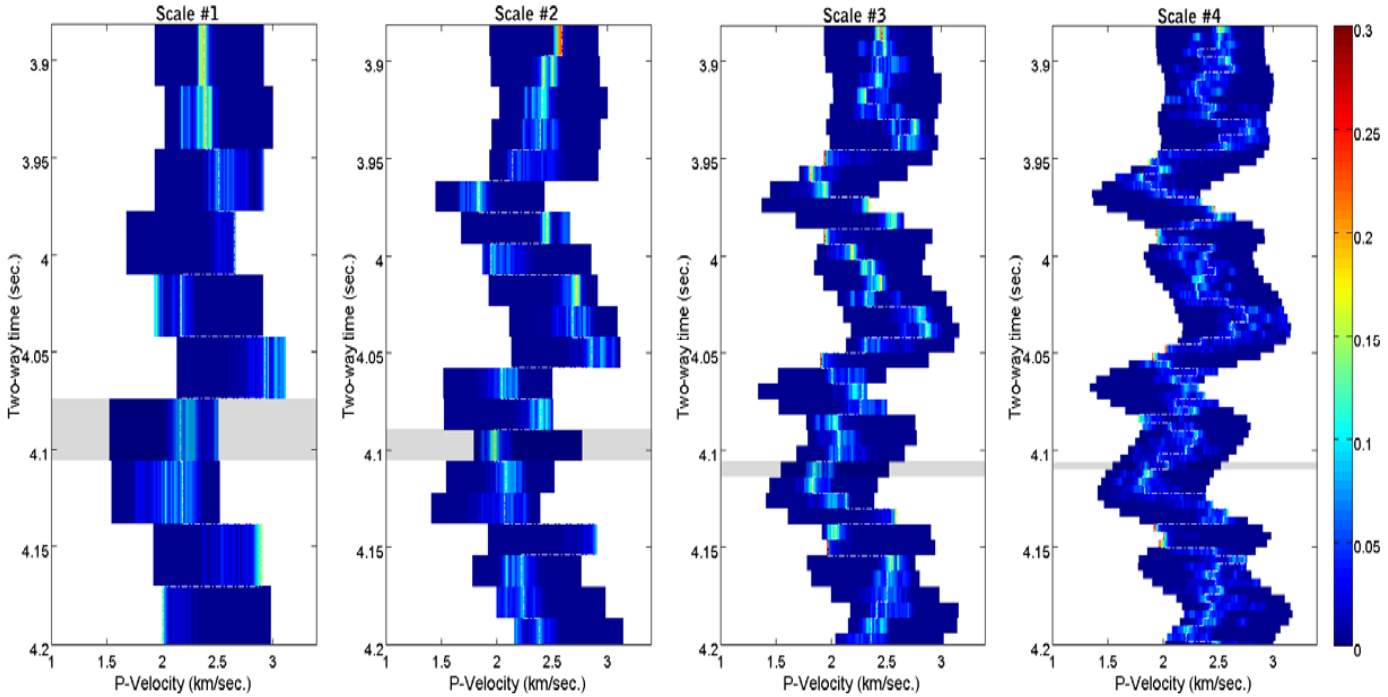

Figure 3.6: Inferred marginal PPDs and MAP models of *Vp* by the multi-scale GA-type MCMC for 4 scales. Dashed white lines represent the final converged MAP models as shown in Figure 2.5. Color bar represents the posterior probability. Marginal PPDs, together with CDFs, of the grayed layers are illustrated in Figure 3.7, which provide both the most probable values and uncertainty bounds.

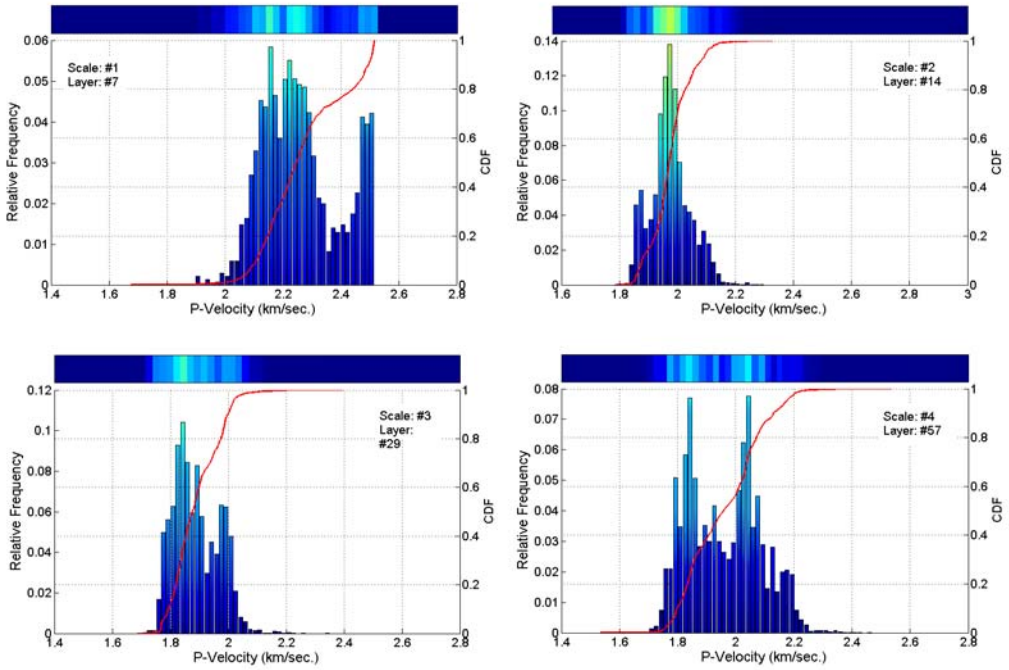

Figure 3.7: CDFs and marginal PPDs of *Vp* of the grayed layers in Figure 3.6 for the 4 scales, which indicates both the most probable (highest frequency) values and the corresponding uncertainty bounds. Similar figures can be shown on *Vs* and density.

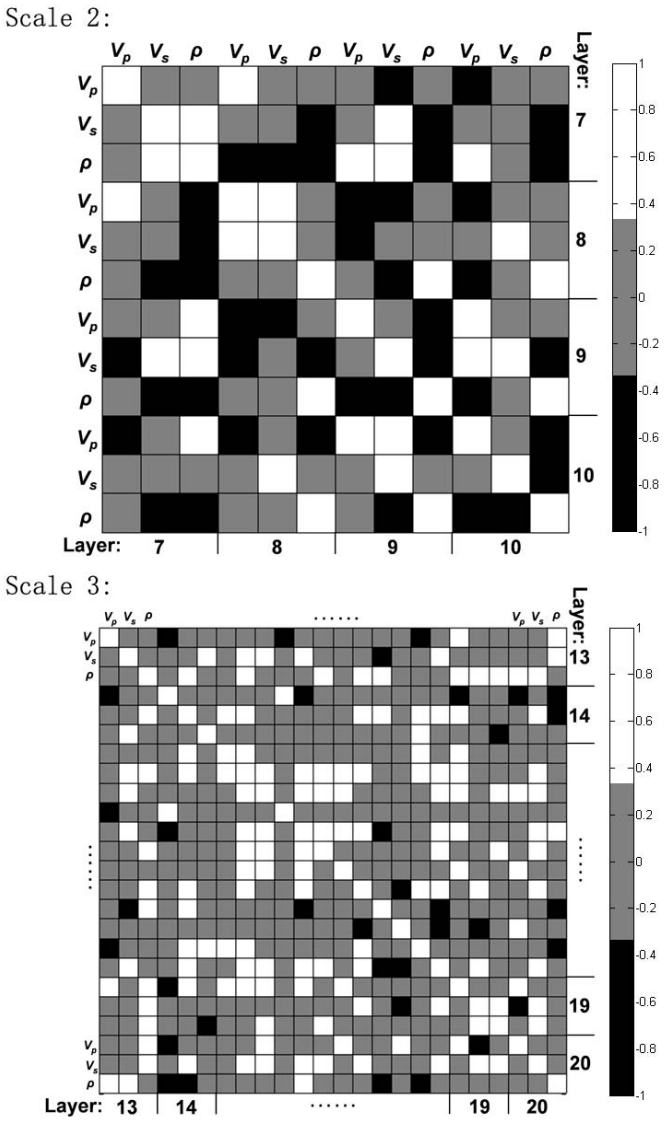

Figure 3.8: Parts of the posterior correlation matrices for scale #2 ranging from 7th to 10 the layer, and for scale #3 ranging from 13th to 20th layer. Black color represents strong negative correlation while white color represents strong positive correlation between parameters. The middle gray color stands for uncorrelation between parameters.

Chapter 4

Integrated 4D seismic history matching

Developing accurate reservoir models is a key objective of oil production companies. A properly constrained reservoir model can be used to accurately evaluate the total volume of recoverable hydrocarbon reserves in place and to predict the dynamic fluid flow in a reservoir, which facilitates future drilling plans and optimization of production strategy. Reservoir flow simulation is often used to help understand the changes in reservoir conditions throughout the production history.

Reservoir characterization is usually achieved using a combination of multiple datasets, such as seismic data, well production data, well logs and sometimes sensor measurements. Each of these datasets represents imperfect measurements having a different spatial and temporal resolution and also corresponding to different physics. How these datasets are used plays a very important role in determining the quality and accuracy of the final reservoir model estimation.

The concept of using time-lapse (4D) seismic data for reservoir delineation has been studied for years; the technology has had rapid acceptance as a practical reservoir management tool. Time-lapse 4D seismic data comprise a set of 3D seismic data volumes acquired at different calendar dates over the same area, with the objective of monitoring changes occurring in a producing hydrocarbon reservoir over time. Changes in reservoir

properties due to production or injection can be recorded by 4D seismic data because seismic velocities and impedance are affected by the changes in reservoir pore fluids, pressure and temperature (Nur and Simmons, 1969; Wang and Nur, 1992; Wang, 2000). The overall elastic moduli of a rock change with the type of fluid in the pores, the effective pressure acting on the rock and the temperature to which the rock is subjected. Due to the change in elastic moduli, the rock becomes more or less resistant to wave-induced deformations; therefore, seismic velocities experience an increase or decrease in magnitude. These observations form the basis of using 4D seismic data in predicting fluid saturation and pore pressure changes in a reservoir.

Providing a spatial distribution of property changes over time in a producing reservoir, 4D seismic data can be treated as dynamic data, and hence can be history matched along with historical well production data in the reservoir modeling workflow. This quantitative use of seismic data leads to the so-called seismic history matching scheme, which calls for forward modeling the 4D seismic response and comparing it with the observed field observations.

In this chapter, a new workflow is proposed to simultaneously integrate 4D seismic data, well production data and sensor measurements for reservoir model estimation as well as the corresponding uncertainty analysis (Hong *et al*., 2007b; Hong and Sen, 2008b). In this integrated workflow, the basic steps involve running the forward modeling process on a starting reservoir model using a flow simulator to generate flow parameter distributions at multiple legacy times. These output parameter distributions are converted to synthetic monitor seismic surveys at the legacy times using a rock physics

model and a simple seismic convolution algorithm. Finally, the difference between the synthetic and observed 4D seismic and production data is minimized using a global optimization algorithm until the reservoir model converges to a good estimation. By repeating this workflow for multiple runs, many realizations of the reservoir model are obtained and can be further used for uncertainty appraisal.

4.1 4D seismic

Time-lapse seismic monitoring is an integrated reservoir exploitation technique. Differences over time in seismic attributes are due to changes in pore fluids and pore pressure during the drainage of a reservoir under production. The detection of areas with significant changes or with unaltered hydrocarbon-indicative attributes facilitates future drilling plans by helping target the hydrocarbon remaining after a certain time period of production.

Time-lapse seismic usually consists of multiple seismic surveys made within the same area and at different times. The surveys can be 2D seismic, 3D seismic, Vertical Seismic Profiles (VSP) and cross-well seismic. The first survey is often called the base survey, and the subsequent repeated surveys are called monitor surveys. Seismic data for a single survey are sensitive to both static reservoir parameters, such as lithology, porosity, permeability and shaliness, and dynamic flow related parameters, such as fluid saturation, pore pressure and temperature. Compared to 3D seismic, 4D seismic data capture both the static and dynamic reservoir features and therefore provide a means to

decouple these two sources of effects. By comparing seismic data obtained from surveys at different times, it is possible to reduce the effects from static parameters and qualitatively focus on dynamic changes coming from pressure and saturation redistributions due to the on-going production (Landro, 2001; Chapin *et al.*, 2002; Landro and Stronen, 2003; McInally *et al.*, 2003; Gouveia *et al*., 2004).

The first study of time-lapse seismic can be traced back to the late 1980's and 1990's, when three different papers describing how multiple seismic surveys were used to monitor reservoirs under production, by comparing data recorded from different surveys (Dong, 2005). The first two applications were to monitor the steam movement in enhanced oil recovery projects (Wayland and Lee, 1986; Greaves and Fulp, 1987). A few years later, the first paper documenting the application of time-lapse seismic data to monitor fluid flow under an isothermal condition was published by Dunlop *et al*. (1991). Other applications to reservoir monitoring and management were published in the past decade (Ross *et al.*, 1996; Lumley *et al.*, 1999; Behrens *et al.*, 2002; Waggoner *et al.*, 2002; Gouveia *et al.*, 2004). According to Lumley (2001), through 2001, there have been 75 active time-lapse seismic projects worldwide, costing on the order of (US) $50 million to (US) $100 million, and such seismic projects have been increasing.

Making sure that seismic differences are related to fluid flow is critical for a complete time-lapse seismic study. Differences in data acquisition, survey orientation, processing and quality of datasets can introduce significant noise in the 4D analysis. Whether the time-lapse seismic data are used for reservoir monitoring or additional input for reservoir characterization, comparison between base survey and legacy surveys is

always useful, commonly subtracting one from the other. To ensure that the image obtained at beginning time is comparable to the subsequent images and therefore to make a meaningful comparison, two important issues need be addressed:

- Signals from locations with no physical property changes should be kept as similar as possible in different surveys;
- Differences related to saturation, pressure or stress should be as large as possible in different surveys.

The primary factors influencing these two important issues may come from the in-situ conditions of the reservoir, such as reservoir depth, elastic properties of reservoir rocks and fluid phase changes between surveys, etc. These in-situ conditions generally determine whether the reservoir property changes will be large or not. Whenever possible, it is beneficial to apply the same geometry alignment and acquisition parameters over the same survey area, and to follow the same processing flows in both the base and the legacy surveys. In 4D seismic terminology, these are the feasibility and repeatability studies in the design phase and cross equalization in the processing phase.

4.1.1 Feasibility study

As reservoir fluids are produced or fluids are injected, the reservoir's seismic properties may change. If these changes are sufficiently large, repetitive seismic surveys can help researchers map fluid distributions and fluid fronts and pressure and temperature changes that indicate that the time-lapse seismic reservoir monitoring may be successful.

According to Wang (1997), the feasibility of 4D seismic monitoring consists of two aspects: the physical and the seismic. It is obvious that not all reservoir conditions or recovery processes are suitable for 4D seismic analysis because the combined total effect of all the changes in the reservoir is not necessarily large enough to be differentiated seismically at a given resolution. Both Wang (1997) and Lumley *et al.* (1997) provide excellent discussions of what kind of reservoirs tend to be good candidates for a time-lapse seismic project.

Wang (1997) identified some critical physical factors listed in Table 4.1. These factors have a significant influence on the success of a time-lapse seismic project.

- *Frame elastic properties of reservoir rocks*. The frame of a reservoir rock is defined as a rock with empty pores. Relatively low frame elastic properties (bulk and shear moduli) are a first-order requirement for the expected success. Rocks with low frame elastic properties typically include unconsolidated or poorly consolidated rocks, high-porosity rocks with open fractures, low aspect-ratio pores or many grain-to-grain contacts, and rocks under low net overburden (effective) pressure.
- *Contrast in pore fluid compressibility*. It is assumed that lithology is not changed in a producing reservoir, and the main objective of time-lapse monitoring is to image the fluid change and other related changes such as pore pressure and temperature. For this purpose, a contrast in pore fluid compressibility is required between the original reservoir fluids and the subsequent reservoir fluids. High-contrast in pore fluid compressibility is

another first-order requirement for the expected success. Table 4.2 lists several cases in which high contrast in compressibility may exist.

- *Nature of recovery process.* This represents the changes in reservoir seismic properties that are caused by the production/injection process. As not all recovery processes can be monitored seismically, it is extremely important for reservoir geophysicists to understand the nature of the recovery process and the changes that it may produce in the seismic properties of the reservoir.
- *Reservoir parameters.* Changes in reservoir temperature induce changes in both rock frame and pore fluid properties. In seismic monitoring of water injection, high original reservoir temperature is always advantageous because of the large dependence of oil compressibility, but small dependence of water compressibility, on temperature. High porosity tends to weaken the reservoir rock, so it is a positive factor in time-lapse seismic monitoring.

CRITICAL FACTORS IN PHYSICAL FEASIBILITY
1) Frame elastic properties of reservoir rocks
2) Contrast in pore fluid compressibility
3) Nature of recovery process
4) Reservoir parameters (depth, pressure, temperature, etc.)

Table 4.1: Critical factors in the physical feasibility of time-lapse seismic reservoir monitoring (Wang, 1997).

CASES OF HIGH CONTRAST IN COMPRESSIBILITY		
Reservoir fluids	**From/to**	**Reservoir fluids**
Liquid (water, oil)	↔	Gas (hydrocarbon, gas, steam)
Oil/water	↔	CO_2 (liquid or gas)
Live oil	↔	Water/brine
Oil (live or dead)	↔	High-salinity brine
Live oil	↔	Dead oil
Low-temperature oil	↔	High-temperature oil

Table 4.2: Cases of possible high contrast in pore fluid compressibility (Wang, 1997).

Based on the above analysis, Wang (1997) developed a list of good candidate reservoirs as shown in Table 4.3, which represents a qualitative or rule-of-thumb assessment of the physical feasibility for time-lapse seismic reservoir monitoring. In addition, Lumley *et al*. (1997) discussed the seismic feasibility and proposed some ideal parameters of seismic operations as in Table 4.4 that are likely to give rise to a successful time-lapse seismic project.

GOOD CANDIDATE RESERVOIRS
1) Reservoirs with weak rocks
2) Reservoirs undergoing large pore fluid compressibility changes
3) Reservoirs undergoing rock compressibility changes
4) Reservoirs undergoing large temperature changes

Table 4.3: Good candidate reservoirs for time-lapse seismic reservoir monitoring (Wang, 1997).

SEISMIC	IDEAL CASES
Dominant frequency (*Hz*)	High
Average resolution (*ft*)	Low
Image quality (1- lowest; 5- highest)	5
Fluid contact visibility (1- lowest; 5- highest)	5
Predicted travel time changes (samples)	> 4
Predicted impedance change (%)	> 4

Table 4.4: Ideal parameters of time-lapse seismic operations (Lumley *et al*., 1997).

4.1.2 Repeatability study

Repeatability does not mean to make multiple seismic surveys as similar as possible since the very important purpose of time-lapse seismic is to highlight the reservoir changes over production time. Therefore, the expected success usually requires that differences in regions altered by production should be maximized while regions unaltered by production should be minimized. These differences are sensitive not only to changes in reservoir rock properties but also to differences in acquisition and processing. Some of the factors that affect repeatability include:

- Acquisition geometry differences such as sail line orientation and heading, source-receiver spacing, streamer feather, and coverage due to obstruction
- Near-surface conditions resulting in statics and receiver coupling variations

- Sea level, sea state and swell noise, water temperature and salinity
- Residual multiple energy
- Ambient and shot-generated noise
- Geological factors such as shallow gas and steep geological dip

The central point of 4D seismic acquisition and processing is to minimize differences in the seismic data that are unrelated to production and to preserve the differences that are due to production. A number of strategies have been developed to maximize acquisition repeatability, and the permanent monitoring systems can lead to high repeatability. A key to successful 4D processing is continual comparison of the base and legacy surveys to ensure that repeatability is not being compromised. The 4D processing can be described as a parallel processing of base and monitor surveys, which implies:

- Controlled amplitude and phase
- Early equalization of geometry to facilitate QC (quality control) comparisons
- Application of the same algorithms and parameters as appropriate

In many situations, the base and monitor surveys are taken from different seismic surveys in the same field. Thus, the acquisition systems are commonly different in geometrical alignment, and the data processing workflows are different in parameters. Since it is impractical to re-shoot the surveys, the only feasible way to reduce these artificial distinctions is through data re-processing by applying the so-called cross equalization scheme. Cross equalization is normally required when a time-lapse seismic

project involves legacy data, and it also helps in well design by facilitating discrepancy adjustment between surveys. Although specific steps may vary in cross equalization processing (Ross *et al.*, 1996; Rickett and Lumley, 2001), the following three steps are commonly used:

- Realignment of geometry of acquisition system
- Time, bandwidth, and phase equalization
- Amplitude equalization

4.2 Post-stack seismic modeling and flow simulation

Fluid flow simulations are routinely used as the main input to the economic evaluation of hydrocarbon recovery. Synthetic predictions from these simulations have proven to be sensitive to reservoir characterization that is usually known from geology and petrophysics at well locations. Because the latter is based primarily on sparsely-spaced wells, there is considerable uncertainty in the characterization and therefore uncertainty in the performance prediction.

Quantitative information contained in seismic data has been used for reservoir characterization (Debeye *et al.*, 1996; Pendrel and van Riel, 1997). The seismic data sample the entire reservoir and thereby offer the possibility of filling the spatial gap between usually sparse well locations. However, without coupling with the flow simulation, the seismic data alone cannot be used to directly describe reservoir engineering features. Therefore, reservoir simulation is usually based on a reasonably

simplified reservoir model to approximate the flow parameter distributions, which are then transformed into seismic elastic parameters, such as wave velocities and density, by performing a certain rock physics model.

4.2.1 Post-stack seismic modeling

As mentioned before, 4D seismic data are generally multiple 3D seismic data obtained throughout the production time and over the same area. Thus, 4D seismic modeling is actually a series of 3D seismic modeling of the same producing reservoir that has changed through time due to fluid movement.

Relationships between stress, strain and velocity of propagation lead to the seismic wave equations that are fundamental to understanding the behavior of earth materials. For an isotropic medium, the relationship can be characterized by only two independent elastic parameters, λ and μ, known as the Lame constants. The corresponding Hooke's Law is

$$\boldsymbol{\tau} = \lambda(\nabla \cdot \mathbf{u})\mathbf{I} + \mu\left[\nabla\mathbf{u} + (\nabla\mathbf{u})^T\right], \tag{4.1}$$

where **I** is the identity tensor; $\boldsymbol{\tau}$ and **u** are stress and displacement tensor, respectively.

Application of Newton's second law of motion (assuming no body forces) yields:

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} = \nabla \cdot \boldsymbol{\tau}, \tag{4.2}$$

For a homogeneous, isotropic elastic medium, the above equation (4.1) may be expressed in another form as:

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} = (\lambda + 2\mu)[\nabla(\nabla \cdot \mathbf{u})] - \mu(\nabla \times \nabla \times \mathbf{u}), \tag{4.3}$$

where **u** is the displacement vector, ρ is density, λ is the Lame's constant, and μ is shear modulus or sometimes called rigidity. λ, μ and other seismic elastic parameters, such as bulk modulus or wave propagation velocities, are related to each other (Sheriff, 1984).

Compressional-wave (P-wave) propagation is described as a particular case of equation (4.3) and is obtained by taking divergence on it, that is

$$\rho \frac{\partial^2 (\nabla \cdot u)}{\partial t^2} = (\lambda + 2\mu)[\nabla \cdot \nabla(\nabla \cdot u)] \tag{4.4}$$

By defining the divergence of the displacement as $\phi = \nabla \cdot u$, yields

$$\nabla^2 \phi = \frac{1}{V_p^{\,2}} \frac{\partial^2 \phi}{\partial t^2} \tag{4.5}$$

where V_p is the compressional-wave velocity and

$$V_p = \sqrt{\frac{\lambda + 2\mu}{\rho}} \tag{4.6}$$

Similarly, the equation of shear-wave propagation can be obtained by taking the curl of equation (4.3), that is

$$\rho \frac{\partial^2 (\nabla \times \mathbf{u})}{\partial t^2} = \mu[\nabla^2(\nabla \times \mathbf{u})] \tag{4.7}$$

Defining the curl of the displacement as a vector of $\boldsymbol{\psi} = \nabla \times \mathbf{u}$, yields

$$\nabla^2 \boldsymbol{\psi} = \frac{1}{V_s^{\,2}} \frac{\partial^2 \boldsymbol{\psi}}{\partial t^2} \tag{4.8}$$

where V_s is the shear-wave velocity and

$$V_s = \sqrt{\frac{\mu}{\rho}} \qquad (4.9)$$

Figure 4.1 graphically shows the reflection and transmission of an incident plane P-wave at a planar interface separating two homogeneous and isotropic media. In the most general case, a single incident plane wave at an interface gives rise to four types of wave: reflected and transmitted compressional waves, reflected and transmitted shear waves. Of these four waves, the primary echoes or reflections are most commonly used to infer features and properties of the target subsurface (Yilmaz, 2000). The ratio of the amplitude of a reflected wave compared to the amplitude of an incident wave is called the reflection coefficient. If the incidence angle is zero, i.e., the incident plane wave if directed normal to the planar interface, the reflected and transmitted waves do not change directions but do change signs. For the situation of normal incidence, the reflection coefficient at the planar interface is given as

$$r_{pp}^{1}(\theta = 0) = \frac{Z_2 - Z_1}{Z_2 + Z_1}, \qquad (4.10)$$

where Z_i stands for the P-wave acoustic impedance of layer i, and the acoustic impedance is the product of density and P-wave velocity; subscripts 1 and 2 indicate the particular medium above (1) and below (2) the interface; r_{pp}^{1} stands for the reflection coefficient of the first layer. For this 1D layered medium, the convolution of the time series of the calculated reflection coefficients with the source wavelet and this convolution results in a zero-offset synthetic seismogram, often called post-stack seismogram. Mathematical representation of this convolution is given as

$$s(t) = w(t) \otimes r_{pp}^{1} \delta(t - t_1) \qquad (4.11)$$

where *w(t)* is the source wavelet, and delta function of $\delta(t - t_1)$ accounts for the phase delay; t_1 is the two-way travel time of a specific layer. The response of a multi-layered model can be computed by summing the response of individual layers as computed by equation (4.10). This is the method that will be used in this research for the post-stack case.

This equation neglects transmission losses, internal multiples and frequency dispersion. In the frequency domain, the convolution equation (4.11) is equivalent to a product of the Fourier transforms of the two functions, namely,

$$S(\omega) = W(\omega)R(\omega) \qquad (4.12)$$

where ω is angular frequency, and *S*, *W*, and *R* stand for the Fourier transforms of the functions *s*, *w* and *r*, respectively. Compared to pre-stack seismic modeling as used in chapters 2 and 3, the disadvantage of such post-stack modeling is that the acoustic impedance does not provide independent sensitivity to the medium's bulk density and compressional-wave velocity.

4.2.2 Flow simulation

Reservoir simulation is a tool for predicting future production rates from a given reservoir engineering model. The theory is based on conservation of mass and energy equations and the mass transport mechanism equations (Darcy's law) in a porous medium.

Suppose that the reservoir under study is rectangularly shaped over a volume Ω,

$$\Omega = \{(x, y, z) | 0 < x < L_x, 0 < y < L_y, 0 < z < L_z\} \qquad (4.13)$$

where L_x, L_y, and L_z are reservoir lengths along x, y and z directions, respectively. The general equation of mass conservation in equation (4.14) shows that any local change in the flow of mass is because of a local source/sink or a temporal change in density or porosity. The source/sink term $\tilde{q}$ accounts for a local production/injection well. The vector **v** is the fluid flux, ϕ is porosity and ρ is fluid density (Ertekin *et al*., 2001).

$$-\nabla \cdot (\rho \mathbf{v}) = \frac{\partial(\phi\rho)}{\partial t} + \tilde{q}; \qquad (4.14)$$

Darcy's law is a gradient expression that relates the flux velocity to pressure and gravity, as shown in the following equation

$$\mathbf{v} = -\frac{K}{\mu}(\nabla p - \rho g \nabla z); \qquad (4.15)$$

where K is permeability, and μ is the viscosity. The second term on the right-hand side incorporates the gravitational effect. The generalized form of Darcy's law to the flow of multi-phases (oil, water and gas) in a porous rock is given as

$$v_j = -\frac{K k_{rl}}{\mu_j}(\nabla p_j - \gamma_j \nabla z) \qquad j = o, w, g \qquad (4.16)$$

where k_{rl} is the relative permeability of each phase, which is a function of saturation.

The combination of Darcy's law with the mass conservation equation yields the following multi-phase fluid flow equations:

$$\begin{cases} \nabla\left[\gamma_o\left(\nabla p_o - \rho_o g \nabla z\right)\right] = \dfrac{\partial\left(\phi S_o / B_o\right)}{\partial t} + q_o \\ \nabla\left[\gamma_w\left(\nabla p_w - \rho_w g \nabla z\right)\right] = \dfrac{\partial\left(\phi S_w / B_w\right)}{\partial t} + q_w \\ \nabla\left[R_S\gamma_o\left(\nabla p_o - \rho_o g \nabla z\right) + \gamma_g\left(\nabla p_g - \rho_g g \nabla z\right)\right] = \dfrac{\partial\left(\phi R_S S_o S_g /(B_o B_g)\right)}{\partial t} + R_S q_o + q_{fg} \end{cases} \tag{4.17}$$

These equations relate differences in pressure gradient (left-hand side) to temporal changes in pore saturation (right-hand side), and they are usually complemented by three additional constraints that account for saturation of three fluids and empirical relationships between capillary pressure and individual saturations:

$$\begin{cases} S_o + S_w + S_g = 1 \\ p_o - p_w = f\left(S_w, S_g\right) \\ p_g - p_o = f\left(S_w, S_g\right) \end{cases} \tag{4.18}$$

In equation (4.17), parameters B_j relate reservoir property values at surface condition to reservoir condition, and they are defined based on the PVT constant at thermodynamic equilibrium; R_S accounts for the mass transfer between the oil and gas phases; coefficients γ_j are transmissibilities and are defined as $\gamma_j = \dfrac{k_{rl}K}{\mu_j B_j}$. The derivation of equation (4.17) uses the black-oil model, which assumes that the only fluid phases are oil, gas and water, and the gas can dissolve into oil, but oil cannot be vaporized into gas, water; water and oil are immiscible.

Since the PDEs in equation (4.17) are nonlinear, it is difficult to solve them analytically. When the finite difference scheme is applied to solve them, the reservoir volume Ω is usually discretized into N_b grid blocks and $N_b = N_x \times N_y \times N_z$, where N_x, N_y, and N_z stand for the number of grid blocks along x, y, and z directions, respectively. For each of the N_b blocks, there exist three finite difference equations for oil, gas, and water. In total, there will be $3N_b$ equations representing the mass balance over the whole reservoir volume. In addition, N_w well equations will have to be added if there are N_w wells. Each well can have a different constraint, such as constant bottomhole pressure, constant oil production rate, and constant total production rate. If bottomhole pressure is applied as a constraint in wells, the corresponding phase production rate can be computed using the Peaceman equation (Peaceman, 1983). These $3N_b+N_w$ equations plus the boundary and initial conditions consist of a complete set of reservoir simulation equations. The solution of the system for pressure and saturation distributions of each phase in each gridblock at different times is the so-called reservoir simulation. The primary variables to be solved in each gridblock are case dependent. For example, p, S_o and S_g are typically the primary variables in a 3-phase system, but if no free gas is present, solution gas-oil ratio, R_s, may replace S_g as one of the primary variables.

The flow simulator that is used in chapter 5 is a standard IMPES (implicit in pressure and explicit in saturation) simulator. At each time step, the set of finite difference equations for the overall pressure equation are solved to obtain gridblock pressures, and the finite difference form of the water flow equations are solved for

gridblock water saturations. The pressure finite difference equations are solved using a sparse matrix routine.

4.3 Seismic rock physics

Seismic responses are affected in coupled ways by many factors, such as porosity, fluid type, pressure, saturation etc. Currently, seismic data collaborated with other measurements are commonly analyzed for reservoir monitoring, lithology discrimination, and hydrocarbon detection through rock physics relationships. These relationships transform reservoir petrophysical properties into seismic elastic attributes, which are further used for simulation of wave propagation (Tatham and McCormack, 1991; Batzle and Wang, 1992; Mavko *et al*., 1998). Because rock physics bridges seismic data and reservoir parameters, it has been instrumental in the development of technologies such as 4D seismic reservoir monitoring, seismic lithology discrimination, and direct hydrocarbon detection with "bright-spot" and angle-dependent reflectivity analysis.

Seismic compressional-wave and shear-wave velocities have been related to the Lame's constants as in equations (4.6) and (4.9) respectively, and they can also be expressed differently as follows:

$$V_p = \sqrt{\frac{K_u + \frac{4}{3}\mu}{\rho_u}} \qquad (4.19)$$

$$V_s = \sqrt{\frac{\mu}{\rho_u}} \qquad (4.20)$$

where K_u is the saturated (undrained) rock's bulk modulus, ρ_u is the saturated (undrained) rock's shear modulus, and ρ is the saturated rock's density. Wang and Nur (1992) developed a series of empirical relationships to calculate fluid seismic velocities and densities utilizing reservoir flow parameters. The calculated fluid seismic velocities and densities can be further used to derive fluid bulk modulus, and then the saturated rock bulk modulus (K_u) can be calculated through the fluid substitution Gassmann equation (Wang and Nur, 2000). The saturated rock density (ρ_u) can be calculated from the different fluid densities and reservoir parameters such as porosity and fluid saturation, etc.

The Gassmann (1951) equation has been used to calculate the effect of fluid substitution on seismic response of low-frequency waves, while Biot's equation (1956a, 1956b) extends the effect to full frequency range but predicts very little velocity dependence on frequency for most reservoir rocks. Geertsma and Smit (1961) developed equations valid for the complete frequency range based on Biot's work, and the results are summarized as follows.

$$\begin{cases} V_p^2 = \left[(K_u + \frac{4}{3}\mu) + \dfrac{\dfrac{\phi\rho_u}{k\rho_f} + (1-\dfrac{K_u}{K_d})(1-\dfrac{K_u}{K_d}-\dfrac{2\phi}{k})}{(1-\phi-\dfrac{K_u}{K_d})(\dfrac{1}{K_d}) + \dfrac{\phi}{K_f}} \right] \cdot \left[\dfrac{1}{(\rho_u - \dfrac{\phi\rho_f}{k})} \right]; \\ V_s^2 = \left[\dfrac{\mu}{(\rho_u - \dfrac{\phi\rho_f}{k})} \right]; \end{cases} \tag{4.21}$$

where V_p is the P-wave velocity, V_s is the S-wave velocity; K_u is the bulk modulus, μ is the bulk shear (rigidity) modulus, K_d is the rock's dry bulk modulus, K_f is the modulus of an oil-water-gas fluid mixture; ϕ is the porosity and k is the mass coupling factor, which varies from 1 (no fluid–solid coupling) to infinity (perfect coupling); ρ_u is the saturated porous bulk density, ρ_f is the fluid density.

As discussed in Wang and Nur (2000), some basic assumptions in Gassmann's equation are: (1) the porous rock is macroscopically homogeneous and isotropic; (2) all the pores are interconnected or communicating; (3) the pores are filled with a frictionless fluid (including gas); (4) the rock-fluid system under study is closed or undrained; (5) the relative motion between the fluid and the solid rock is negligibly small compared to the motion of the saturated rock itself when the rock is excited by a wave; (6) the pore fluid does not interact with the solid in a way that would soften or harden the frame. For heavy oil saturated unconsolidated sands, assumption (2) is very well satisfied, but assumption (3) is violated. Wang and Nur (2000) compared laboratory data with Gassmann's predictions and showed that for the sands and sandstones under low effective pressure (10MPa), the Gassmann-predicted V_p is lower than the measured V_p by as much as 8%.

To calculate P- and S-wave velocities based on the Gassmann equation (4.21), ρ_u, ρ_f and K_f are needed. These three terms can be calculated according to equations (4.22)-(4.24) respectively

$$\rho_f = S_g \rho_g + S_o \rho_o + S_w \rho_w \qquad (4.22)$$

$$\rho_u = \rho_s (1-\phi) + \rho_f \phi \quad . \qquad (4.23)$$

Here, ρ_o, ρ_g, ρ_w, ρ_s, ρ_u, and ρ_f respectively stand for the densities of oil, gas, water, solid grains, saturated reservoir rock, and fluid mixture at reservoir condition. S_g, S_o, and S_w are the saturations of gas, oil and water. The ρ_o, ρ_g, ρ_w at the reservoir condition have to be calculated first in order to calculate the adiabatic gas, oil, and water bulk modulus K_g, K_o, and K_w following the equations developed by Batzle and Wang (1992) based on the known reservoir pressure (P), temperature (T), gas specific gravity (G) and water salinity (S).

For gas,

$$\rho_g = \frac{28.8GP}{ZRT_a} \qquad (4.24)$$

where

$$Z = [0.03 + 0.00527(3.5 - T_{pr})^3]P_{pr} + (0.642T_{pr} - 0.007T_{pr}^4 - 0.52) + E$$
$$E = 0.109(3.85 - T_{pr})^2 \exp\{-[0.45 + 8(0.56 - 1/T_{pr})^2]P_{pr}^{1.2}/T_{pr}\}$$
$$P_{pr} = P/(4.892 - 0.4048G), \qquad T_{pr} = T_a/(94.72 + 170.75G)$$

T_a is the absolute temperature and T_a=T (ºC)+273.15. The units used here are: *MPa* for pressure and bulk modulus; g/cm^3 for density; *m/s* for velocity.

$$K_g = \frac{P}{\left(1 - \frac{P_{pr}}{Z}\frac{\partial Z}{\partial P_{pr}}\right)_T}\gamma \qquad (4.25)$$

where the subscript T means to do partial differentiation with respect to P, and

$$\gamma = 0.85 + \frac{5.6}{(P_{pr} + 2)} + \frac{27.1}{(P_{pr} + 3.5)^2} - 8.7\exp[-0.65(P_{pr} + 1)]$$

For oil,

$$\rho_o = \frac{\rho_t^p}{B_o}\left(1 + 0.001 R_G\right)^{-1} \tag{4.26}$$

where

$$B_o = 0.972 + 0.00038\left[2.49 R_G \left(\frac{G}{\rho_o}\right)^{1/2} + T + 17.8\right]^{1.175}$$

$$\rho_t^p = \rho_p / \left[0.972 + 3.81\times 10^{-4}\left(T + 17.78\right)\right]^{1.175}$$

$$\rho_p = \rho_o^{St} + \left(0.00277P - 1.71\times 10^{-7} P^3\right)\left(\rho_o^{St} - 1.15\right)^2 + 3.49\times 10^{-4} P$$

$$V_o = 2070\left(\frac{\rho_o}{2.6 - \rho_o}\right)^{1/2} - 3.0T + 4.64P + 0.0115\left[4.12\left(1.08\rho_o^{-1} - 1\right)^{1/2} - 1\right]TP \tag{4.27}$$

ρ_o^{St} is oil density at the standard condition (15.6ºC and atmospheric pressure). V_o is oil P-wave velocity. R_G is gas-oil ratio at the standard condition.

For water,

$$\rho_w = \rho_{pw} + S\left\{0.668 + 0.44S + 10^{-6}\left[300P - 2400PS + T\left(80 + 3T - 3300S - 13P + 47PS\right)\right]\right\} \tag{4.28}$$

where

$$\rho_{pw} = 1 + 10^{-6}(-80T - 3.3T^2 + 0.00175T^3 + 489P - 2TP + 0.016T^2P - 1.3\times 10^{-5}T^3P - 0.333P^2 - 0.002TP^2)$$

$$V_w = V_{pw} + S(1170 - 9.6T + 0.055T^2 - 8.5\times 10^{-5}T^3 + 2.6P - 0.0029TP - 0.0476P^2) + S^{1.5}(780 - 10P + 0.16P^2) - 1820S^2 \tag{4.29}$$

Here S is water salinity (the weight fraction of sodium chloride, *ppm*/1000000). V_w is the P-wave velocity for water, and $V_{pw} = \sum_{i=0}^{4} \sum_{j=0}^{3} W_{ij} T^i P^j$ is P-wave velocity for pure water, where

$$W_{00} = 1402.85, \quad W_{02} = 3.437 \times 10^{-3}, \quad W_{10} = 4.871, \quad W_{12} = 1.739 \times 10^{-4},$$
$$W_{20} = -0.04783, \quad W_{22} = -2.135 \times 10^{-6}, \quad W_{30} = 1.487 \times 10^{-4}, \quad W_{32} = -1.455 \times 10^{-8},$$
$$W_{40} = -2.197 \times 10^{-7}, \quad W_{42} = 5.23 \times 10^{-11}, \quad W_{01} = 1.524, \quad W_{03} = -1.197 \times 10^{-5},$$
$$W_{11} = -0.0111, \quad W_{13} = -1.628 \times 10^{-6}, \quad W_{21} = 2.747 \times 10^{-4}, \quad W_{23} = 1.237 \times 10^{-8},$$
$$W_{31} = -6.503 \times 10^{-7}, \quad W_{33} = 1.327 \times 10^{-10}, \quad W_{41} = 7.987 \times 10^{-10}, \quad W_{43} = -4.614 \times 10^{-13}$$

After equations (4.24) to (4.29) are carried out, ρ_f can be calculated based on equation (4.22), and the saturated bulk density ρ_u can be calculated based on equation (4.23).

It is assumed that fluids have no effect on the estimated shear modulus, that is μ_o=μ_g=μ_w=0. K_g, K_o, and K_w can be calculated by K=V_p^2*ρ, where V_p is the P-wave velocity of fluids. The average fluid bulk modulus for a multiphase system depends on the fluid distribution. Wood's equation as in equation (4.30) can be used to calculate the combined fluid bulk modulus.

$$\frac{1}{K_f} = \frac{S_g}{K_g} + \frac{S_o}{K_o} + \frac{S_w}{K_w} \qquad (4.30)$$

By this point, the only remaining unknown in Gassmann's equation is K_d. It was assumed that K_d does not vary with different fluid saturation, but it is affected by effective pressure P_e (overburden pressure minus pore pressure) and temperature. The same assumption is applied to the shear modulus. Consequently, the saturated shear modulus μ_u equals the dry shear modulus μ_d.

4.4 Integrated 4D seismic history matching workflow

Scientists from different disciplines usually view reservoir characterization differently. Geologists normally work with geologic observations, such as well logs and outcrop information, and mainly consider reservoir characterization from the viewpoint of geological stratigraphy and architecture. Geophysicists mainly work with geophysical data, such as seismic and electromagnetic data, and focus on reservoir shape, structure as well as some geophysical parameter distributions. Engineers are interested in reservoir rock and fluid properties, well conditions, etc., that are directly related to reservoir production and management.

Different data types generally present different benefits but also have different inherent limitations. Geologic observations usually only provide static information about a reservoir but cannot give any dynamic indication inside a reservoir. Static geologic knowledge helps build a rough initial reservoir model and also set up a reasonably good reservoir model space. In contrast to geologic information, 4D seismic and well production data capture both the static and dynamic reservoir features. Well measurements usually provide a very good vertical sampling on reservoir heterogeneity, but they are usually performed at sparsely distributed locations and therefore have great constraints in lateral resolution. Although not as good as well measurement in terms of vertical resolution, seismic data provide invaluable information about reservoir characterization because of the extensive spatial coverage and dense lateral sampling.

Thus, the most appropriate way to quantitatively characterize a reservoir with reduced uncertainty is to integrate all the available data sources, such as geologic, geophysical and engineering data, and thereby take advantage of their different benefits.

4.4.1 History matching

For typical oil, gas and water flow problems, the governing PDEs are nonlinear, which makes it almost impossible to provide an analytical solution for the saturation and pressure distributions. As discussed in section 4.2.2, reservoir flow simulation is a useful tool for numerically estimating the distributions of oil, gas and water at the partitioned reservoir gridblocks. To estimating reservoir flow parameters is equivalent to deriving reservoir simulation model parameters by generating synthetic production histories and iteratively fitting them to the field well observations. This process is often called history matching. Because reservoirs are usually very heterogeneous, there are numerous gridblocks in a typical reservoir simulation system, in order to capture reservoir parameters with a reasonably high resolution. Manually adjusting parameters in the history matching procedure is unfeasible, and computers are used to automatically perturb the parameters.

Reservoir modeling normally starts by creating a high-resolution 3D geo-cellular model using static data. A hierarchical approach to building the 3D geo-cellular model is presented by Caers (2005). The major steps of this approach are sequentially shown as follows:

- Establish the architecture of the reservoir in terms of horizons and faults that are determined from 3D seismic data and well-markers.
- Build a 3D stratigraphic grid from the structural framework.
- Build a Cartesian grid from the stratigraphic grid. This grid ideally represents the coordinate system for the original depositional environment. A one-to-one relationship is established between each grid-cell in the Cartesian grid and in the stratigraphic grid. All data, well paths, well logs and 3D seismic data are imported into that Cartesian grid.
- Populate the Cartesian grid with facies rock types. Outcrop data and sedimento-logical models provide information on the style of facies architecture; well-log, core and seismic data provide local constraints on the spatial distribution of these facies types.
- Populate each facies type with porosity and permeability. Porosity is assigned to each grid cell of the Cartesian grid based on well-log and core data; permeability is derived from the porosity model. Porosity is usually determined first since the porosity data are more reliable and abundant than permeability data.
- Map the petrophyscial properties back into the stratigraphic grid to provide a high-resolution 3D geo-cellular model.

The 3D geo-cellular model established after completing the above steps honors all the static data. But it does not match the historical well production data until a history matching procedure is applied to finally arrive at a match between the synthetic and the observed production data. The high-resolution geo-cellular reservoir model often consists

of millions of grid cells, which precludes the practical performing of flow simulation and history matching. To make the flow simulation applicable, the number of grid cells in a geo-cellular model is usually reduced by upscaling it to a manageable dimension. This upscaled reservoir model is the input of the flow simulation and is iteratively perturbed to obtain a history match.

The history matching process can be thought of as a minimization problem, in which the objective is to derive a reservoir model by iteratively minimizing the mismatch between synthetic and observed well production histories. However, in a mathematical formulation, the history matching process usually results in an ill-conditioned problem. This means that there are not a sufficient number of observed data available to determine all the reservoir parameters uniquely. Estimates of reservoir properties obtained from production data are typically well resolved only near the wellbores. For regions far away from wells, model estimation is usually poorly constrained, especially when the target reservoir is large and the amount of production data is limited. If another type of data can be incorporated into the history matching scheme to help sample and account for the inter-well heterogeneity, then the corresponding nonuniqueness and uncertainty in reservoir model estimation will be reduced. Following this integrated history matching workflow, the derived reservoir model will simultaneously honor all the different types of data.

4.4.2 Seismic history matching

The need for reservoir characterization technology within the oil and gas industry is mostly driven by the reality that a better defined reservoir usually leads to greater drilling success and lower development cost. This is because a properly constrained reservoir model can be used to more accurately quantify hydrocarbons in place and to efficiently optimize hydrocarbon production. Therefore, better reservoir characterization technology calls for integration of all available subsurface data. Currently, this is typically achieved using a combination of static geologic data, 4D seismic data and well production data. Each of these data corresponds to a different physics and represents a certain level of resolution. So the methodology used to integrate these different data sources and to mathematically formulate them determines the quality of the final derived reservoir model.

Technology using 4D seismic data for reservoir characterization experiences a fast development and ranges from purely qualitative evaluation to a quantitative constraint in rigorous numerical model optimization process (Wayland and Lee, 1986; Dunlop *et al.*, 1991; Landa and Horne, 1997; Lumley *et al.*, 1999; Arenas *et al.*, 2001; Huang *et al.*, 2001; Gosselin *et al.*, 2003; Kretz *et al.*, 2004; Mezghani *et al.*, 2004; Dong and Oliver, 2005). The type of time-lapse seismic data used for reservoir property estimation varies among previous researches. For example, seismic attribute difference was used in work by Huang et al. (1997) and Arenas et al. (2001), whereas changes of fluid saturation and/or pore pressure were used in studies by Landa and Horne (1997),

Tura and Lumley (1999), Landro (2001) and Meadows (2001). In this dissertation, the whole waveform of any particular trace is used as the seismic constraint on the history matching process. In other words, not only the two-way travel time and amplitude but also the entire shape of any trace's seismogram is matched to derive an optimal reservoir model.

To simultaneously match the 4D seismic and well production data, the reservoir model is stochastically perturbed and an objective function, also called the error function, is evaluated. The objective function comprises two normalized terms of the mismatches in production data and 4D seismic data respectively. Different weights are usually assigned on the two terms to reflect different confidence in them. The stochastic perturbation and error appraisal process is iterated using an optimization algorithm until a certain stopping criterion is reached.

Assigning weights to the two different terms is more or less subjective. When the weight on seismic term is 0 and the weight on the production term is 1, the resulting problem will be the conventional history matching process with constraint from the well production data only. As discussed before, the derived solution will be highly nonunique and will not capture the heterogeneity in the reservoir with an acceptable resolution. Furthermore, the estimation of residual hydrocarbon distribution and prediction of future fluid flow can be unreliable. If the weights are reversed, that is, weight of 1 on the seismic term and weight of 0 on the production term, the result will be a seismic matching process that ignores the production data. This is actually a quantitative 4D seismic inverse problem for reservoir parameters. If neither of the two weights is 0, the

resulting optimization process makes use of constraints from both of the two datasets to perform quantitative reservoir characterization. This is the so-called seismic history matching, in which a reservoir is derived such that, after both flow simulation and seismic modeling, the synthetic 4D seismic and production data match the recorded 4D seismic data and the well production data at the well locations. If the match is poor, the reservoir will be stochastically perturbed to improve the data fit. Before this matching process, the available static geologic scenario and well logs can be used to geostatistically generate the starting reservoir in order to initiate the seismic history match workflow.

4.5 2D example

Specialized sensors measure fluid and rock properties at a high local resolution, which, in conjunction with time-lapse seismic technique, possesses enormous potential to reduce the uncertainty in reservoir characterization and future production performance prediction (Klie *et al*., 2006). Therefore, sensor information is also incorporated in the proposed integrated seismic history matching workflow to provide additional constraints on pressure, saturation and flux velocity at some given locations. The resulting integrated workflow is illustrated in Figure 4.2. In this workflow, an initial reservoir model is first constructed based on available static information and is used to generate synthetic data according to different physics. The calculated data are then compared against the observed data to do error appraisal. If the error is acceptable, the workflow is terminated and a reservoir model is derived. Otherwise, the reservoir model will be stochastically

perturbed to improve data fit and the optimization loop is repeated until a satisfactory error is obtained.

As a preliminary test of the applicability and effectiveness of this integrated history matching scheme, a 2D synthetic example is designed based on a 2D cross section. This reservoir model has a grid dimension of 20x100x1, that is, of 2000 grid blocks in total. The size of each grid block is 2.5x25x25 ft^3. A fixed production strategy is adopted with one water injection well located at the leftmost side with flow rate specified and an oil production well at the opposite side with pressure specified as shown in Figure 4.3(a). In this application, permeability is taken as the primary reservoir model unknown. Porosity for each grid block is assumed constant at a value of 0.2 and pressure and fluid saturation are assumed to be independent of each other. Table 4.5 lists the flow and rock parameter values used in this example. Observations of time-lapse seismic data and production data are numerically generated at 10 legacy time steps through 1000 days. Specifically, a flow simulator is used on the true permeability distribution as in Figure 4.3(a) to generate the well production data at the producing well location. Outputs of pressure, saturation from the flow simulator are used as the inputs of the Gassmann model as introduced in section 4.2.2 to calculate seismic elastic parameters and hence seismic P-impedance at every grid block. The calculated P-impedance distribution is used to generate post-stack seismograms at the specified 10 legacy time steps using a convolution algorithm, as described in section 4.2.1.

PARAMETER	VALUE	UNIT
Oil pressure	1000.0	*kPa*
Oil concentration	43.15	%
Oil compressibility	4.0e-5	kPa^{-1}
Water compressibility	3.3e-6	kPa^{-1}
Oil density	847.00	Kg/m^3
Water density	1000.0	Kg/m^3
Oil viscosity	0.92	*cp*
Water viscosity	1.00	*cp*
Background *Vp*	3.5	*Km/s*
Background density	2100	Kg/m^3
Oil bulk modulus	1.57e+9	*Pa*
Water bulk modulus	2.25e+9	*Pa*
Porosity	0.20	*fraction*

Table 4.5: Rock and flow parameter values for the 2D seismic history matching test.

In this numerical example, the integrated seismic history matching workflow constitutes an inverse problem. It is formulated into an optimization framework in which the best-fitting reservoir model is targeted by minimizing the objective function defined in equation (4.31) as:

$$Err(\mathbf{s},\mathbf{p},\mathbf{c},\mathbf{u},\mathbf{q}) = \sum_{i=1}^{T}\left[\left\|\mathbf{w}_{s,i}\left(\mathbf{s}_i^d - \mathbf{s}_i\right)\right\|_2 + \left\|\mathbf{w}_{p,i}\left(\mathbf{p}_i^d - \mathbf{p}_i\right)\right\|_2 + \left\|\mathbf{w}_{c,i}\left(\mathbf{c}_i^d - \mathbf{c}_i\right)\right\|_2 + \left\|\mathbf{w}_{u,i}\left(\mathbf{u}_i^d - \mathbf{u}_i\right)\right\|_2 + \left\|\mathbf{w}_{q,i}\left(\mathbf{q}_i^d - \mathbf{q}_i\right)\right\|_2\right]$$

where **s**, **p**, **c**, **u** represent seismic, pressure, concentration and flux velocity vectors at different legacy times respectively; **q** denotes production data at well locations. Different weights are put on different terms, reflecting corresponding different confidences in them. In this example, the seismic data used stand for the post-stack seismic waveforms; the

production data stand for the oil-water ratio and cumulative oil production at the producing well location.

The flow simulator used in this test is the Integrated Parallel Accurate Reservoir Simulator (IPARS). IPARS is a three-dimensional general subsurface simulator that uses a multiphysics, multiphase flow reservoir model coupled with the petrophysical model and can be run serially or in parallel. It currently includes multiple physical models: single phase, two-phase oil-water and air-water, and a reactive transport, compositional model as well as the black-oil model. It also contains several numerical discretizations of these physical models, including mixed finite elements and discontinuous Galerkin finite elements for two-phase flow. In addition, different time discretizations have also been implemented in IPARS, including implicit, semi-implicit and sequential schemes. For more comprehensive descriptions of IPARS, please refer to Wheeler (2002) and Minkoff *et al*. (2003).

The Simultaneous Perturbation Stochastic Approximation (SPSA) method is performed as the optimization algorithm to minimize the error function defined in equation (4.31). Like the global optimization methods of Simulated Annealing and Genetic Algorithm, SPSA is a stochastic optimization algorithm, and it uses only an objective function measurement of multivariate systems. This contrasts with algorithms requiring direct measurements of the gradient of the objective function, which are often difficult or impossible to obtain. Formal theoretical and numerical comparisons of SPSA with other state-of-the-art optimization methods have been performed and have shown that SPSA to be competitive (Maryak and Chin, 2001; Spall, 2003). The essential feature

of SPSA, which accounts for its power and relative ease of implementation, is the underlying gradient approximation that requires only two measurements of the objective function regardless of the dimension of the problem. This feature allows for a significant decrease in the cost of optimization, especially in problems with a large number of variables.

In this test, SPSA is applied to iteratively fit time-lapse seismic data, well production data and the sensor measurements through 35 iterations and following the workflow shown in Figure 4.2. The resulting history of error function as in equation (4.31) is shown in Figure 4.4. As shown, the data integration leads to a quick convergence of the optimization process to a fairly good estimation. The final synthetic data compared against the true data are illustrated in Figure 4.5 for production data and Figure 4.6 for seismic data. In Figure 4.5, both the derived synthetic cumulative oil production data and the oil-water ratio data match the observed data quite well. This is not a difficult task because they are only evaluated in a single location, that is the production well location on the rightmost side. Figure 4.6 illustrates the comparison of synthetic and observed time-lapse seismic data for the model cross-section after three different time steps of 100, 500 and 1000 days. They generally match each other fairly well. It is also shown that the injected high-density water increases the seismic reflectivity and hence amplitude. This increase propagates laterally along with the injected water front over production/injection time.

In this test, multiple runs of the proposed integrated scheme are performed to derive multiple realizations of the reservoir model. For each run, the starting reservoir

model is randomly chosen within a certain bound. Figure 4.3 shows the true permeability distribution, the inverted mean permeability and corresponding standard deviation distributions based on the derived multiple reservoir model realizations. The inverted reservoir permeability model is clearly very close to the actual model, with only a small deviation.

As long as the reservoir permeability model is obtained as in Figure 4.8(b), it can be further used to run the flow simulation to image the reservoir's dynamic evolutions of fluid pressure and oil concentration as they have direct relationships to the reservoir's drainage and recovery strategy. As shown in Figure 4.7, the pressure is higher on the left side due to the water injection while it is lower on the right side due to the production. There exists a decreasing gradient between the injection and production wells. Evolution of oil concentration is illustrated in Figure 4.8 at three different legacy times. As shown, along with injection, the injected water moves laterally and pushes the oil rightward to the production well on the right side. Therefore, imaging of the water-oil contact is obtained, which indicates the inside dynamic fluid flow.

The derived dynamic distribution of fluid pressure and concentration can be further used as inputs to the Gassmann equation to derive the dynamic seismic features. Figure 4.9 shows the resulting P-impedance distributions at three different times of 100, 500 and 1000 days. As illustrated, the injected high-density water increases the P-impedance, and this increase propagates laterally rightward along with the water-oil front movement as shown in Figure 4.8. The observation of P-impedance evolution with time

is consistent with the inverted time-lapse seismograms at the same time steps as shown in Figure 4.6.

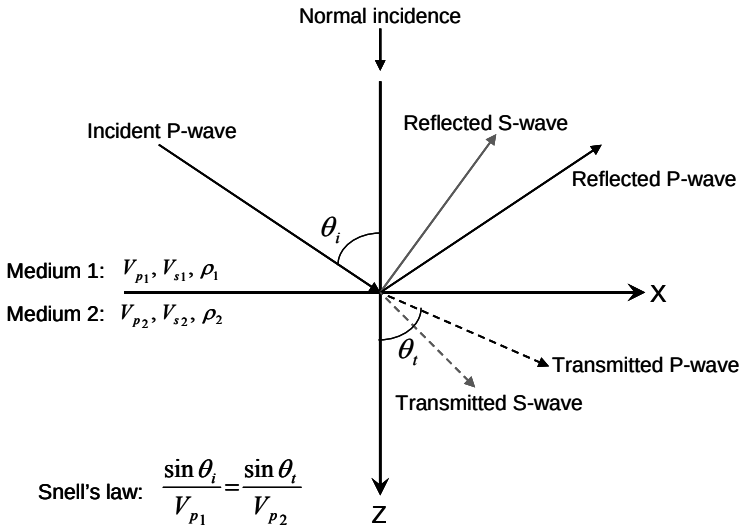

Figure 4.1: Graphic description of the reflection and transmission phenomenon of an incident plane P-wave at a planar interface (Z=0) which separates two different elastic media.

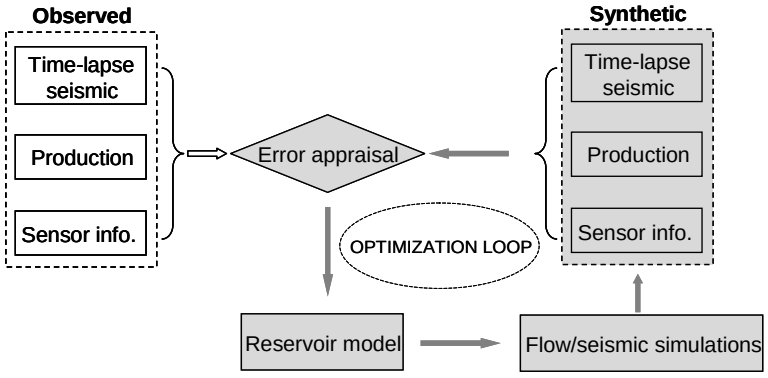

Figure 4.2: The proposed integrated workflow to simultaneously history match time-lapse seismic data, well production data and sensor information.

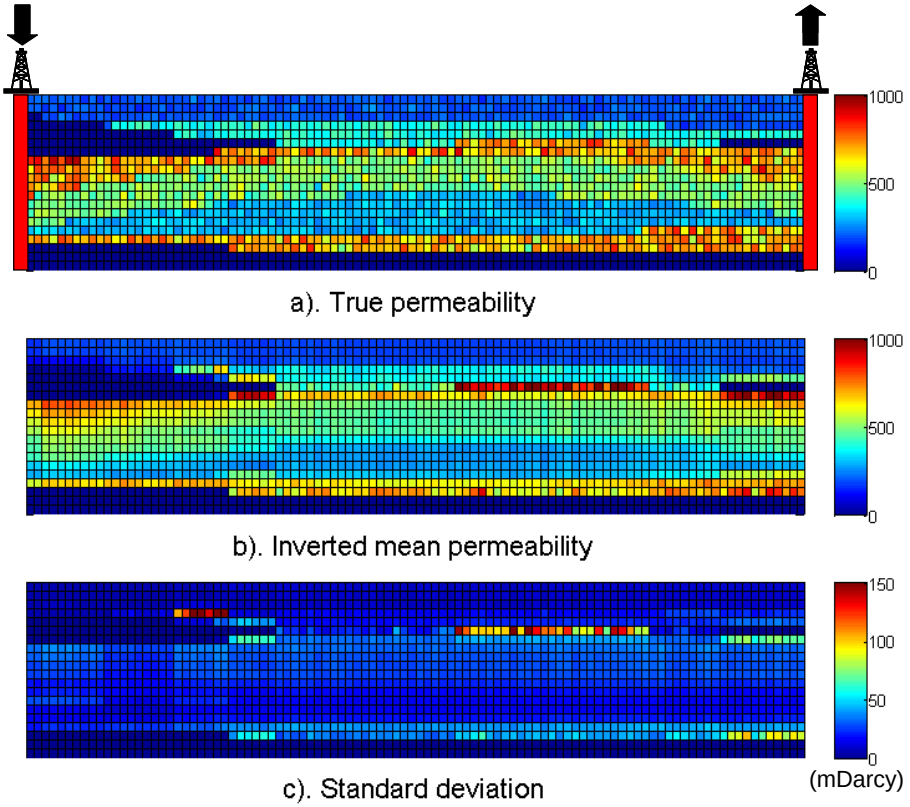

Figure 4.3: (a) True permeability model, (b) Inverted mean reservoir permeability model, and (c) Corresponding standard deviation. The well control strategy is also shown.

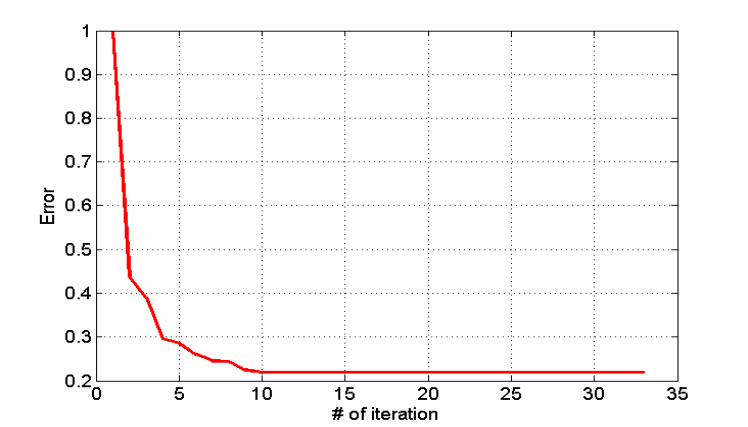

Figure 4.4: Normalized error history through 35 iterations. Integrating time-lapse seismic, well production data and sensor information leads to a fast convergence to a fairly good reservoir model estimation.

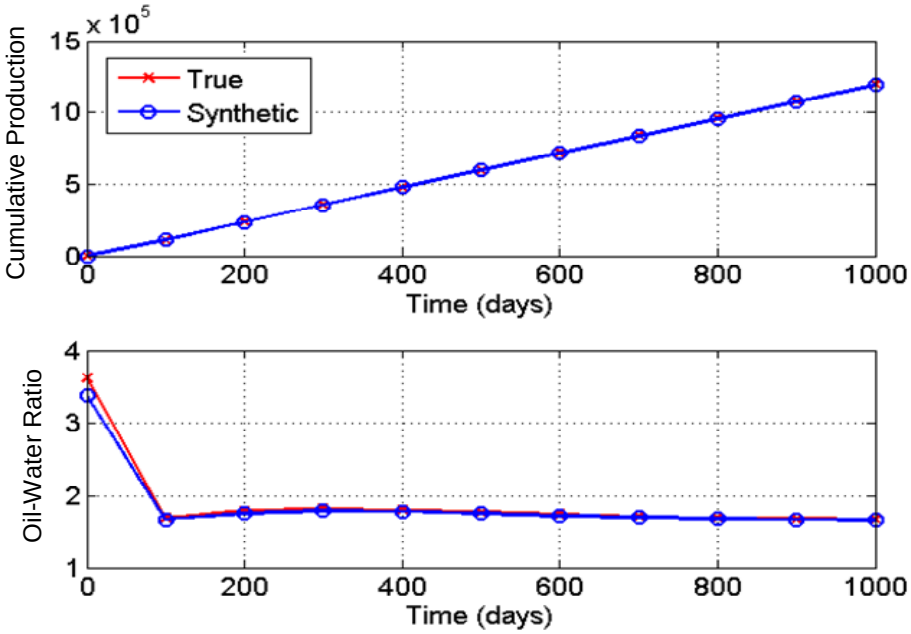

Figure 4.5: Comparison of observed and synthetic data fits of well cumulative production (STB) and oil-water ratio measurements.

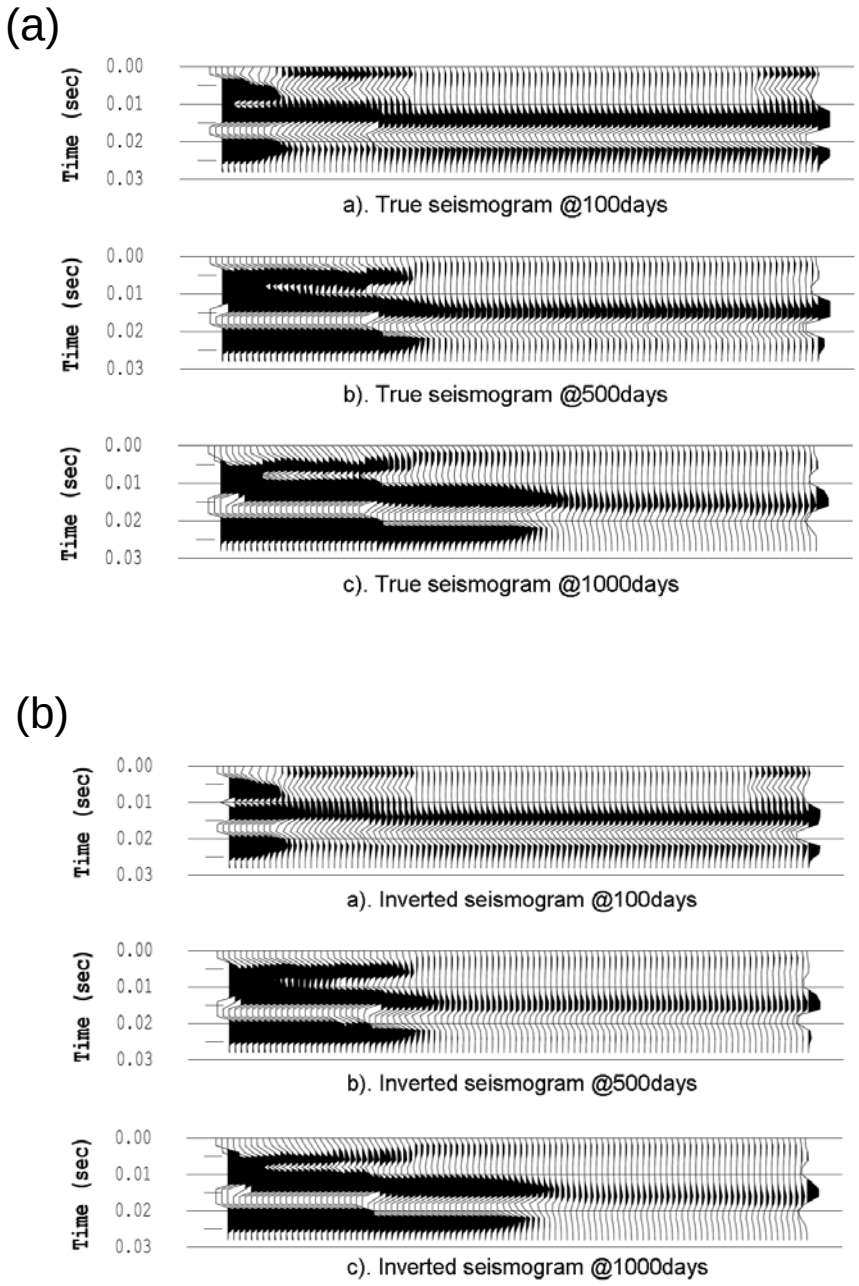

Figure 4.6: Comparison of (a) observed and (b) synthetic time-lapse data fits for the reservoir cross-section at three different legacy times.

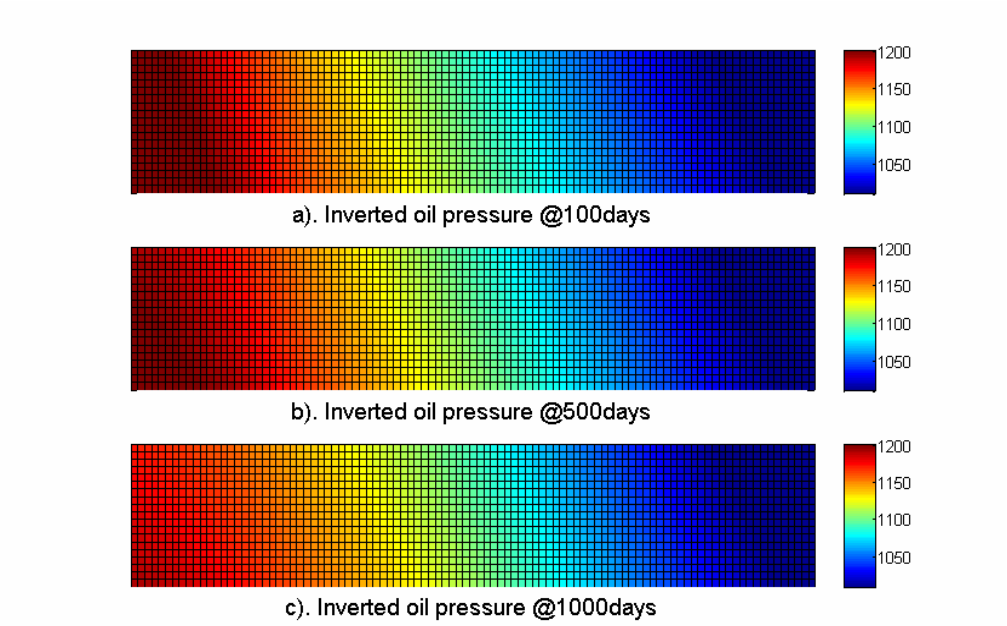

Figure 4.7: Distributions of fluid pressure (*kPa*) at three different monitor times. As shown, there exists a decreasing gradient between the injection and production wells.

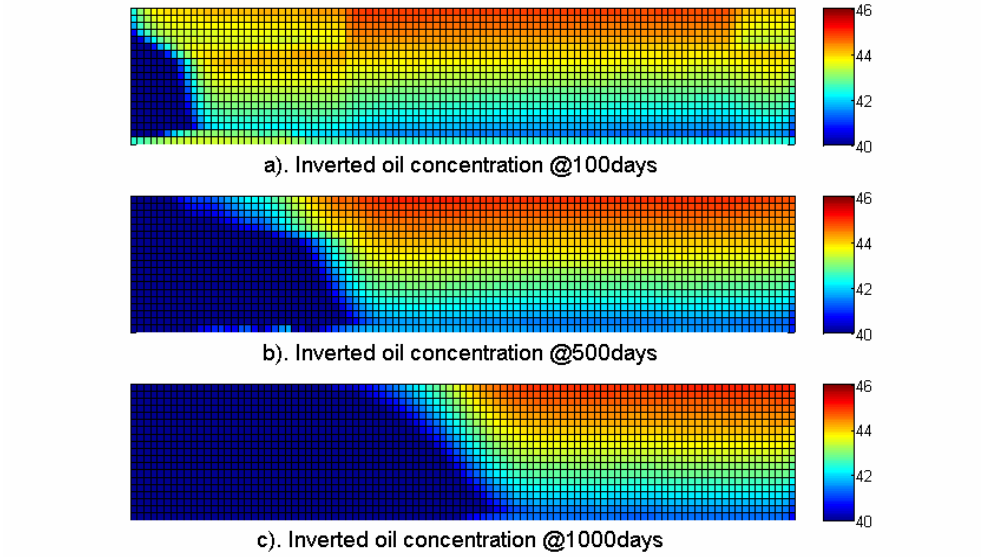

Figure 4.8: Distributions of oil concentration (%) at three different monitor times. As shown, the injected water moves laterally and pushes the oil rightward to the production well.

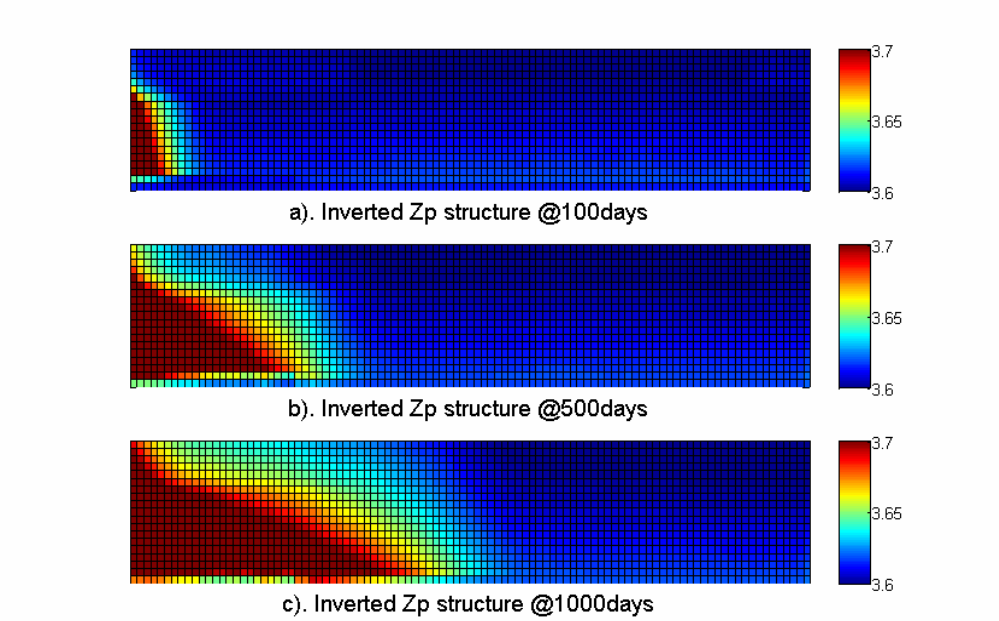

Figure 4.9: Distributions of the resulting P-impedance ($10^3 Kg/m^2 sec.$). The injected high-density water increases the P-impedance and this increase propagates laterally along with the water-oil front movement as in Figure 4.8.

Chapter 5

Integrated 4D seismic reservoir characterization and uncertainty analysis in a Bayesian framework

In petroleum reservoir appraisal, it is expedient to forecast future production performance under various recovery strategies and eventually to decide on a management strategy. The procedure for achieving this goal consists of two steps: first, establish a representation of the target reservoir characteristics based on available information, and second, simulate the future production evolution by running a numerical flow simulator on the derived reservoir characteristics.

In chapter 4, an integrated 4D seismic history matching workflow was proposed, in which the purpose is to derive a best-fitting reservoir model by quantitatively and simultaneously matching multiple datasets such as 4D seismic data and well production data. This optimization process constitutes a joint inverse problem, that is, a mapping from the joint data space to the model space. Although the incorporation of seismic data, which have better spatial resolution, in addition to the well production data, which have better local vertical resolution, helps to provide more constraints and reduce the inherent nonuniqueness, the number of data is nevertheless often less than the number of model parameters that need be quantified. Thus, the joint inverse problem is often under-determined and ill-posed, so it is almost impossible to deterministically solve it for a unique solution. From this point of view and in order to account for the uncertainty

associated with model estimation, a stochastic reservoir model may be defined, and the joint inverse problem may be formulated into a statistical framework and fully solved using a sampling algorithm.

To make the reservoir characterization and hence the future performance prediction as accurate as possible with reduced uncertainty, the stochastic reservoir model under study is often conditioned to two major types of information: (1) general reservoir information and (2) reservoir-specific observations (Tjelmeland, 1996; Hong *et al*., 2007; Hong and Sen, 2008b). The former consists of the static information as introduced in chapter 1, such as core data, well log data, geologic outcrop analog data, interpretation of horizons and faults, etc., which are often qualitatively used to construct a starting reservoir model. The latter consists of the dynamic observations on the target reservoir such as well production history and time-lapse seismic data, which can be quantitatively used to condition the starting model for an optimal estimation. However, since different data usually have different intrinsic resolution and correspond to different physics, the challenge is to integrate these dissimilar types of information, especially the diverse types of dynamic data sources, in an optimal manner for an accurate reservoir model characteristics and reliable production performance prediction. In addition, because this is a time-consuming iterative process, the issue of computational efficiency also needs to be specifically addressed.

Considering the unique feature of Bayesian inference in data integration and uncertainty handling, in this chapter, the integrated 4D seismic history matching problem or the joint inverse problem is formulated into a Bayesian framework and fully solved by

stochastically constructing the posterior distribution or PPD. The derived PPD describes all the consistent models along with corresponding likelihoods for the given joint datasets. Therefore, besides the most likely reservoir model, it also provides a convenient way to quantify corresponding uncertainty based on equations (1.6), (1.7) and (1.8). This requires that a sufficient sampling from the resulting PPD be performed, and the new multi-scale MCMC developed in chapter 3 is used as a sampling tool in this study.

In this chapter, a synthetic 3D reservoir, which is parameterized into both fine and coarse scales for computational efficiency, is used to justify the applicability and good performance of the proposed integrated Bayesian inference scheme. As will be illustrated, this scheme leads to a reasonably accurate estimation of the static petrophysical model and imaging of the inside dynamic evolutions. In addition, based on the PPD samples, corresponding uncertainty is also quantified in a theoretically correct way.

5.1 Integrated Bayesian reservoir characterization

Suppose the reservoir model **m** under study can be modeled as a random field in which the model parameters, such as porosity and permeability, are all random variables. As mentioned above, to quantify these variables, available information consists of general static knowledge and reservoir-specific dynamic measurements such as time-lapse seismic and well production data. Based on the former, a prior model can be constructed. This prior model is further conditioned to the seismic and production data to generate the posterior stochastic model and therefore to update the state of knowledge about the target model. In this way, the central point in the integrated seismic reservoir

characterization is to obtain a good description of the posterior properties, either analytically or using a sampling tool.

5.1.1 Likelihood functions

Suppose the 4D seismic data are denoted by $\mathbf{S}^{obs}$, and well production data are denoted by $\mathbf{P}^{obs}$. These observations are usually linked to the target reservoir model **m** through likelihood functions. For seismic data, the likelihood function corresponds to a conditional probability $p(\mathbf{S}^{obs}|\mathbf{m})$, which specifies the probability density function (PDF) of $\mathbf{S}^{obs}$ given that the reservoir model **m** is actually the true model. It can also be represented in a different form as: $\mathbf{S}^{obs} = g_s(\mathbf{m}) + \boldsymbol{\varepsilon}_s$, where $\mathbf{S}^{obs}$ is seen as a deterministic function of **m** plus a random error $\boldsymbol{\varepsilon}_s$ which accounts for the fact that the recorded observations are contaminated by noise; $g_s(\mathbf{m})$ is the so-called forward operator and represents the expected seismic response recorded by acquisition equipments to a given reservoir model **m**, which can be derived using an appropriate seismic wave propagation simulator. However, as described in chapter 4, seismic modeling cannot directly work on reservoir model parameters, so a rock physics model has to be incorporated to transform the flow parameters into the seismic elastic parameters. Correspondingly, $p(\mathbf{P}^{obs}|\mathbf{m})$ is the likelihood function for well production data, and $\mathbf{P}^{obs} = g_p(\mathbf{m}) + \boldsymbol{\varepsilon}_p$. The associated $g_p(\mathbf{m})$ is defined from the mass conservation law and Darcy's law. It represents the expected production measurement at well locations and can be derived using a flow simulator directly on the reservoir model parameters.

In practice, it is very difficult to represent the likelihood function in the form of a conditional probability as it is very difficult to know the intrinsic error distribution of the data. To handle this issue, an appropriate probability distribution is usually assumed for the data. Of course, identical distribution or different distributions may be applied to various data types. For example, if the observation error distribution is assumed to be Gaussian with zero mean, then the PDF of the error can be written as:

$$\begin{aligned} p(\boldsymbol{\varepsilon}) &\propto \exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}^T \mathbf{C}_D^{-1}\boldsymbol{\varepsilon}\right) \\ &\propto \exp\left(-\frac{1}{2}(\mathbf{d}_{syn}-\mathbf{d}_{obs})^T \mathbf{C}_D^{-1}(\mathbf{d}_{syn}-\mathbf{d}_{obs})\right), \end{aligned} \tag{5.1}$$

where $\mathbf{C}_D$ is the observation error covariance matrix, which defines the correlation among noise. For the most general case, $\mathbf{C}_D$ is usually not a diagonal matrix and the off-diagonal elements are not all zeros. $\mathbf{d}_{syn}$ is the expected response to a given model $\mathbf{m}$, and it can be derived through the forward simulator $g(\mathbf{m})$. Since the intrinsic noise is random, the resulting synthetic data in the model $\mathbf{m}$ has to be random and can be described in a form of PDF as:

$$p(\mathbf{d}_{obs}\,|\,\mathbf{m}) \propto \exp\left(-\frac{1}{2}(g(\mathbf{m})-\mathbf{d}_{obs})^T \mathbf{C}_D^{-1}(g(\mathbf{m})-\mathbf{d}_{obs})\right) \quad . \tag{5.2}$$

Since each type of data contains different information about the reservoir property and corresponds to different physics, it is reasonable to assume them to be conditionally independent of each other for a given reservoir model. In this way, the overall likelihood function incorporating both seismic and production data can be represented as:

$$p(\mathbf{S}^{obs}, \mathbf{P}^{obs}\,|\,\mathbf{m}) = p(\mathbf{S}^{obs}\,|\,\mathbf{m}) \cdot p(\mathbf{P}^{obs}\,|\,\mathbf{m}), \tag{5.3}$$

in which the likelihood definition for different data types can be inserted. For integration of multiple data types that are not independent of each other, other alternative schemes may be used to replace equation (5.3) (Journel, 2002; Caers *et al*., 2006; Castro, 2007).

5.1.2 Prior model

Statistical solution of the ill-posed joint inverse problem usually starts with a prior stochastic reservoir model, which is often represented by a PDF $p(\mathbf{m})$. The prior model summarizes all available information about the target reservoir before the reservoir-specific seismic and production data come into play. Hence, they are based on general knowledge of the reservoir and its fluid flow processes. This general reservoir information typically comes from the static data, such as geological setting, analogue observations and experience from neighboring comparable reservoirs. It is crucial for the reservoir model derivation that $p(\mathbf{m})$ gives a realistic description of the prior uncertainty about the model $\mathbf{m}$. Although the starting reservoir model can be relatively easily constructed by using a hierarchical workflow such as the one in Caers (2005), it is still challenging to translate this qualitative information into a probability distribution.

To date, most previous studies assume the random reservoir model to be Gaussian, so the PDF of the reservoir model can be written as,

$$p(\mathbf{m}) \propto \exp\left(-\frac{1}{2}(\mathbf{m}-\mathbf{m}_{prior})^T \mathbf{C}_M^{-1}(\mathbf{m}-\mathbf{m}_{prior})\right), \qquad (5.4)$$

where $\mathbf{m}_{prior}$ is the estimated prior reservoir model based on the static data and following a hierarchical workflow; $\mathbf{C}_M$ is the model variable covariance matrix, which is usually established through geostatistical tools.

On the other hand, if we simply know the lower and upper bounds of the model variables, it may be a good choice to use a uniform distribution to encode the prior information. The lower and upper bounds are not difficult to obtaine by satisfying certain physical criteria. Compared to other distributions, it is believed that the uniform distribution helps to avoid more biased constraints to be incorporated (Curtis and Lomax, 2001). In this way, the prior PDF is given as:

$$p\ (\mathbf{m}\ ) \propto \prod_{i=1}^{N} \frac{1}{m_{\max i} - m_{\min i}}, \qquad (5.5)$$

where m_{min} and m_{max} are the lower and upper bounds of the ith model variable; N is the total number of model variables.

5.1.3 Posterior model

With the likelihood functions and prior model defined, the posterior distribution, which is the updated prior distribution conditioned to the available observations, can be expressed as

$$p(\mathbf{m} | \mathbf{S}^{obs}, \mathbf{P}^{obs}) \propto p(\mathbf{S}^{obs} | \mathbf{m}) \cdot p(\mathbf{P}^{obs} | \mathbf{m}) \cdot p(\mathbf{m}) \qquad (5.6)$$

by assuming $\mathbf{S}^{obs}$ and $\mathbf{P}^{obs}$ to be conditionally independent for the given reservoir model $\mathbf{m}$. Analytical treatment of this posterior distribution is feasible only in very few cases. On the other hand, Bayesian inference approaches are frequently used in situations where

information is incomplete and a deterministic solution cannot be obtained (Gregory, 2005). In a Bayesian framework, the posterior distribution of the reservoir stochastic model conditioned to conditionally independent datasets of seismic and production observations is shown in equation (1.5), where the denominator $p(\mathbf{S}^{obs}) \cdot p(\mathbf{P}^{obs})$ acts as a normalizing factor to guarantee that the derived posterior probability is between 0 and 1. This normalizing factor is therefore the integration of the numerator over the entire model space, which poses challenges in realistic applications due to high-dimensionality. For example, for the problem of reservoir characterization, the stochastic reservoir model **m** is of extremely high dimension because of its highly heterogeneous nature. The corresponding likelihood functions are normally very complex and require solving a large set of differential equations. Hence, reliable solutions to these equations are computationally expensive, and they entail tremendous computer resources.

Instead of struggling with deterministic approaches for the derivation of posterior distribution, a sampling tool may be used to sample from the posterior distribution. Tools such as the maximum a posterior (MAP) estimation and uncertainty quantification can be used to facilitate further processing. Various sampling algorithms can be used to sample from the posterior PDF as described in More *et al*. (1999). In this study, the new multi-scale MCMC developed in chapter 3 is applied to sample from the PPD. In addition, the Bayesian inference approach has a unique capability in dynamic data integration. With another new dataset added, the derived PPD may be taken as a new prior for further updating and for generating a new PPD additionally conditioned to the new set of data. This technique is very useful in reservoir evaluation because it is a dynamic process with

frequent updates of the posterior stochastic model when more reservoir-specific measurements become available through infill drilling, additional seismic surveys and current well production history. Sometimes, the MAP is of primary interest, and it can be obtained by minimizing the derived posterior distribution.

From the perspective of reservoir decision making and management, determined values of the reservoir model variables are usually open to uncertainties considering that the results of the joint inverse problem are non-unique. Thus, it is also very important to quantify uncertainties associated with reservoir model estimation and future performance prediction. This can be done based on the model realizations through the procedure of posterior distribution sampling. Typically, these realizations are used to derive the model covariance $\mathbf{C}_M$ as well as to compute the histograms of model variables of interest. These histograms are then taken as the estimated marginal posterior distributions for particular variables, which leads to the most likely model parameter values and the corresponding uncertainty bounds. Finally, the obtained model realizations can be fed into a reservoir flow simulator under a certain recovery strategy to generate realizations of future production observations. Based on these realizations of future production, the most probable performance prediction and its associated uncertainty can be derived.

5.2 MCMC sampling from the PPD

The posterior distribution shown in equation (1.5) is the complete solution of the joint inverse problem. It describes all the consistent reservoir models conditioned to data $\mathbf{S}^{obs}$ and $\mathbf{P}^{obs}$. The denominator is an integration of the numerator over the entire model

space. For most problems of seismic reservoir characterization, this integral cannot be determined analytically over the multivariate, spatial random variable **m**. In addition, the likelihood functions usually are highly non-linear and iterative computation of them is hence extremely challenging in terms of the computational cost. Therefore, it precludes direct sampling from the PPD.

Sampling from a PPD can be efficiently done whenever the PPD can be factorized into lower dimensional PDFs, preferable one-dimensional ones, in that some sequential algorithms may be performed (Gomez-Hernandez and Journel, 1993; Omre *et al*., 1993). However, this requires strong constraints and assumptions so that it is only applicable in very particular cases. Sampling from a complex PPD has been of intensive study in the statistical community for decades. MCMC techniques based on the Metropolis-Hastings rule have been the central topic. Although application of MCMC in seismic history matching is very resource demanding, sampling from the correct posterior distribution is ensured in the limit, and therefore calculation of normalization integration is avoided. As briefly described in chapter 1, all the MCMC algorithms apply an iterative process to draw samples. These samples are then used to approximate the target PPD, in which each iteration contains a proposal and an acceptance/rejection step. In this study, the developed new multi-scale MCMC (see chapter 4) is performed as the sampling tool to estimate the PPD. The resulting samples are also further analyzed for uncertainty analysis by estimating the marginal PDF for individual model variables.

It is very important in an MCMC application to decide when a sufficient number of samples have been obtained and a sufficiently good estimation has therefore been

reached. One approach is to find the theoretical bounds on the number of iterations necessary to go through within a given distance from the specified distribution (Meyn and Tweedie, 1994; Rosenthal, 1994). However, thus far, no good bounds seem to exist for the complex models necessary in integrated reservoir characterization. An alternative method frequently used in complex models is output analysis. In output analysis, important univariate characteristics of the realizations are plotted against the number of iterations until they seem to have been stabilized statistically (Ripley, 1981). The latter approach is also used in this study, that is, the multi-scale MCMC is performed to continuously draw samples until the resulting marginal PPDs of certain variables do not change.

5.3 Example: Bayesian inference of reservoir properties and uncertainty analysis

In this section, a numerical example is presented to derive reservoir characteristics by integrating 4D seismic data and well production data. A stochastic 3D reservoir is defined and an integrated seismic history matching workflow illustrated in Figure 5.1 is followed for reservoir model parameter estimation and corresponding uncertainty quantification. In this workflow, a reservoir model is found such that, after both flow simulation and seismic modeling, the synthetic seismic and production data are matched against the recorded well data. If the match is poor, the reservoir model is stochastically perturbed to improve the data fit. As introduced earlier, it is very important in reservoir characterization to use a realistic prior description. This is usually done using a

hierarchical scheme, and most previous researchers assumed the random reservoir field to be Gaussian as the prior constraint. This is a more or less overly strong constraint and may give rise to artificial bias in the final model estimation. In this example, a good trial is that, instead of assuming a Gaussian random field, the reservoir model space is specified based on the roughly determined upper and lower bounds on model variables. Thus, the starting reservoir model is just randomly picked from the pre-defined model space, which relaxes the stringent constraint and does not influence the quality of the final model estimation as a convergence–guaranteed algorithm is used.

The integrated objective function used to measure the mismatch between synthetic and observed datasets is given as

$$Err\ (\mathbf{m}) = \sum_{t=1}^{T} \left(W_{seis} \cdot \left\| \mathbf{S}_t^{obs} - \mathbf{S}_t^{syn} \right\|_2 + W_{prod} \cdot \left\| \mathbf{P}_t^{obs} - \mathbf{P}_t^{syn} \right\|_1 \right), \qquad (5.7)$$

where $\mathbf{S}_t^{obs}$ and $\mathbf{S}_t^{syn}$ respectively represent observed and synthetic seismic data at time step t; $\mathbf{P}_t^{obs}$ and $\mathbf{P}_t^{syn}$ respectively represent observed and synthetic production data at time step t. There are T legacy time steps in total. Different weights W_{seis} and W_{prod} are put on the two normalized terms to reflect corresponding confidence on them. If W_{seis} is set to zero, only production data are used and this corresponds to the conventional history matching problems. If W_{seis} is set to zero, only the time-lapse seismic data are used, and this corresponds to a pure 4D seismic inversion problem. In the integrated 4D seismic history matching, neither of them is zero. As noted, an L_2 norm is used to measure the mismatch between synthetic and observed 4D seismic data, whereas an L_1 norm is used to measure the mismatch between synthetic and observed production data. Other

researchers incorporate some other terms, called “regularization terms” or “smoothing terms”, in the objective function to impose extra constraints in order to reduce non-uniqueness. Nevertheless, in this study, this is not the case. Only the two terms of data fit are included in the objective function because it will be shown that, by using a multi-scale MCMC algorithm, the coarse-scale model works like a smoothing operator to smooth the fine-scale model and make it more realistic.

In this study, the proposed integrated 4D seismic history matching workflow (Figure 5.1) is formulated into a Bayesian framework. The posterior distribution (PPD) of the stochastic reservoir model is estimated based on samples drawn by the new multi-scale MCMC method.

5.3.1 Reference model and model parameterization

In this synthetic example, the 3D reservoir contains oil and water and has a regular grid system with a dimension of 20x20x5 over an area of 1600x1600x50 ft^3 (Figure 5.2(a)). The two phase black-oil model is used to describe the inside oil and water. Reservoir permeability is assumed to be a constant distribution, and porosity is taken as the primary unknown. Hence, each grid block contains only one stochastic variable, the porosity. The 3D porosity in Figure 5.3, which is shown in the representation of a 3D cube and slices as well, is assumed to be the actual distribution. It is used in this example for comparison in order to evaluate the quality of the final model estimation and to appraise the performance of the new multi-scale MCMC algorithm developed in chapter

3. The porosity value ranges between 0.05 and 0.35 through the entire reservoir. Porosity distributions for three of the five layers are shown in Figure 5.4, in which the white dots represent the five well locations. The initial water and oil saturations are assumed to be constant at 0.19 and 0.81, respectively. Other rock and flow parameters are also assumed to be constant and are shown in Table 5.1.

PARAMETER	VALUE	UNIT
Initial pressure	1300.0	*psi*
Oil compressibility	1.0e-5	psi^{-1}
Water compressibility	3.2e-6	psi^{-1}
Oil density	52.88	lbm/ft^3
Water density	62.40	lbm/ft^3
Oil viscosity	0.92	*cp*
Water viscosity	1.00	*cp*
Background *Vp*	3.5	*Km/s*
Background density	2100	Kg/m^3
Oil bulk modulus	1.57e+9	*Pa*
Water bulk modulus	2.25e+9	*Pa*
Radius of well	0.3	*ft*
Skin factor	0.0	-

Table 5.1: Rock and flow parameter values that are assumed to be constant throughout the production history.

Figure 5.2 also shows the fixed well control strategy, that is, a water injection well is in the middle and four symmetrical oil production wells are around the reservoir. The reservoir is depleted by a constant water injection rate of 1780STB/D in a well at the

lateral grid (10, 10) and oil production with a constant flow rate of 400STB/D respectively in wells at the lateral grids (5, 16), (16, 5), (5, 5) and (16, 16). The production history is for 700 *days*, which is discretized into 7 legacy time steps at (30, 100, 300, 400, 500, 600, 700) *days*. The production data used in this example represent the water and oil flow rates observed at the four production wells; the seismic data used comprise the post-stack seismic traces, which means that not only the amplitude and vertical two-way travel but also the entire shape of a trace seismogram is used to condition the reservoir model evaluation.

To generate 4D seismic data and well production data and use them as the "true" observations, the reference 3D porosity model in Figure 5.3 is used as the input to initiate the flow simulation and hence to compute the water and oil flow rates in the four production wells at the seven time steps. The generated true production data are shown in red in Figure 5.5. Outputs of porosity, fluid saturation and pore pressure distributions are used as inputs to the Gassmann equation introduced in chapter 4 to derive the corresponding elastic parameter (acoustic impedance) distributions at the seven legacy time steps. The generated true post-stack seismic waveforms for a particular cross-section (X=10) and respectively after 30, 300 and 700 *days* are shown in Figure 5.6.

Although the observed data are only of a single scale, a coarse-scale reservoir model of 10x10x5 is incorporated in addition to the fine-scale model of 20x20x5 in this example in order to apply the multi-scale MCMC. Compared to the fine-scale setting, the same well control strategy is applied except that the injection well is changed to the lateral grid of (5, 5) and the production wells are changed to grids of (3, 8), (8, 3), (3, 3)

and (8, 8) (Figure 5.2(b)). In this way, for a particular layer, four neighboring fine-scale grid blocks correspond to a single coarse-scale block, as shown in Figure 5.2. Therefore, during the multi-scale MCMC sampling, if it attempts to trade information between the fine and coarse reservoir models, it actually attempts to exchange the parameter values of the single coarse block with the average parameter value of the four fine blocks. Thus, a single coarse block works like a regularization operator to smooth the corresponding four fine-scale blocks.

5.3.2 Bayesian formulation

To derive the PPD based on available data, a prior distribution is required. The most often assumed prior in reservoir characterization is Gaussian distribution, as shown in equation (5.4). In this example, the lower and upper bounds on the model variable, that is porosity, are known, and they are 0.05 and 0.35, respectively. This may be used as a loose constraint to narrow the model space and also to construct the prior distribution. Supposing that every point in the model space has an equal probability to be drawn as the reservoir stochastic model, the resulting uniform prior distribution is given in equation (5.5).

The choice of likelihood functions relates to the forward processes and depends on the distributions of the noise or error in the data. Based on the Central Limit Theorem in statistics, it is appropriate the Gaussian error distribution in seismic data (Gregory, 2005). As shown in the objective function in equation (5.7), an L_2 norm is proposed to

measure the mismatch for seismic part. The correlation relationship of equation (2.4) is used again in here as the L_2 norm to measure the fitness between synthetic and observed seismic data, which is described as

$$Fit\_seis(\mathbf{m}) = \frac{\mathbf{S}^{obs} \otimes g_s(\mathbf{m})}{\left(\mathbf{S}^{obs} \otimes \mathbf{S}^{obs}\right)^{1/2} \cdot \left(g_s(\mathbf{m}) \otimes g_s(\mathbf{m})\right)^{1/2}}. \tag{5.8}$$

Basically, fitness is one minus error. Thus, for the multi-scale MCMC application, the resulting likelihood function for the seismic part is expressed as follows

$$p(\mathbf{S}^{obs} | \mathbf{m}) \propto \exp\left(\frac{1}{2}(g_s(\mathbf{m}) - \mathbf{S}^{obs})^T \mathbf{C}_{sD}^{-1}(g_s(\mathbf{m}) - \mathbf{S}^{obs})\right), \tag{5.9}$$

where $\mathbf{C}_{sD}$ is the seismic data error covariance matrix, and g_s is the seismic forward simulation operator.

As illustrated in equation (5.7), an L_1 norm is used to measure the mismatch of production data. Accordingly, to use the multi-scale GA based MCMC, the fitness is calculated based on an expression as

$$Fit\_prod(\mathbf{m}) = 1 - \frac{2 \cdot \sum_{i=1}^{T}\sum_{j=1}^{N}\left|y_{i,j} - x_{i,j}\right|}{\sum_{i=1}^{T}\sum_{j=1}^{N}\left|y_{i,j} + x_{i,j}\right| + \sum_{i=1}^{T}\sum_{j=1}^{N}\left|y_{i,j} - x_{i,j}\right|}, \tag{5.10}$$

where T stands for the index of legacy time steps; N stands for the index of model grid block; and $y_{i,j}$ and $x_{i,j}$ represent observed and synthetic data, respectively. As L_2 norm bears the same relationship to Gaussian distribution, L_1 norm bears the same relationship to exponential distribution (Menke, 1984). Therefore, the resulting likelihood function by using equation (5.10) for production data is

$$p(\mathbf{P}^{obs} | \mathbf{m}) \propto \exp\left(\left|g_p(\mathbf{m}) - \mathbf{P}^{obs}\right|\right), \tag{5.11}$$

where g_p is the flow forward simulation operator.

Using the uniform prior distribution, Gaussian distribution for seismic likelihood function and exponential distribution for production likelihood function, the resulting posterior distribution at a certain scale i is

$$p^{(i)}(\mathbf{m}^{(i)} \mid \mathbf{S}^{obs}, \mathbf{P}^{obs}) \propto \exp\left(\frac{1}{2}\left(\mathbf{S}^{obs} - g_s(\mathbf{m}^{(i)})\right)^T C_{sD}^{-1}\left(\mathbf{S}^{obs} - g_s(\mathbf{m}^{(i)})\right)\right) \cdot \exp\left(\left|g_p(\mathbf{m}^{(i)}) - \mathbf{P}^{obs}\right|\right) \cdot \prod_{j=1}^{N} \frac{1}{m_{\max j}^{(i)} - m_{\min j}^{(i)}} \quad . \quad (5.12)$$

Again, the solution models at different scales are assumed to be conditionally independent of each other for the given data. The nonlinear forward operators g_s and g_p usually lead to a highly multi-modal posterior distribution with a very complicated topology of many hills and valleys. This creates large challenges in applications and calls for a powerful MCMC method in order to sufficiently explore the posterior within an acceptable CPU time period.

5.3.3 Static petrophysical model and uncertainty quantification

In this example, the fine-scale 20x20x5 model is of primary interest because it provides more reliable information about the target reservoir. The main reason for additionally incorporating the coarse-scale 10x10x5 model is to yield a more tractable posterior distribution and a faster forward simulation. This allows more MCMC updates per unit CPU time and helps to sufficiently explore the posterior. In addition, for the coarse scale, stochastic update steps over the model space are usually large, which helps

to avoid the sampling process being trapped at a certain local optimum. In the implementation, 8 MCMC chains in total are run through 2500 generations, that is, four chains are defined at either of the two scales. By iteratively fitting the 4D seismic data and well production data, the multi-scale MCMC is applied to update the current models and propose new realizations, which are accepted or rejected based on the Metropolis-Hastings rule until a certain stopping criterion is reached. Through the 2500 generations, all the obtained model realizations of different scales are tracked and stored separately for further analysis.

5.3.3.1 Data fits

For the fine-scale case and after 2500 generations, the resulting synthetic production data (in blue) are plotted and compared against the true well production data (in red) in Figure 5.5. Excellent match is obtained between the true and synthetic production data, that is, water and oil flow rates at the four production wells. This very good production data fit is relatively easy to obtain in a short time as they are evaluated only at sparse well locations. From the fine-scale model, the finally obtained synthetic seismic data for the same particular cross-section (X=10) and after the different time periods used in Figure 5.6 are shown in Figure 5.7 for evaluation. By comparing Figure 5.6 to Figure 5.7, it is found that a reasonably good seismic data fit is also obtained. Also, by carefully assessing the seismic data for a particular cross-section during the reservoir production time, it is found that the injected high-density water increases the amplitudes of the seismogram. This increase propagates laterally along with the injected and laterally

moving water. Corresponding P-impedance (acoustic) distributions for the same cross-section and after the same time steps are shown in Figure 5.8. The P-impedance is also changed by the production process, and this change directly leads to the amplitude changes observed in Figure 5.7.

Figure 5.9 shows the fitness histories through the 2500 generations for the two different scales. Both of them increase gradually, finally arriving at a steady state. As expected, fitness from the coarse-scale model is necessarily low, as it starts with a value of about 0.54 and finally rises to a slightly better value of less than 0.60. This low fitness is because the coarse-scale model parameterization is relatively too rough to capture detailed reservoir information. In comparison, the fine-scale model leads to higher fitness values starting at a point around 0.70 and finally reaching a very good value of larger than 0.85.

5.3.3.2 Porosity models

Through the 2500 generations of sampling, many MCMC samples of reservoir model realizations are obtained separately for the fine scale and the coarse scale. These samples are used to approximate the posterior distributions for the two different scales. Based on the estimated posterior distributions, the most likely model, also called the maximum a posteriori (MAP), can be further determined. MAPs are used to represent the derived reservoir models. Figure 5.10 shows the finally derived fine-scale model, and Figure 5.11 shows the finally derived coarse-scale model. Both are shown in a 3D cube as well as slices. As expected, the derived coarse-scale model does not provide reliable

reservoir description. However, additional incorporation of the coarse scale does help to speed up the convergence of the fine-scale model to a reasonably good estimation. In addition, by applying the multi-scale MCMC algorithm and periodically exchanging information between the two different scales, the coarse-scale model works like a regularization or smoothing operator to smooth the fine-scale model and make it more realistic. As shown in Figure 5.10 and compared to Figure 5.3, a fairly accurate reservoir model is obtained without much perturbation, which is the usual phenomenon if a regularization or smoothing term is not included in the objective function. Figure 5.12 shows the derived porosity distributions for three of the five layers. Although this derived model is not as smooth as the reference model in Figure 5.4, the distributions of high-porosity and low-porosity zones match fairly well. Therefore, it could be said that a fairly accurate static petrophysical model, that is, a reservoir porosity model, has been obtained.

To confirm the improvement in computational efficiency by using multi-scaling, an experiment is also done on the fine scale only, that is, all eight chains are defined on the scale of 20x20x5. In addition, a regularization term is incorporated in the objective function to do model smoothing, which is similar to the work of Li and Oldenburg (2000). The single-scaling result is shown in Figure 5.13. It is found that, in order to obtain the similarly good result of multi-scaling case as in Figure 5.12, the single-scale sampling takes at least twice as much total CPU time as the multi-scale sampling. This means that the computer resources have to be at least doubled. Hence, using multi-scaling and additionally incorporating the coarse scale improves the computational efficiency while retaining the estimation accuracy.

To test the effectiveness of integrating 4D seismic data and well production data in improving the accuracy and hence reducing uncertainty in reservoir characterization, different experimental setups are also tested. First, the multi-scaling MCMC sampling process is applied with only production data in use. In equation (5.7), this equals to set the weight on the seismic part to be zero. The finally derived porosity distributions for the three layers are shown in Figure 5.14. Although well data have better local resolution, they are only from sparse wells, and the corresponding data fit evaluations are only performed at the four far distributed production locations. Thus, they cannot provide reliable constraints on the reservoir's spatial heterogeneity. Therefore, the derived porosity distribution has large uncertainties and does not reliably reflect the actual reservoir characteristics. In addition, the case of only seismic data in use is also tested, which is equivalent to set the weight on the production term to be zero. The finally derived distributions are shown in Figure 5.15.

Compared to Figure 5.14, this scheme results in a better global characterization by taking advantage of the benefits from seismic data, that is, the extensive spatial coverage and dense lateral sampling. Nevertheless, this result is still less accurate with more uncertainty compared to the result in Figure 5.12, in which the seismic data and production data are integrated to simultaneously condition the reservoir characterization. In other words, this confirms that the integration of 4D seismic data and well production data leads to increased accuracy and reduced uncertainty in reservoir characterization.

5.3.3.3 Uncertainty quantification

Besides the mostly reservoir model, the estimated posterior distribution can also be used to quantify corresponding uncertainties. For this very high-dimensional problem, it is impossible to directly show the multivariate posterior graphically, the marginal posterior for particular parameters are usually computed and illustrated instead. Typically, the histogram of a particular parameter calculated from the obtained MCMC samples is used to estimate the marginal posterior for that parameter. For example, Figure 5.16 shows the computed histograms of four particular model parameters with the grid coordinates of (17, 7, 5), (3, 17, 3), (5, 5, 3) and (6, 13, 2), respectively. From those estimated PPDs, the most likely porosity values for the four different grid blocks that correspond to the maximum frequency can be determined. In addition, the corresponding uncertainty bounds can also be estimated. If this marginal PPD is estimated for all the reservoir model blocks, a 3D cube of deviation may be generated. Figure 5.17 illustrates the relative deviation cube with respect to the derived most likely model cube in Figure 5.10. For this estimated percentage deviation cube, three of the five layers are presented in Figure 5.18. As shown, although uncertainties are relatively large at some points, they mostly localize below 15%.

5.3.4 Dynamic pore pressure and fluid saturation

Reservoir pore pressure and fluid saturation changes are directly related to reservoir recovery process. Therefore, reliable estimation of these dynamic features

greatly influences reservoir decision making and management. Prediction of pore pressure is very important for planning and safety reasons when drilling new wells, particularly in deep-water reservoirs where operating costs are very high. Fluid saturation provides a direct indication of the inside fluid flow, so robust description of the saturation evolution will lead to well-founded future predictions of production performance and thus will increase ultimate recovery from the reservoir and reduce production costs.

The derived porosity model (Figure 5.10) can be further used as the input to trigger the flow simulator in order to image the dynamic evolutions of pore pressure and fluid saturation in the reservoir. Figure 5.19 illustrates the dynamic pore pressure distributions for a particular layer (Z=3) after 30, 300 and 700 *days*, respectively. Figure 5.20 shows the pore pressure distributions after the same three different time periods but for a particular cross-section (X=10). As shown, the overall pore pressure lowers gradually as the production progresses, and there are decreasing gradients between the water injection well and the four production wells. In addition, pore pressure around the producing wells is especially low due to the production but especially high around the injection well. The differential decreases in pore pressure for the specified layer and cross-section during certain production time periods are shown in Figure 5.21. Evolution of the water saturation for the same layer (Z=3) after 30, 300 and 700 *days* is also imaged in Figure 5.22. Corresponding evolution for a particular cross-section (X=10) is illustrated in Figure 5.23. As shown, the injected water basically propagates laterally, and the water front movement through production time is clearly imaged. Similarly, the

differential increases in water saturation for the specified layer and cross-section during certain production time periods are also shown in Figure 5.24.

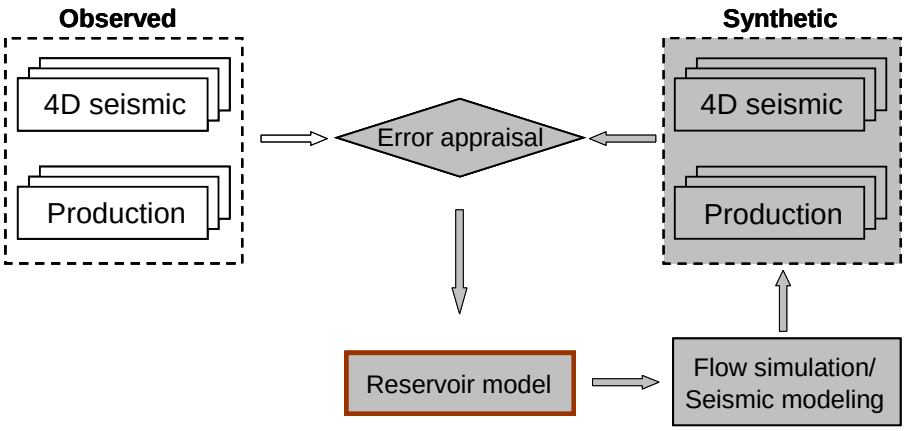

Figure 5.1: Integrated workflow used to derive reservoir model parameters by simultaneously integrating 4D seismic and well production data. The idea of seismic history matching is applied in this workflow and the starting model is usually constructed based on certain general static information about the reservoir.

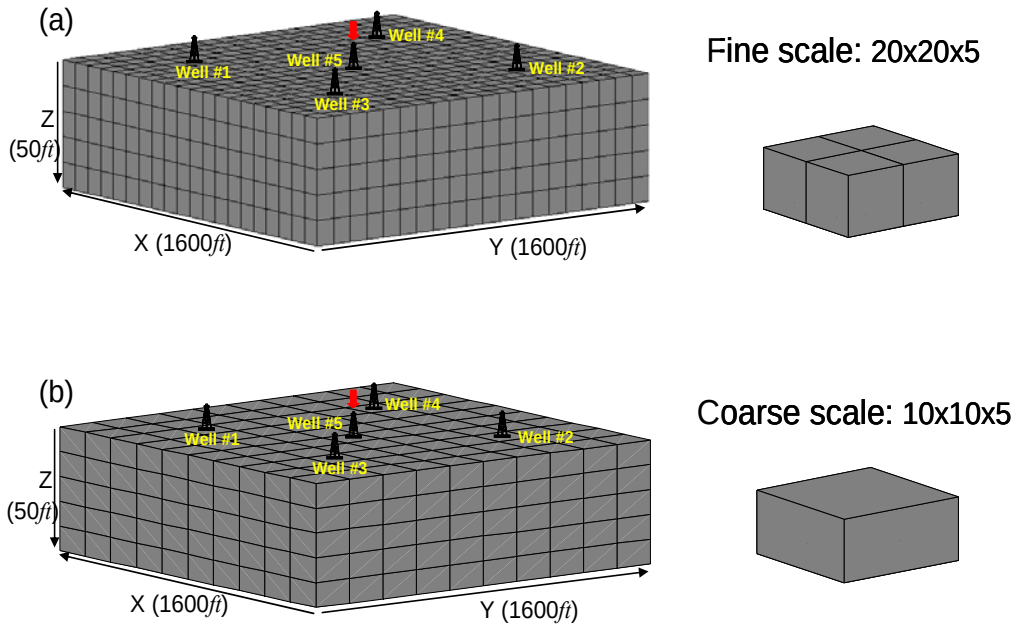

Figure 5.2: 3D reservoir model used in the study, which is parameterized into two different scales of 20x20x5 and 10x10x5. In this way, for each layer, four neighboring fine-scale blocks correspond to a single coarse-scale block. The well control strategy is also illustrated, that is a water injection well in middle and four production wells around.

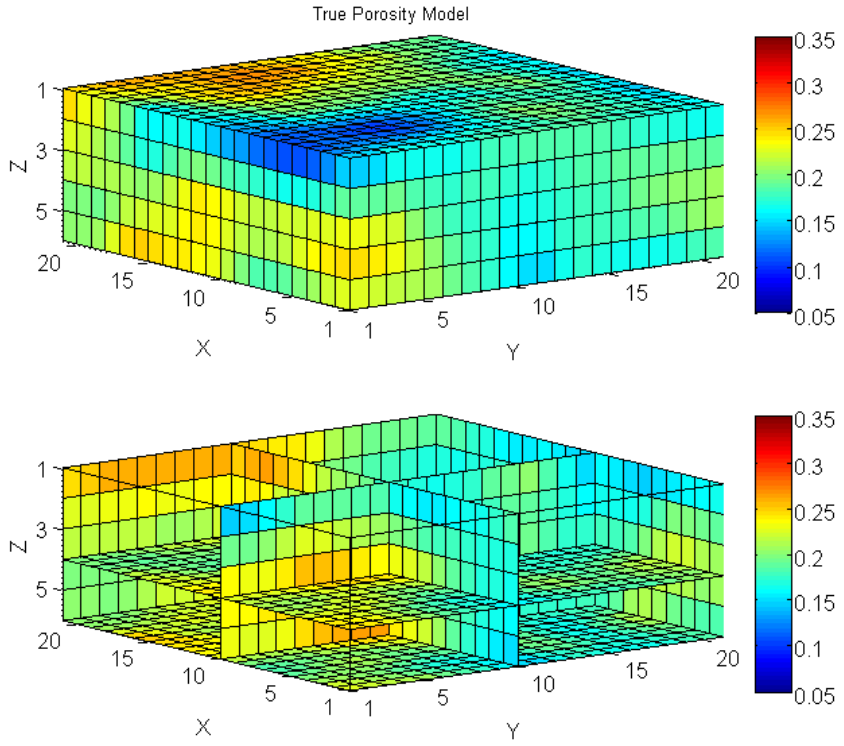

Figure 5.3: Reference porosity model that is assumed to be the actual reservoir model that is taken as the target to estimate. The model is shown in a 3D cube (upper) as well as slices (bottom). Through the reservoir, porosity value is between 0.05 and 0.35.

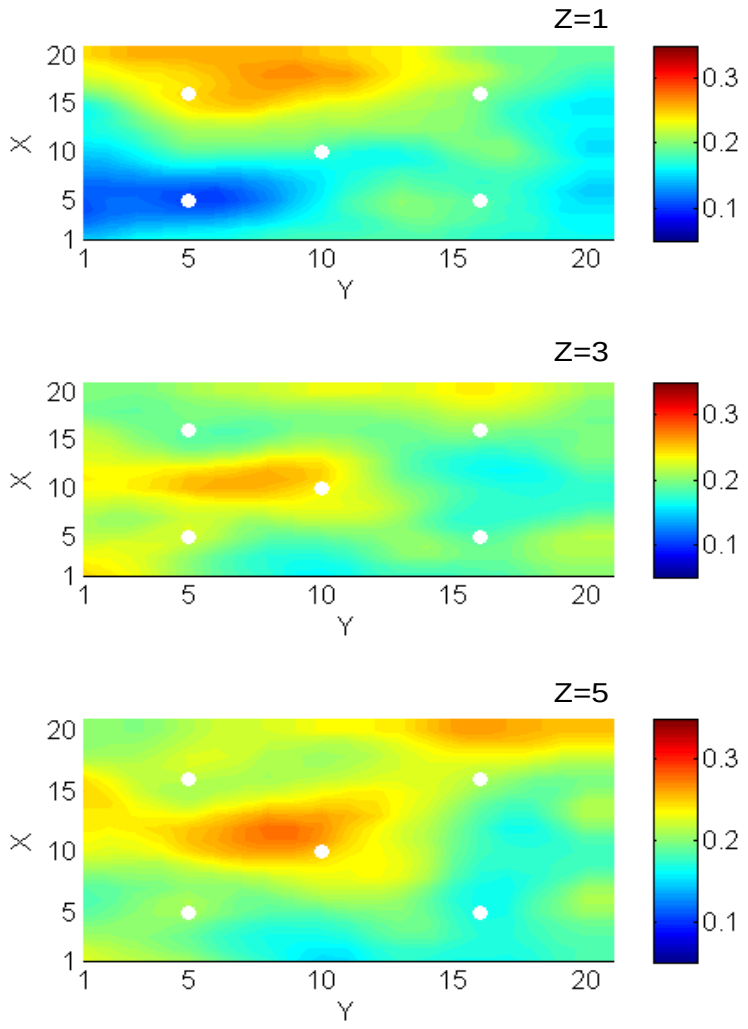

Figure 5.4: Reference porosity distributions for three different layers that are extracted from the 3D reference model in Figure 5.3. Some high porosity and low porosity zones are clearly shown.

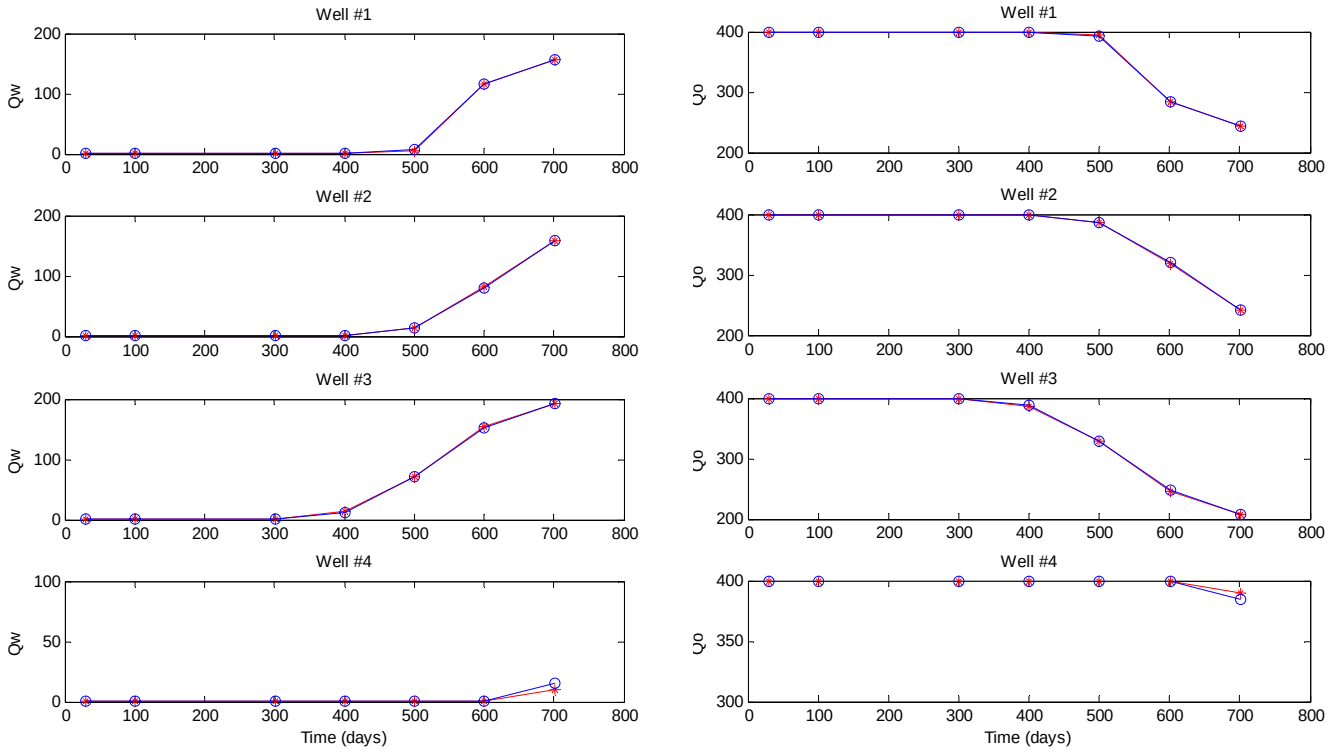

Figure 5.5: Observed (red) and synthetic (blue) well production data, that is, water (left) and oil (right) rates. As shown, an excellent match is obtained between the observed and synthetic data.

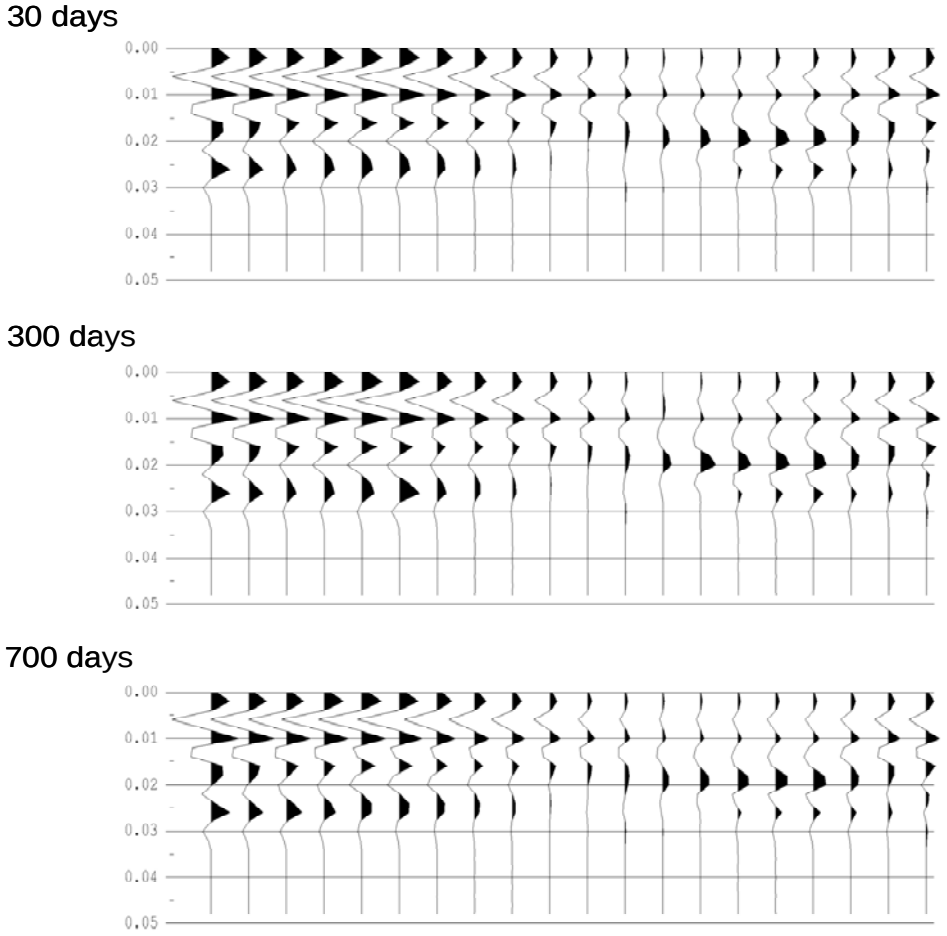

Figure 5.6: Observed time-lapse seismic data for a particular cross-section (X=10) after different time steps, namely, 30, 300 and 700 *days*.Amplitudes are changed by the producing process with high-density water injected in.

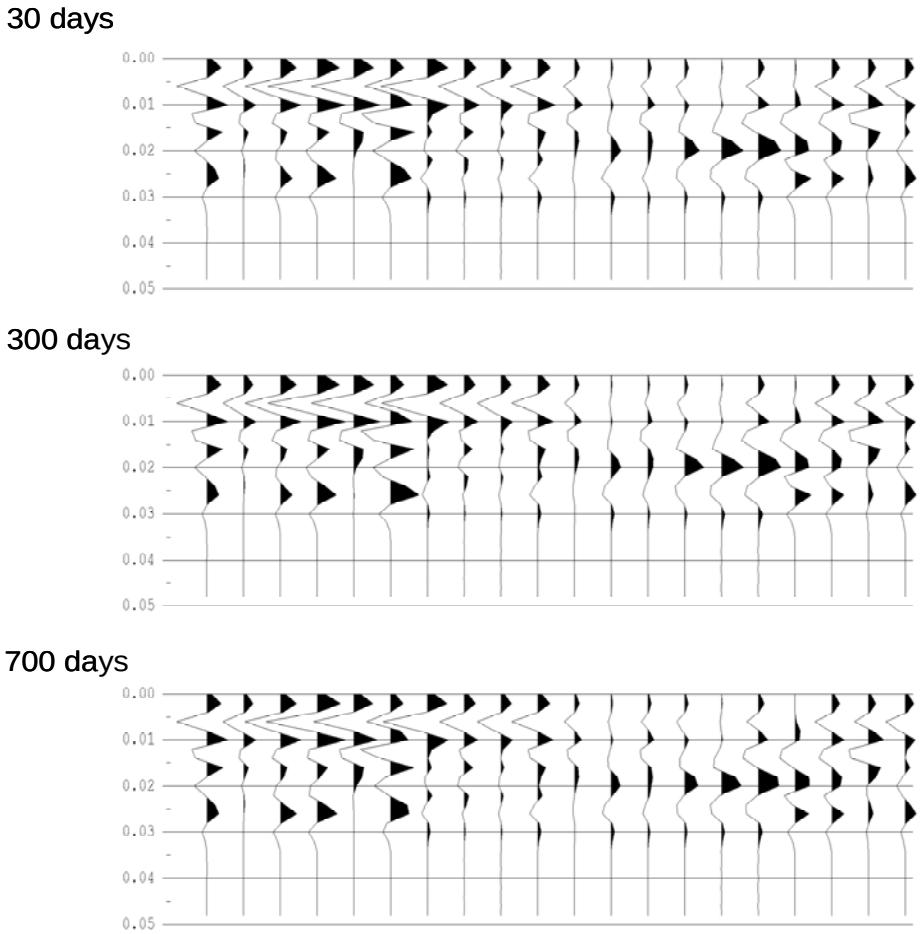

Figure 5.7: Synthetic time-lapse seismic data for a particular cross-section (X=10) after 30, 300 and 700*days* respectively. Compared to Figure 5.6, a fairly good match is obtained and this derived result also confirms the observation of increased amplitude by the injected high-density water.

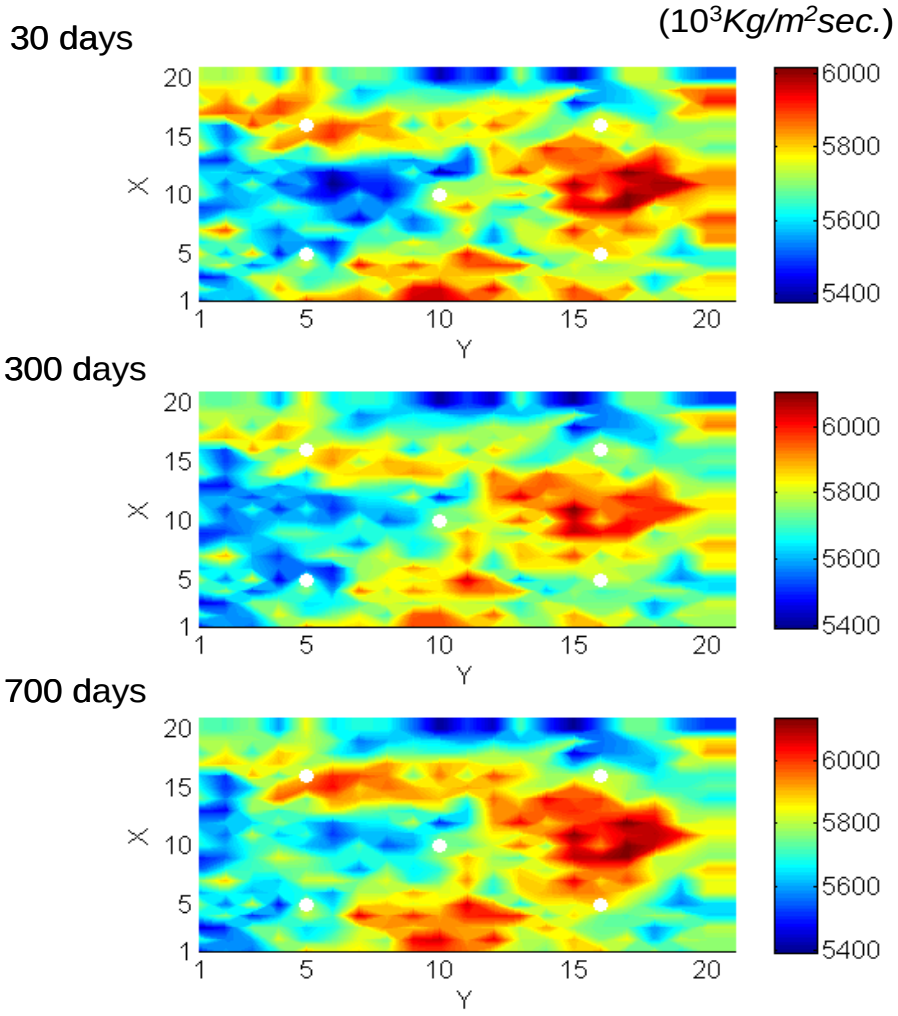

Figure 5.8: The distributions of P-impedance (acoustic) for the same cross-section and after the same different time periods as in Figure 5.7. As shown, P-impedance is changed by the production process over time, and this change directly leads to the amplitude changes observed in Figure 5.7.

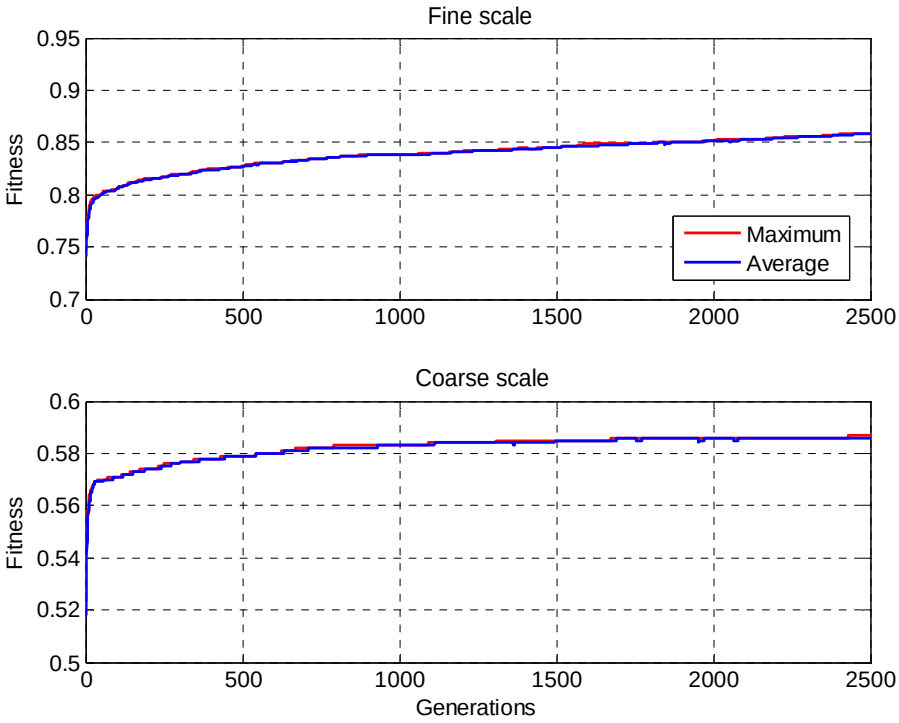

Figure 5.9: Fitness histories for both the fine-scale and the coarse-scale models. By iteratively fitting the seismic and production data, both the fitness increase gradually and finally converge to better values. Compared to the coarse scale, the fine scale model leads to much better data fit.

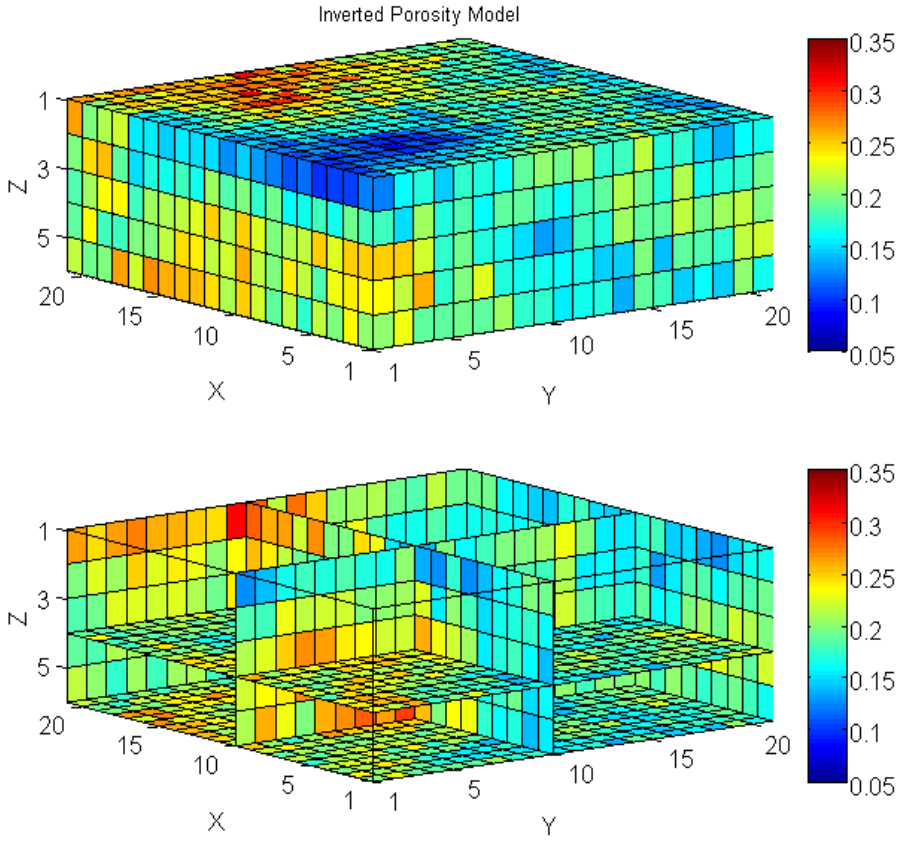

Figure 5.10: The most likely fine-scale reservoir model that is derived based on the multi-scale MCMC samples defined on the scale of 20x20x5. It is represented in a 3D cube and slices.

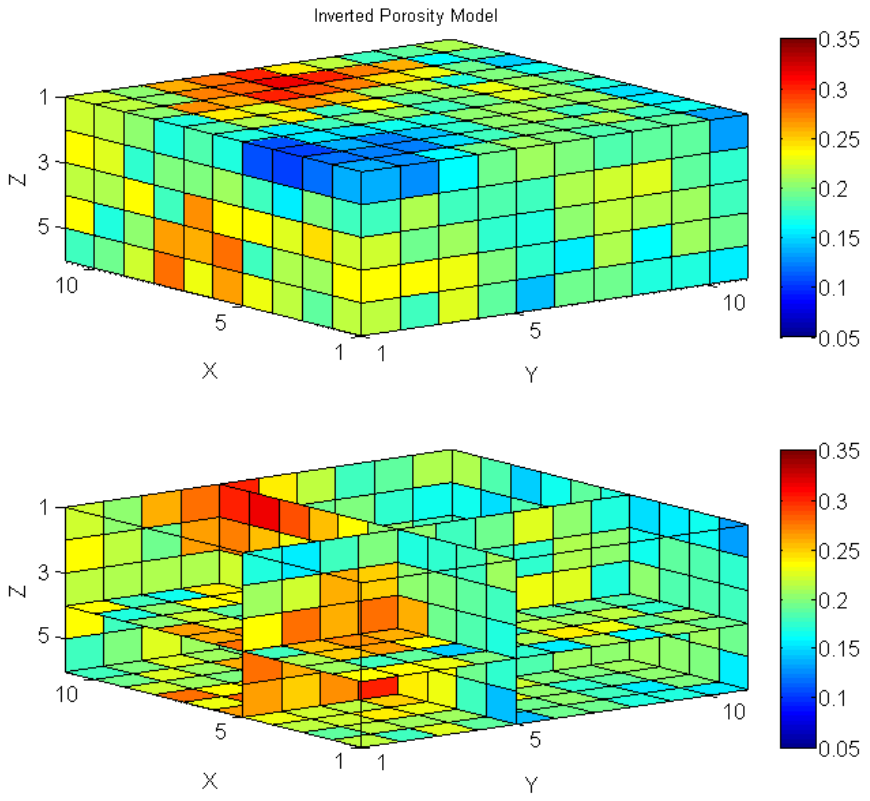

Figure 5.11: The most likely coarse-scale reservoir model that is derived based on the multi-scale MCMC samples defined on the scale of 10x10x5. It is represented in a 3D cube and slices. As expected, this coarse scale model does not provide reliable reservoir description.

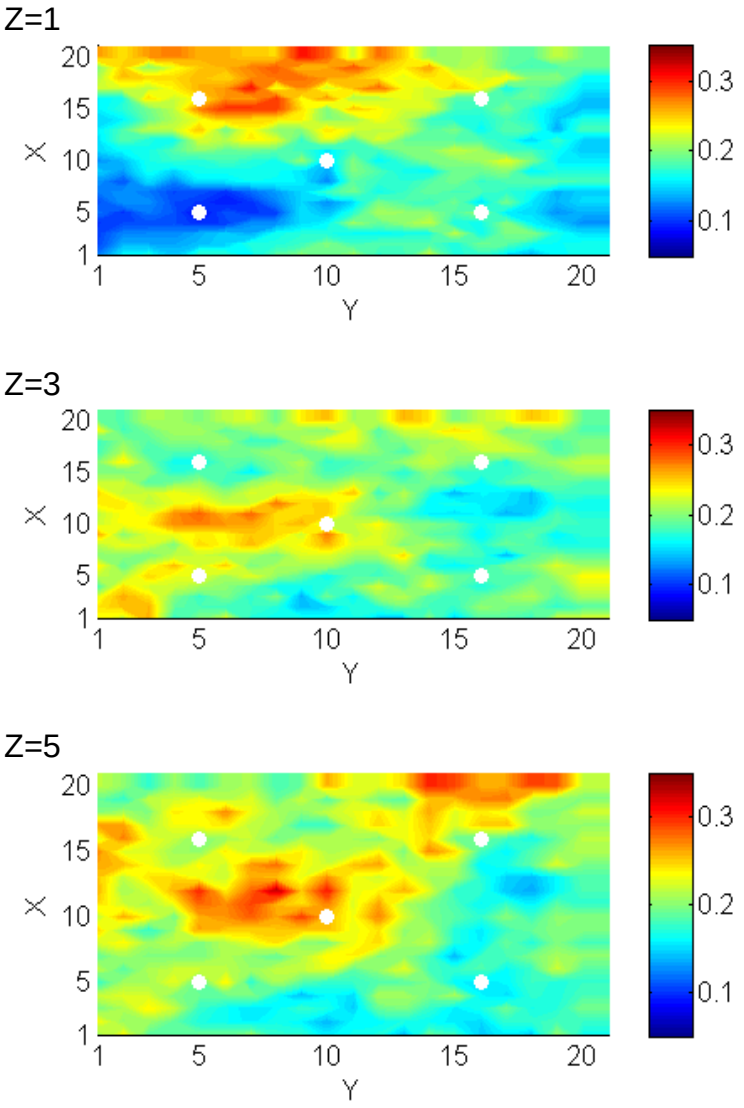

Figure 5.12: The derived porosity distributions using the multi-scale MCMC and for three different layers that are extracted from the derived fine-scale 3D reservoir porosity model in Figure 5.10. Although it is not as smooth as the reference distributions in Figure 5.4, the distributions of high and low porosity zones match to each other quite well.

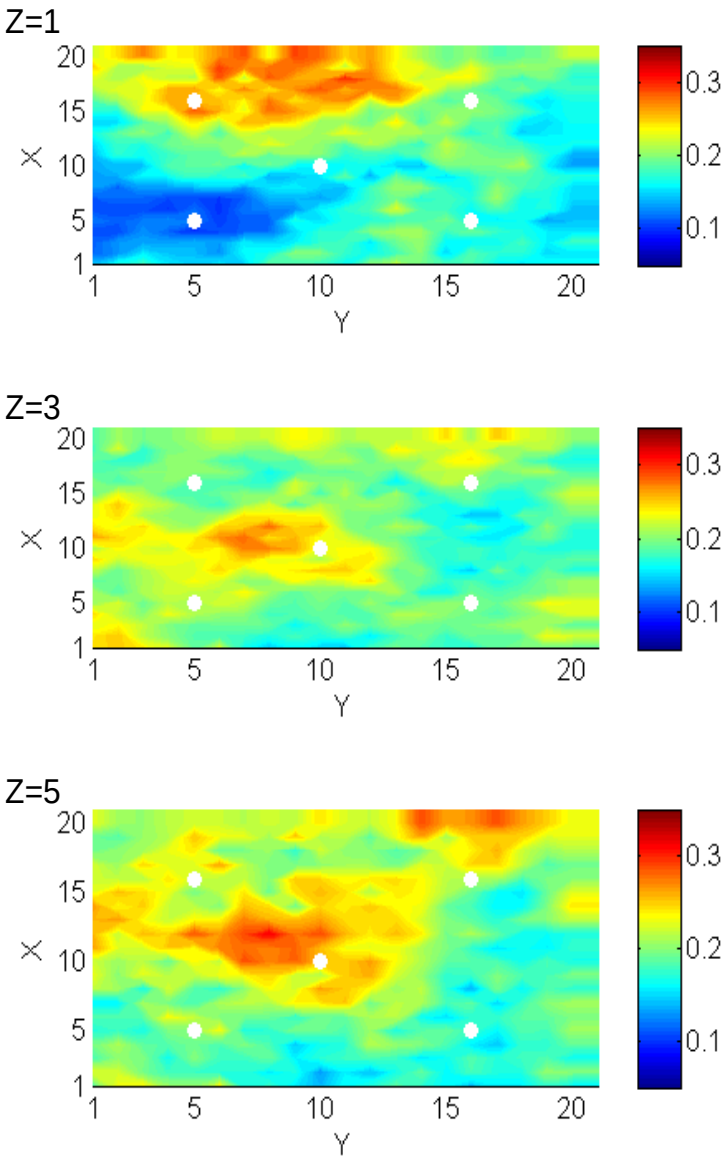

Figure 5.13: The derived porosity distributions using the single-scale MCMC and for three of the five layers in the fine-scale model. Although, compared to Figure 5.12, a similarly accurate porosity model is obtained, the single-scale sampling algorithm calls for much more computer resources, namely at least twice as much the total CPU time than the multi-scale sampling.

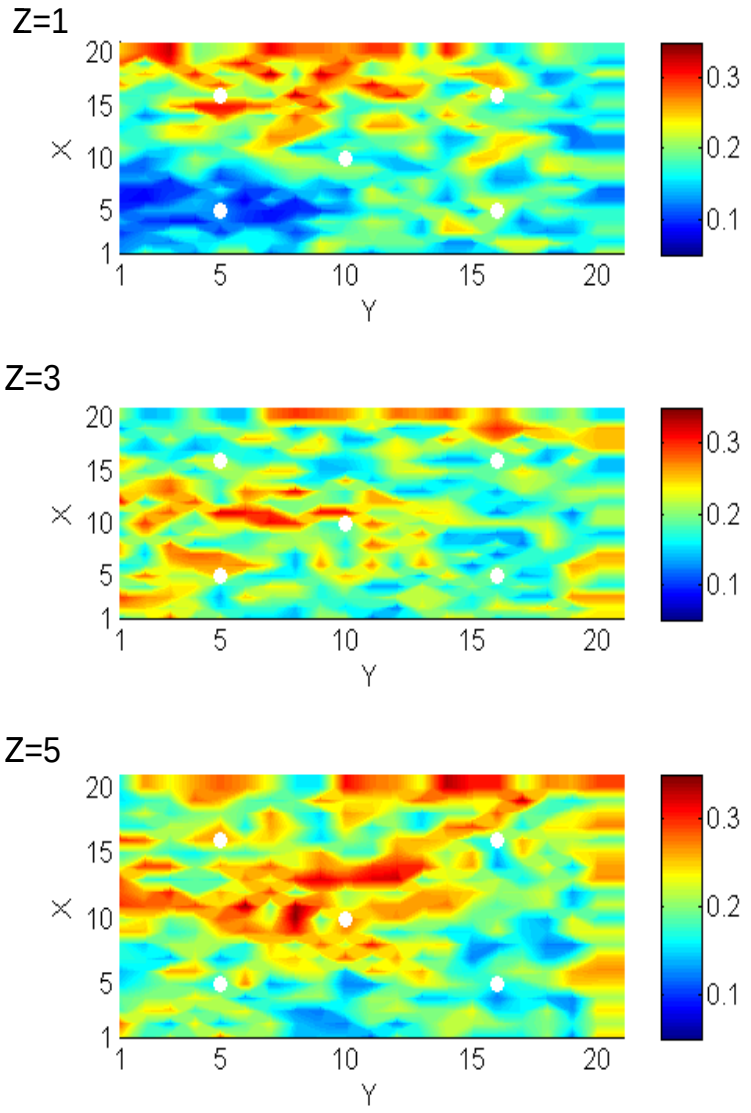

Figure 5.14: The derived porosity distributions using the multi-scale MCMC for three of the five layers but with well production data in use only. Due to the sparse spatial distribution, well information does not provide reliable constraints on reservoir's heterogeneity and hence does not result in accurate characterization.

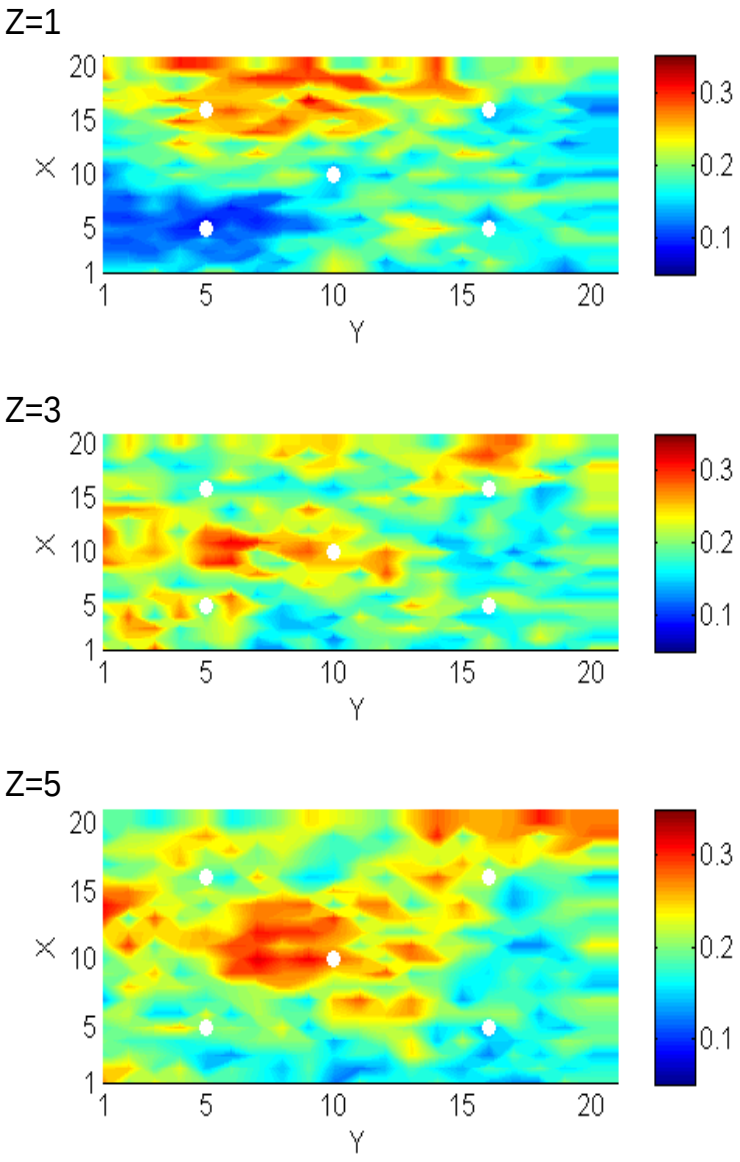

Figure 5.15: The derived porosity distributions using the multi-scale MCMC for three of the five layers but with 4D seismic data in use only. With benefits of extensive spatial coverage and dense lateral sampling in seismic data, this scheme leads to a better global characterization compared to Figure 5.14.

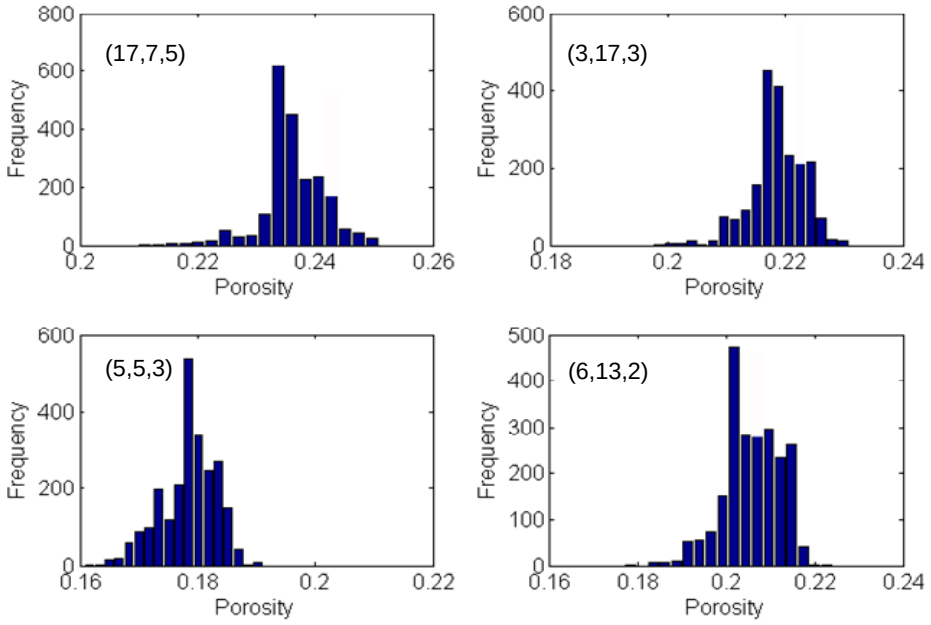

Figure 5.16: Calculated histograms of four particular grid blocks based on the multi-scale MCMC samples on the fine scale, which are used as the estimated marginal PPDs to determine the most likely porosity values and to quantify corresponding uncertainties.

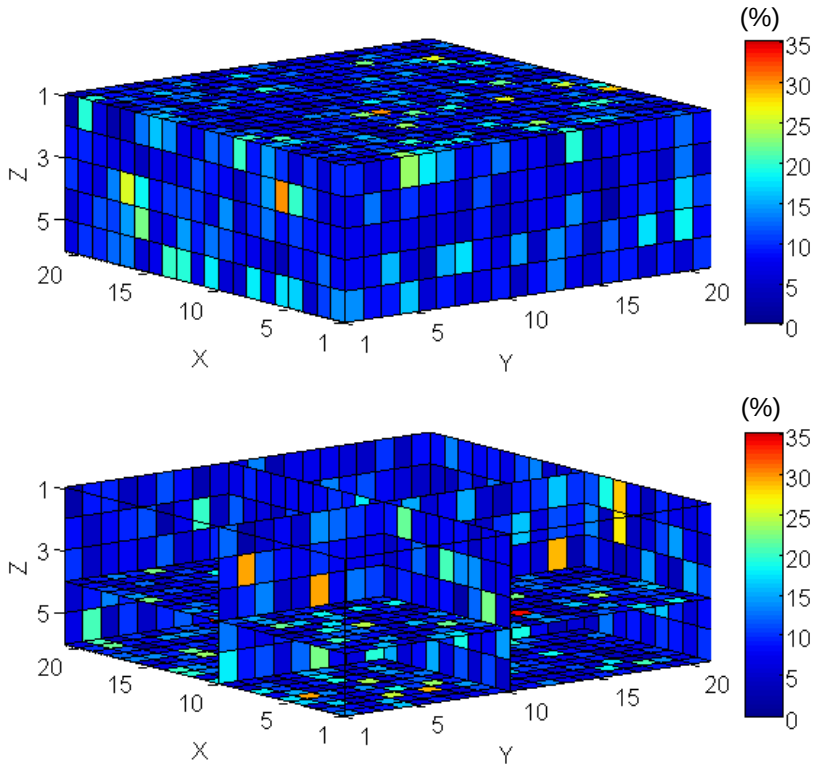

Figure 5.17: Derived relative porosity deviation model of scale 20x20x5 and with respect to the most likely porosities, which is represented in a 3D cube and slices as well.

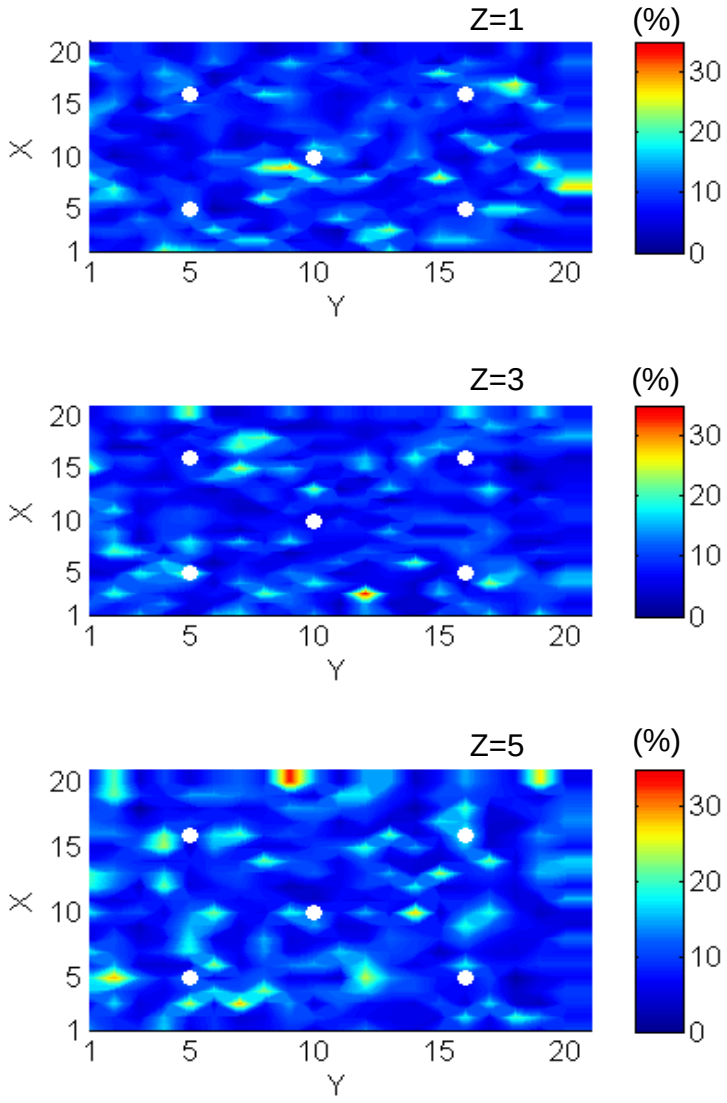

Figure 5.18: Derived relative porosity deviation for three different layers extracted from the 3D deviation cube in Figure 5.17. As shown, although uncertainties are relatively large at some points, they mostly localize below 15%.

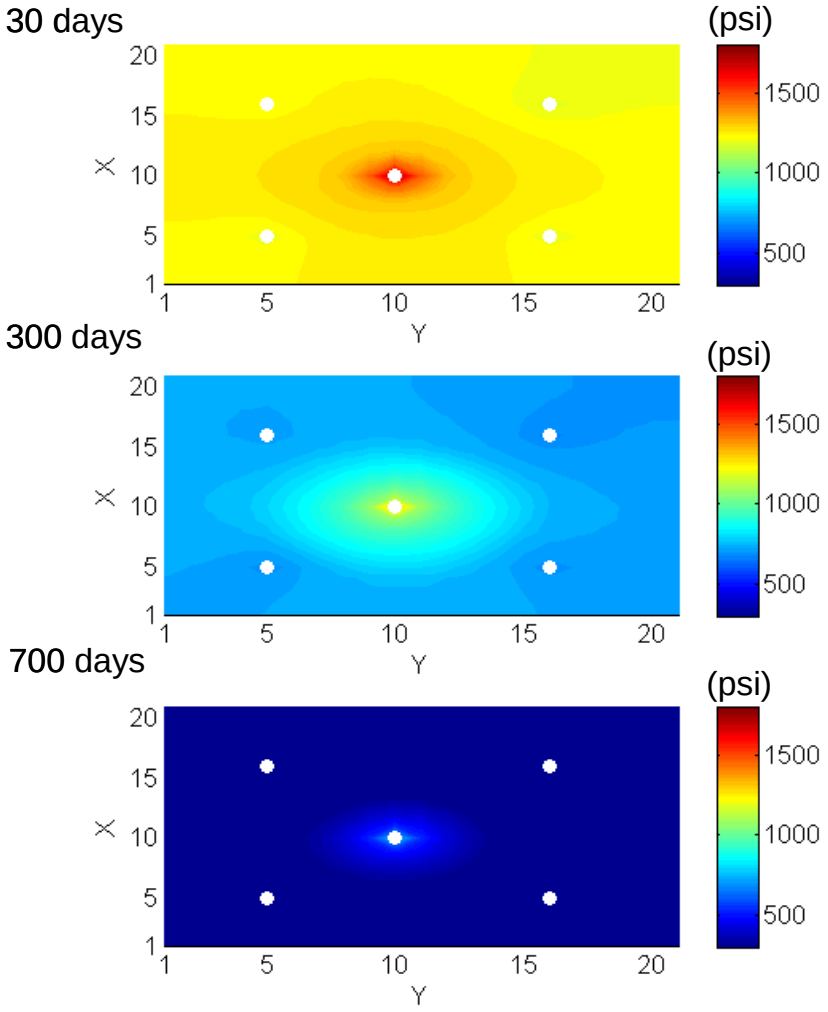

Figure 5.19: Distributions of the dynamic pore pressure for a particular layer (Z=3) after 30, 300 and 700 *days* respectively. As shown, the overall pore pressure drops through production time, and there are decreasing gradients between the injection well and production wells. In addition, pore pressure is especially low around production wells and is especially high around the injection well.

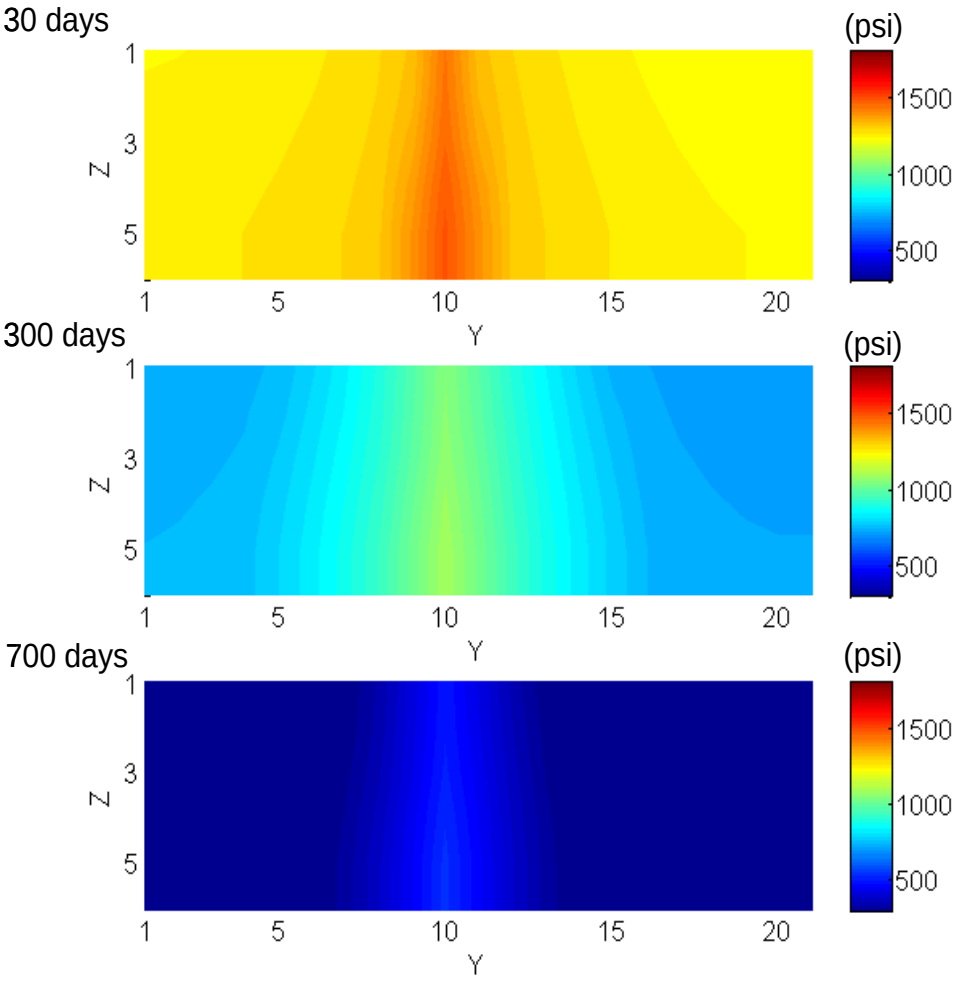

Figure 5.20: Distributions of the dynamic pore pressure for a particular cross-section (X=10) after 30, 300 and 700 *days* respectively. The same observations are obtained as those from Figure 5.19.

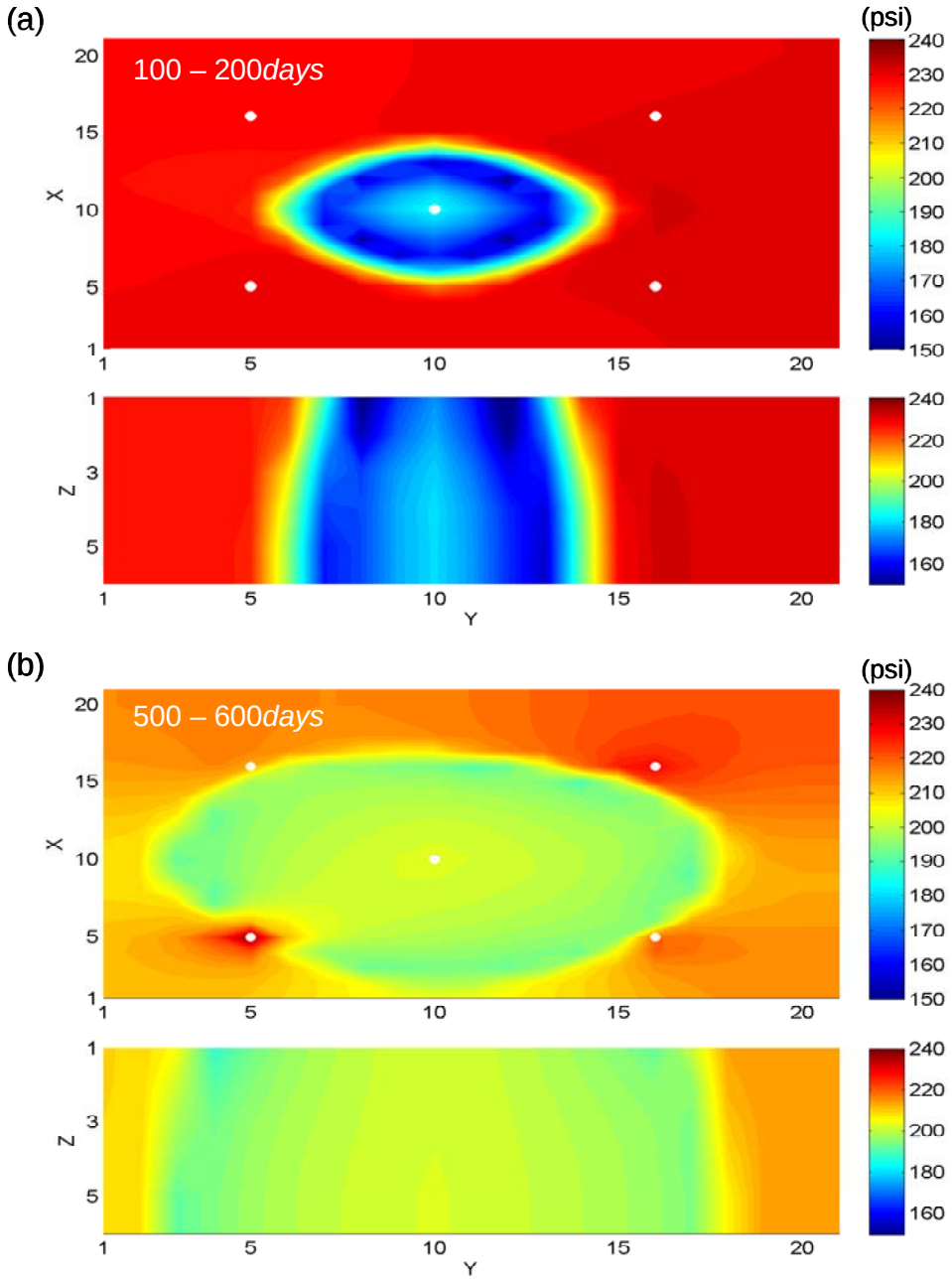

Figure 5.21: Differential decreases in pore pressure for a particular layer (*Z*=3) and a particular cross-section (*X*=10) respectively from (a) 100 to 200 *days*, and (b) 500 to 600 *days*.

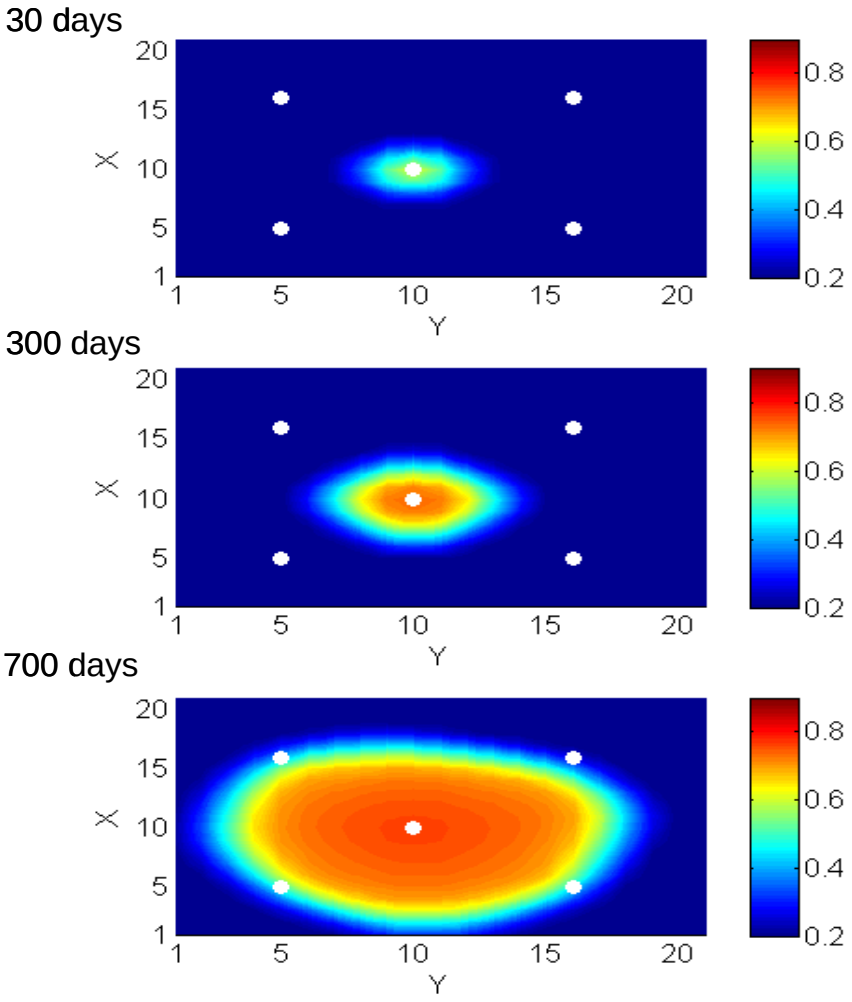

Figure 5.22: Distributions of the dynamic water saturation for a particular layer (Z=3) after 30, 300 and 700 *days* respectively. The injected water propagates laterally and the movement of water front is imaged.

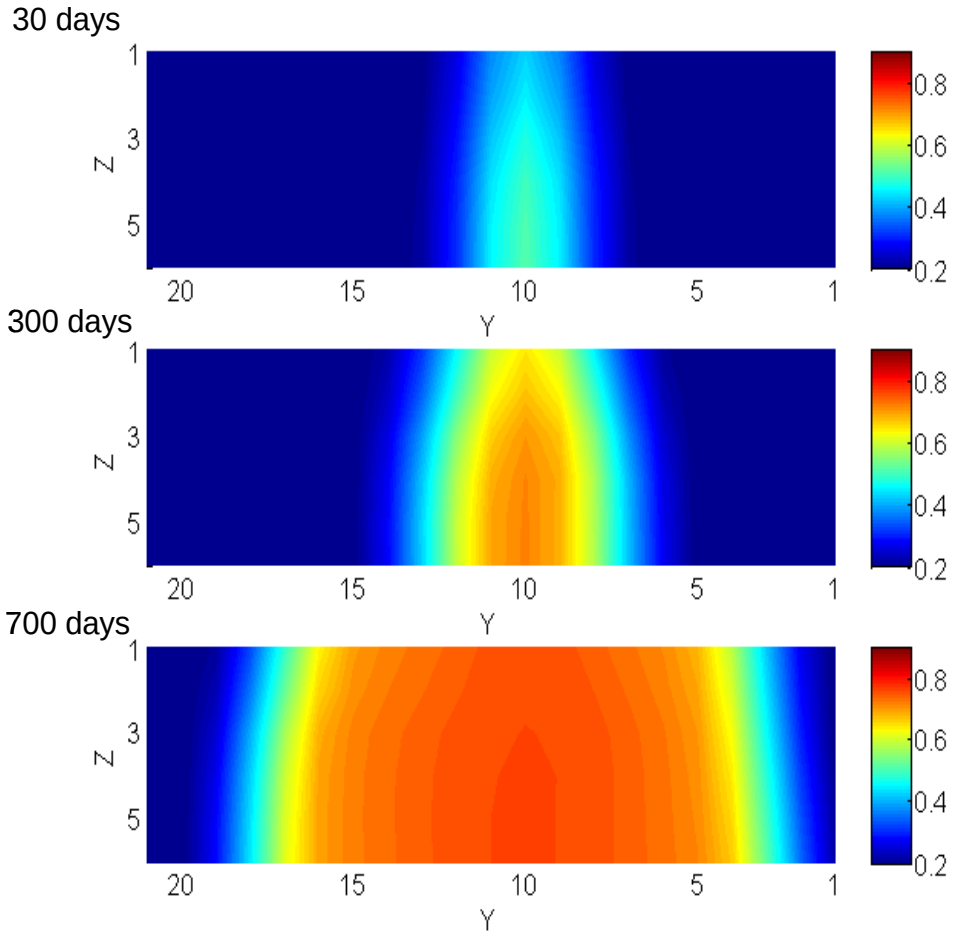

Figure 5.23: Distributions of the dynamic water saturation for a particular cross-section (X=10) after 30, 300 and 700 *days* respectively. The same phenomenon as that in Figure 5.21 is observed.

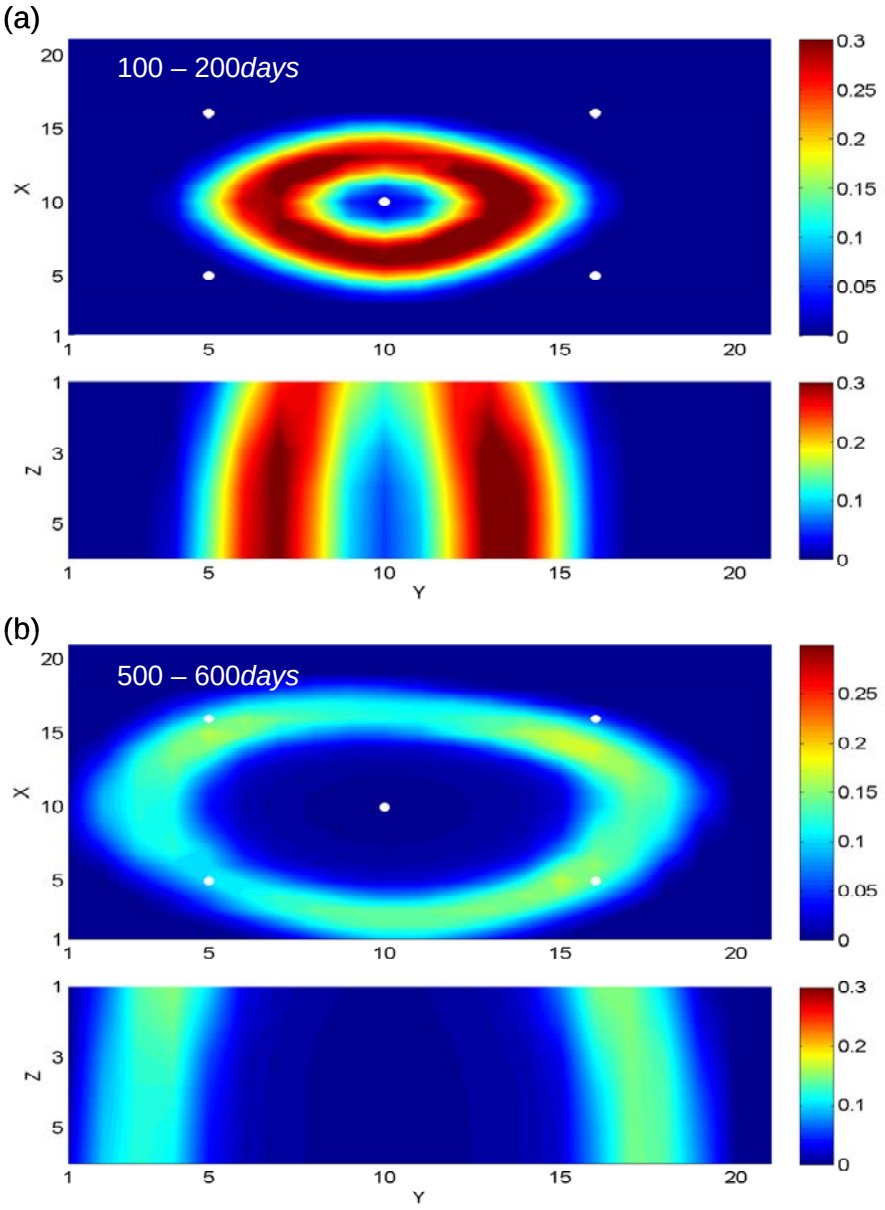

Figure 5.24: Differential increases in water saturation for a particular layer (Z=3) and a particular cross-section (X=10) respectively from (a) 100 to 200 *days*, and (b) 500 to 600 *days*.

Chapter 6

Conclusions and future directions

6.1 Conclusions

6.1.1 Markov Chain Monte Carlo (MCMC) algorithm

In this dissertation, some new developments are incorporated into a conventional Genetic Algorithm (GA) to design a new global optimization method and a new MCMC sampling algorithm. With a nonlinear pre-stack seismic waveform inversion problem, the applicability and usefulness of these new algorithms were demonstrated. In particular, a multi-scale hybrid GA was developed to rapidly exploit the optimal model. Although parallel GAs have been used to exploit the model space, application of our new multi-scale GA to seismic waveform inversion shows better performance compared to the results from a standard single-scale GA: both computational efficiency and estimation accuracy are improved. It was found that the multi-scaling helps overcome the model parameterization issue, as the inferred optimal models of different parameterizations can be investigated to determine the best parameterization without separately solving a model selection problem in a Bayesian setting.

Although PPD on the fine scale is our primary interest, incorporation of the coarser MCMC chains provides additional benefits by yielding more tractable PPDs, facilitating faster mixing and speeding up of the forward simulation. While inverting seismic data, over-parameterization results in non-uniqueness that can be addressed by algorithms that are computationally expensive, whereas, under-parameterization may be fast but can often result in an inadequate data fit. To take advantage of the benefits from all different scales, multiple MCMC chains of different scales are run simultaneously and realizations from different chains are uniquely combined to explore possibly unexplored portions. Although its applicability for correct PPD estimation and uncertainty analysis in practical problems is only demonstrated using a 1D seismic inversion example with as many as about 250 free parameters, it can be easily adapted to, and is equally applicable to, other complex 2D or 3D models. In addition, the parallelism inherent in multi-scaling and coupled MCMC makes our algorithms very adaptive to a parallel computing environment.

6.1.2 Integrated seismic history matching in a Bayesian framework

Integrated history matching of 4D seismic and well production data for reservoir characterization usually gives rise to an ill-posed joint inverse problem. This is primarily caused by the high-dimensional representation of the reservoir characteristics and the strong non-linear nature of the forward simulation processes. In addition, the need for dynamic updating due to incoming observations has to be accounted for. Hence, the

reservoir model estimation is typically not deterministic, and stochastic representation of the joint inverse problem is promising. In this dissertation, an integrated history matching workflow has been proposed, in which 4D seismic and well production data are integrated to condition reservoir characterization and corresponding uncertainty quantification. This integrated workflow is formulated in a Bayesian framework to take advantage of its unique capability in data integration and uncertainty analysis.

Assessing the uncertainty in reservoir characterization, i.e., deriving the entire posterior distribution, is a very difficult and very computer-intensive task. This calls for a powerful MCMC algorithm that can sufficiently explore the model space for PPD construction within a satisfactory CPU time period. In this study, the new multi-scale MCMC algorithm is applied for PPD sampling. Accordingly, a 3D reservoir model is parameterized into a fine scale and a coarse scale as well. Although the coarse scale does not provide reliable information about the target reservoir, incorporation of it helps to speed up the convergence of the fine-scale model to a good estimation. In addition, by periodically exchanging information between the two different scales, the coarse-scale model works as a smoothing operator to make the derived fine-scale model smooth and more realistic. Better performance by using multi-scaling is justified compared to using single-scaling.

Based on the numerical results estimated from MCMC samples, it is confirmed that integrating seismic and production data leads to a reasonably accurate reservoir porosity model compared to using either seismic or production data alone. This is because that the integration of these two datasets takes advantage of the benefits from

both of them, that is the greater local resolution from production data and global extensive coverage and dense sampling from seismic data. Besides the static porosity model, corresponding uncertainty is also derived based on the MCMC samples. Uncertainty in reservoir model description is one of the most common uncertainties that lead to risks in reservoir development and management. Therefore, the uncertainty quantification as described in this study will facilitate risk assessment associated with reservoir decision making. In this study, the derived static reservoir petrophysical model is further used to run the flow simulation to image the inside dynamic evolutions. Dynamic pore pressure and water saturation evolutions over production time have been obtained, and they provide reliable description of the inside fluid flow. This is also very important as they lead to well-founded production prediction and optimal recovery strategy.

6.2 Future directions

6.2.1 MCMC algorithm

As introduced, in the new multi-scale MCMC algorithm, multiple chains are run simultaneously, with each chain acting as a standard GA run. Therefore, one advantageous feature of this new MCMC algorithm is the parallelism; that is, some of the operations can be performed in parallel. In simplest form, a MCMC algorithm cannot be performed in parallel for updating because each parameter update has to be finished

before the next generation of the algorithm can start. In other words, the preceding step must be completed to determine the full conditional distribution for guidance on the next step. Nevertheless, in some cases of independence of model variables, parallel implementation is a much more efficient scheme, and the new multi-scale coupled MCMC algorithm can create this independence.

The increasing ability of computers to perform large numbers of calculations in a limited amount of time has allowed many improvements in computer intensive problems such as the integrated reservoir characterization problem in a Bayesian framework. As an extended development, the multi-scale MCMC algorithm can be further adapted to a parallel implementation for even better performance.

Finally, it is noted that the schemes of multi-scaling and coupled MCMC are equally adaptable to other global optimization methods such as Very Fast Simulated Annealing (VFSA) and other MCMC methods such as Parallel Tempering (PT). Therefore, applicability of multi-scale VFSA and multi-scale PT may be investigated in the future.

6.2.2 Integrated seismic history matching in a Bayesian framework

Porosity and permeability are two key petrophysical parameters. In integrated seismic reservoir characterization problems, modelers usually take either porosity or permeability as the primary model unknown and assume the other to be of a constant distribution through the simulation time. Although this helps to reduce the number of

free parameters and hence reduce the inherent non-uniqueness, it is intuitively not realistic. Pape *et al*. (1999) presented a relationship to map the porosity to permeability for different geological settings, which can be summarized as

$$k = a \cdot \phi + b \cdot \phi^2 + c \cdot (10 \cdot \phi)^{10} \quad [nm^2], \qquad (6.1)$$

where,

$$\begin{array}{lll} \textit{For average sandstone}: & a = 31, & b = 7463, \quad c = 191 \\ \textit{For shaly sandstone}: & a = 6.2, & b = 1493, \quad c = 58 \\ \textit{For shale}: & a = 0.1, & b = 26, \quad c = 1 \end{array} .$$

Thus, potential future work could incorporate this model in the seismic history matching workflow. For example, this model may be included in the numerical example presented in chapter 5 to derive permeability distribution based on the derived porosity distribution while assuming the permeability to be variable throughout the production time of the target reservoir.

As described in chapter 4, in the seismic history matching workflow, a rock physics model is needed to convert the reservoir flow parameters into seismic elastic parameters. Empirical relationships are typically used for this purpose, in which the coefficients are usually pre-determined based on core measurements or laboratory experiments. These pre-specified empirical coefficients may give rise to another set of uncertainties in reservoir characterization as the reservoir's geological properties vary from field to field. A better way to handle this is to take these coefficients as another set of model unknowns in addition to the primary reservoir variables and to integrate them into the Bayesian inference framework as

$$p(\mathbf{m},\mathbf{a}|\mathbf{S}^{obs},\mathbf{P}^{obs}) = \frac{p(\mathbf{S}^{obs}|\mathbf{m},\mathbf{a})\cdot p(\mathbf{P}^{obs}|\mathbf{m})\cdot p(\mathbf{m})\cdot p(\mathbf{a})}{p(\mathbf{S}^{obs})\cdot p(\mathbf{P}^{obs})} \quad , \qquad (6.2)$$

where **a** represents the coefficients of a rock physics model that is used to link flow parameters to seismic elastics parameters. In this way, the best-fitting coefficients **a** can be derived stochastically while the best-fitting reservoir model is derived. Therefore, **a** is now a data-driven unknown and does not have to be pre-determined based on laboratory experiments or empirical formulae. The same idea applies to the coefficients of *a*, *b* and *c* in equation (6.1). Moreover, both the coefficients of a rock physics model and the coefficients of the porosity-permeability relationship can also be incorporated in the Bayesian inference framework at the same time.

Integrated seismic history matching for reservoir characterization in the Bayesian framework is a computer-intensive problem. One of the main causes comes from the forward flow simulation over a multivariate reservoir model. Thus, an efficient reservoir characterization always calls for an efficient flow simulator, and it is a good trial to make use of a more efficient simulator for potentially better performance.

Appendix A: Metropolis-Hastings algorithm

Suppose the goal is to draw samples from a certain distribution $p(m)$ and $p(m)=f(m)/K$, where the normalization constant K may not be known and may be very difficult to compute. The **Metropolis algorithm** (Metropolis and Ulam, 1949; Metropolis *et al*., 1953) generates a sequence of draws from this distribution as follows:

1. Start with any initial values m_0 satisfying $f(m_0)>0$.
2. Using current m value, sample a candidate point m^* from some jumping distribution $q(m_1, m_2)$, which is the probability of returning a value of m_2 given a previous value of m_1. This distribution is also referred to as the "proposal" or "candidate-generating distribution". The only restriction on the jump density in the Metropolis algorithm is that it is symmetric, i.e., $q(m_1, m_2)= q(m_2, m_1)$.
3. Given the candidate point m^*, calculate the ratio of the density at the candidate (m^*) and the current (m_{t-1}) point,
$$\alpha = \frac{p(m^*)}{p(m_{t-1})} = \frac{f(m^*)}{f(m_{t-1})} \quad .$$
Notice that because we are considering the ratio of $p(m)$ under two different values, the normalizing constant K cancels out.
4. If the jump increases density ($\alpha>1$), accept the candidate point (set $m_t= m^*$) and return to step 2. If the jump decreases the density ($\alpha<1$), then with probability α accept the candidate point, or else reject it and return to step 2.

We can summarize the Metropolis sampling as first computing

$$\alpha = \min\left(\frac{f(m^*)}{f(m_{t-1})}, 1\right) \quad ,$$

and then accepting a candidate point with probability α (the probability of a move). This generates a Markov chain (m_0, m_1, …, m_k, …), as the transition probabilities from m_t to m_{t+1} depends only on m_t and not on (m_0, m_1, …, m_{t-1}). Following a sufficient burn-in period (say, k steps), the chain approaches its stationary distribution, and samples from the vector (m_k, m_{k+1}, …, m_{k+n}) are samples from $p(m)$ (see Gregory (2005) for details).

Hastings (1970) generalized the Metropolis algorithm by using an arbitrary transition probability function $q(m_1, m_2)$=Pr($m_1 \rightarrow m_2$), and setting the acceptance probability for a candidate point as

$$\alpha = \min\left(\frac{f(m^*)q(m^*, m_{t-1})}{f(m_{t-1})q(m_{t-1}, m^*)}, 1\right)$$

This is the **Metropolis-Hastings algorithm**. Assuming that the proposal distribution is symmetric, i.e., $q(m^*, m_{t-1}) = q(m_{t-1}, m^*)$, it recovers the original Metropolis algorithm.

Bibliography

Antonisse, J. (1989), "A new interpretation of schema notation that overturns the binary encoding constraint", *Proc. of the Third Int. Conf. on Genetic Algorithms*, J. David Schaffer (Ed.) (Morgan Kaufmann Publishers, San Mateo), 86-91.

Arenas, E., C. van Kruijsdijk, and T. Oldenziel (2001), "Semi-automatic history matching using the pilot point method including time-lapse seismic data", SPE71634, in 2001 *SPE Annual Technical Conference and Exhibition*.

Back, T. (1993), "Optimal mutation rates in genetic search", In *Proceeding of the Fifth International Conference on Genetic Algorithms*. Morgan Kaufmann: San Mateo, CA.

Backus, G. E. (1962), "Long-wave elastic anisotropy produced by horizontal layering", *J. Geophys. Res.*, **67**(11), 4427-4440.

Baker, J. E., (1987), Reducing bias and inefficiency in the selection algorithm, in J. J. Grefenstette (Ed.), Proceedings of the Second International Conference on Genetic Algorithms, 14-21. Brooks, S. P. (1998) "Markov chain Monte Carlo method and its application." *The Statistician,* **47**(1): 69-100.

Batzle, M., and Z. Wang, (1992), Seismic properties of pore fluids, *Geophysics* **57**(11): 1396-1408.

Behrens, R., P. Condon, W. Haworth, M. Bergeron, Z. Wang, and C. Ecker (2002), "4D seismic monitoring of water flux at Bay Marchand: The practical use of 4D in an imperfect world", *SPE Reservoir Evaluation and Engineering*, **5**(5), 410-420.

Biot, M. A., (1956a), Theory of propagation of elastic waves in a fluid saturated porous solid, I. Low frequency range, *J. Acoust. Soc. Am.*, **28**, 168-178.

Biot, M. A., (1956b), Theory of propagation of elastic waves in a fluid saturated porous solid, II. Higher frequency range, *J. Acoust. Soc. Am.*, **28**, 179-191.

Bouman, C., and B. Liu, (1991). "Multiple resolution segmentation of textured images." *IEEE Transactions on Pattern Analysis and Machine Learning*, **13**(2): 99-113.

Box, G. P., and Tiao, G. C., (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley Publishing Company.

Brooks, S. P. (1998). "Markov chain Monte Carlo method and its application." *The Statistician*, **47**(1): 69-100.

Buland, A., and Y. E. Ouair, (2006), “Bayesian time-lapse inversion”, *Geophysics* **71**(3): R43-R48.

Caers, J. (2005), *Petroleum Geostatistics*, Society of Petroleum Engineers, 96p.

Caers, J., T. Hoffman, S. Strebelle, and X. Wen, (2006) "Probabilistic integration of geological scenario, seismic, and production data- a West Africa turbidite reservoir case study." *The Leading Edge*, 240-244, March.

Carlin, B., and S. Chib, (1993). "Bayesian model choice via Markov Chain Monte Carlo." *Journal of the Royal Statistical Society. Series B,* **57**: 473-484.

Cary, P. W., and Chapman, C. H., 1988, “Automatic 1D waveform inversion of marine seismic reflection data”, *Geophys. J. Int*., **93**, 527-546.

Castro, S. A., (2007), A probabilistic approach to jointly integrate 3D/4D seismic production data and geological information for building reservoir models, *Ph.D Dissertation*, Stanford University.

Chapin, M., P. Swinburn, R. V. Weiden, D. Skaloud, S. Adesanya, D. Stevens, C. Varley, J. Wilkie, E. Brentjens and M. Blaauw, (2002), Integrated seismic and subsurface characterization of Bonga Field, offshore Nigeria, *The Leading Edge*: 1125-1131.

Chatterjee, S., M. Laudato, and L. A. Lynch, (1996), “Genetic algorithms and their statistical applications: an introduction”, *Computational Statistics and Data Analysis*, **22**: 633-651.

Costa, A.P.A., D. J. Schiozer, (2003), Treatment of Geological Attributes in Risk Analysis applied to the Appraisal Phase of Petroleum Fields, Canadian International Petroleum Conference, Calgary, Canada.

Curtis, A., and A. Lomax, (2001). "Prior information, sampling distributions, and the curse of dimensionality." *Geophysics*, **66**(2): 372-378.

Davis, T. E., (1991), Towards an extrapolation of the Simulated Annealing convergence theory onto the Genetic Algorithm, *Ph.D. Dissertation*, University of Florida, Gainesville.

Davis, T. E., and J. C. Principe, (1993), A Markov chain framework for the simple Genetic Algorithm, *Evolutionary Comp*., **1**(3): 269-288.

Debeye, H., and P. M. van der Made (1996), “Stochastic inversion”, *SEG 66th Annual International Meeting*, Denver, Colorado.

DeJong, K. (1975), An analysis of the behavior of a class of genetic adaptive systems, *Ph.D. Dissertation*, University of Michigan, Ann Arbor, MI.

Dong, Y., (2005), Integration of time-lapse seismic data into automatic history matching, *Ph.D. Dissertation*, The University of Oklahoma, Norman, OK.

Dong, Y., and D. S. Oliver (2005), “Quantitative use of 4D seismic data for reservoir characterization”, *SPE Journal*, **10**(1), 91-99.

Dunlop, T. N. B., G. A. King, and E. A. Breltenbach (1991), "Monitoring oil/water fronts by direct measurement" (includes associated papers 23943 and 24103), *Journal of Petroleum Technology*, pp. 596-602.

Earl, D. J., and M. W. Deem, (2005), "Parallel tempering: Theory, application, and new perspectives." *Phys. Chem. Chem. Phys.*, **7**: 3910-3916.

Ertekin, T., J. H. Abou-Kassem, and G. R. King, (2001), *Basic Applied Reservoir Simulation*, Society of Petroleum Engineers Inc.

Floris, F. J. T., M. D. Bush, M. Cuypers, F. Roggero, and A. R. Syversveen, (1999), "Comparison of production forecast uncertainty quantification methods- an integrated study", in *Proc. 1st Symp. Petroleum Geostatist.*, Toulouse, France, Apr. 20-23, 1999.

Gassmann, F., (1951), Elastic waves through a packing of spheres, *Geophysics* **16,** 673–85.

Geertsma, J., and Smit, D. C., (1961), Some aspects of elastic wave propagation in fluid saturated porous solids, *Geophysics*, **26**, 169-181.

Gelfand, A. E., and A. F. M. Smith, (1990), "Sampling based approaches to calculating marginal densities", *Journal of American Statistical Association*, **85**, 398-409.

Gelfand, A. E., A. F. M. Smith, and T. M. Lee, (1992), "Bayesian analysis of constrained parameter and truncated data problems using Gibbs' sampling", *Journal of American Statistical Association*, **87**, 523-532.

Geman, S., and D. Geman, (1984), "Stochastic relaxation, Gibbs' distribution and Bayesian restoration of images", *IEEE Trans.*, PAMI-6, 721-741.

Geyer, C. J., (1991), "Markov Chain Monte Carlo maximum likelihood", in *Proc. Soc. Petroleum Eng. Annu. Tech. Conf.*, 2000.

Goldberg, D. E., (1989), *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley: Reading, Massachusetts.

Goldberg, D. E., B. Korb, and K. Deb, (1989). "Messy Genetic Algorithms: Motivation, analysis, and first results." *Complex Systems*, **3**: 493-530.

Goldberg, D. E. (1991), "Real-coded Genetic Algorithms", Virtual Alphabets, and Blocking. *Complex Systems*, **5**, 139-167.

Gomez-Hernandez, J., and A. Journel (1993), "Joint sequential simulation of multiGaussian fields", in Soares, A. (ed.) *Geostatistics Troia '92*, Kluwer Academic Publishers, 85-94.

Gosselin, O., S. I. Aanonsen, I. Aavatsmark, A. Cominelli, R. Gonard, M. Kolasinski, F. Ferdinandi, L. Kovacic, and K. Neylon (2003), "History matching using time-lapse seismic (HUTS)", SPE84464, in 2003 *SPE Annual Technical Conference and Exhibition*.

Gouveia, W. P., D. H. Johnston, A. Solberg, and M. Lauritzen (2004), "Jotun 4D: Characterization of fluid contact movement from time-lapse seismic and production logging tool", *The Leading Edge*, 1187-1194.

Greaves, R. J., and T. J. Fulp (1987), "Three dimensional seismic monitoring of an enhanced oil recovery process", *Geophysics*, **52**(9), 1175-1187.

Gregory, P., (2005), *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematical Support*, Cambridge University Press.

Haario, H., E. Saksman, and J. Tamminen, (2005), "Componentwise adaptation for high dimensional MCMC", *Computational Statistics*, **20**:265-273.

Hastings, W. K., (1970), "Monte Carlo methods using Markov chains and their applications", *Biometrika*, **57**, 97-109.

Herrera, F., and M. Lozano (2000). "Gradual Distributed Real-Coded Genetic Algorithms." *IEEE Transactions on Evolutionary Computation,* **4**(1).

Herrera, F., M. Lozano, and J. L. Verdegay (1995), "Tuning Fuzzy Logic Controllers by Genetic Algorithms", *International Journal of Approximate Reasoning*, **12**, 299-315.

Herrera, F., M. Lozano, and J. L. Verdegay (1998), "Tacking Real-coded Genetic Algorithms: Operators and Tools for Behavioural Analysis", *Artificial Intelligence Review*, **12**, 265-319.

Higdon, D., H. Lee, and Z. Bi, (2002), "A Bayesian approach to characterizing uncertainty in inverse problems using coarse and fine-scale information", *IEEE transaction on signal processing*, **50**(2): 389-399.

Holland, J. H. (1975). "Adaptation in natural and artificial systems." Ann Arbor: The University of Michigan Press.

Holloman, C. H., (2002), Parameter estimation algorithms for computational intensive spatial problems, *Ph.D. Dissertation*, Duke University.

Holmes, C. C., and B. K. Mallick, (1998). "Parallel Markov Chain Monte Carlo sampling." Technical Report, Imperial College.

Hong, T., and M. K. Sen (2006), Real-coded multi-scale genetic algorithm for pre-stack seismic waveform inversion, Expanded Abstract, 76th SEG Conference, New Orleans.

Hong, T., and M. K. Sen (2007), A new MCMC algorithm for parameter estimation and uncertainty analysis, Expanded Abstract, 77th SEG Conference, San Antonio, 1883-1887.

Hong, T., M. K. Sen, P. L. Stoffa, H. Klie, S. G. Thomas, A. Rodriguez, and M. F. Wheeler (2007), Integrated time-lapse seismic inversion for reservoir's petrophysics and fluid flow imaging, Expanded Abstract, 77th SEG Conference, San Antonio, 1952-1956.

Hong, T., and M. K. Sen (2008a), A new MCMC algorithm for seismic parameter estimation and uncertainty analysis, *Geophysical Journal International*, in press.

Hong, T., and M. K. Sen (2008b), Joint Bayesian inversion for reservoir characterization and uncertainty quantification, Expanded Abstract, 78th SEG Conference, Las Vegas.

Huang, X., L. R. Bentley, and C. Laflamme (2001), "Integration of production history and time-lapse seismic data guided by seismic attribute zonation", SPE68819, in 2001 *SPE Annual Technical Conference and Exhibition*.

Huang, X., L. Meister, and R. Workman (1997), "Reservoir characterization by integration of time-lapse seismic and production data", SPE38695, in 1997 *SPE Annual Technical Conference and Exhibition*.

Janikow, C. Z., and Z. Michalewicz (1991), Experimental comparison of binary and floating point representations in Genetic Algorithms, San Diego, CA, USA: Morgan-Kaufmann Publication Inc.

Journel, A. G. (2002), "Combining knowledge from diverse sources: An alternative to traditional data independence hypothesis", *Mathematical Geology*, **34**(5), 573-596.

Kass, R., and L. Wasserman (1996), "The selection of prior distributions by formal rules." *Journal of American Statistical Association*, **91**: 1342-1370.

Kennett, B. L. N. (1983), "*Seismic wave propagation in stratified media*", Cambridge University Press.

Kirkpatrick, S., Jr. C. D. Gelatt, and M. P. Vecchi (1983), "Optimization by simulated annealing", *Science*, **220**, 671-680.

Klie, H., A. Rodriguez, S. G. Thomas, M. F. Wheeler, and R. Banchs (2006), "Assessing the value of sensor information in 4D seismic history matching" *SEG Expanded Abstracts*.

Kretz, V., M. L. Ravalec-Dupin, and F. Roggero (2004), "An integrated reservoir characterization study matching production data and 4D seismic", *SPE Reservoir Evaluation and Engineering*, **7**(2), 116-122.

Landa, J. L., and R. N. Horne (1997), "A procedure to integrate all well test data, reservoir performance history and 4D seismic information into a reservoir description", SPE38653, in 1997 *SPE Annual Technical Conference and Exhibition*.

Landro, M. (2001) Discrimination between pressure and fluid saturation changes from time-lapse seismic data, *Geophysics* **66**(3): 836-844.

Landro, M., and L. K. Stronen (2003), "4D study of fluid effects on seismic data in the Gullfaks Field, North Sea", *Geofluids*, 3, 233-244.

Li, Y., and D. W. Oldenburg (2000), “3D inversion of induced polarization data”, *Geophysics*, **65**(6), 1931-1945.

Liang, F., and W. H. Wong (2000). "Evolutionary Monte Carlo: Applications to Cp model sampling and change point problem." *Statistica Sinica*, **10**: 317-342.

Liang, F., and W. H. Wong (2001). "Real-parameter evolution Monte Carlo with applications to Bayesian mixture models." *Journal of the American Statistical Association*, **96**(454): 653-666.

Lumley, D. E. (2001), “Time-lapse seismic reservoir monitoring”, *Geophysics*, **66**(1), 50-53.

Lumley, D. E., D. Adams, M. Meadows, S. Cole, and R. Ergas (2003), 4D seismic pressure-saturation inversion at Gullfaks field, Norway, *First Break* **21**: 3-9.

Lumley, D. E., R. A. Behrens, and Z. Wang (1997), “Assessing the technical risk of a 4D seismic project”, *The Leading Edge*, **16**(9), 1287-1292.

Lumley, D. E., A. G. Numms, G. Delorme, and M. F. Bee (1999), “A 4D seismic case study”, *1999 SEG Expanded Abstracts*.

Lumley, D. E., and A. Viejo (2004), “Business and technology challenges for 4D seismic reservoir monitoring”, *The Leading Edge*, **23**(11), 1166-1168.

Malinverno, A. (2000). "A Bayesian criterion for simplicity in inverse problem parameterization." *Geophys. J. Int*., **140**: 267-285.

Malinverno A. (2002). "Parsimonious Bayesian Markov Chain Monte Carlo inversion in a nonlinear geophysical problem." *Geophys. J. Int.*, **151**: 675-688.

Malinverno, A., and W. S. Leaney (2005), "Monte-Carlo Bayesian look-ahead inversion of walkaway vertical seismic profiles", *Geophysical Prospecting*, **53**: 689-703.

Marinari, E., and G. Parisi (1992). "Simulated tempering: A new Monte Carlo scheme." *Europhysics Letters,* **19**: 451.

Maryak, J. L., and D. C. Chin (2001), “Global Random Optimization by Simultaneous Perturbation Stochastic Approximation”, *Proceedings of the American Control Conference*, Arlington, VA.

Mavko, G., T. Mukerji, and J. Dvorkin, (1998), *The rock physics handbook: Tools for seismic analysis in porous media*, Cambridge University Press.

McInally, A. T., T. Rendondo-Lopez, J. Garnham, J. Kunka, A. D. Brooks, L. Stenstrup Hansen, F. Barclay, and D. J. Davies (2003), “Optimizing 4D fluid imaging”, *Petroleum Geoscience*, **9**, 91-101.

Meadows, M. A. (2001), “Enhancements to Landro’s method for separating time-lapse pressure and saturation changes”, *2001 SEG Expanded Abstracts*.

Menke, W. (1984), *Geophysical data analysis: Discrete inverse theory*, Academic Press, Inc., London.

Metropolis, N., and S. Ulam (1949), "The Monte Carlo method", *J. Acous. Soc. Am*., **44**, 335-341.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. Teller, and H. Teller (1953), "Equations of state calculations by fast computing machines", *Journal of Chemical Physics*, **21**: 1087-1091.

Meyn, S. P., and R. L. Tweedie (1994), "Computable bounds for convergence rates of Markov chains", *Ann. Appl. Probab*., **4**, 981-1011.

Mezghani, M., A. Fornel, V. Langlais, and N. Lucet (2004), "History matching and quantitative use of 4D seismic data for an improved reservoir characterization", SPE90420, in 2004 *SPE Annual Technical Conference and Exhibition.*

Minkoff, A. E., C. M. Stone, S. Bryant, M. Peszynska, and M. F. Wheeler (2003), "Coupled fluid flow and geomechanical deformation modeling", *Journal of Petroleum Science and Engineering*, **38**, 37-56.

Mosegaard, K., and A. Tarantola (1995). "Monte Carlo sampling of solutions to inverse problems." *J. Geophys. Res.*, **100**(B7): 12,431-12,447.

Muhlenbein, H., and D. Schlierkamp-Voosen (1993), "Predictive Models for the Breeder Genetic Algorithm I: Continuous Parameter Optimization", *Evolutionary Computation,* **1**, 25-49.

Nur, A., and G. Simmons (1969), "The effect of saturation on velocity in low porosity rocks", Earth and Planetary Sciences Letters, **7**, 183-193.

Omre, H., K. Solna, and H. Tjelmeland (1993), "Simulation of random functions on large lattices", in Soares, A. (ed.) *Geostatistics Troia '92*, Kluwer Academic Publishers, 179-199.

Ono, I, and S. Kobayashi (1997), "A real-coded Genetic Algorithm for function optimization using unimodal normal distribution crossover", *Proceedings of the Seventh Int. Conf. on Genetic Algorithms.* T. Back (Ed.) (Morgan Kaufmann Publishers, San Francisco), 246-253.

Pape, H., C. Clauser, and J. Iffland (1999), "Permeability prediction based on fractal pore-space geometry", *Geophysics*, **64**(5) 1447-1460.

Peaceman, D. W. (1983), "Interpretation of well-block pressures in numerical reservoir simulation with non-square grid blocks and anisotropic permeability", *SPE Journal*, **23**(6), 531-543.

Pendrel, J. V., and van Riel, P. (1997), "Estimating porosity from 3D seismic inversion and 3D geostatistics", *SEG 67th Annual International Meeting*, Dallas, Texas.

Pickering, S., J. Khazanehdari, I. Marin, P. Hodgson, and S. Leaney (2005), "The integration imperative: piecing together the reservoir jigsaw", *First Break* **23**: 69-73.

Radcliffe, N. J. (1992), "Non-linear Genetic Representations", *Parallel Problem Solving from Nature 2*, R. Manner and B. Manderick (Eds.) (Elsevier Science Publishers, Amsterdam), 259-268.

Rickett, J., and D. E. Lumley (2001), "Cross-equalization data processing for time-lapse seismic reservoir monitoring: A case study from Gulf of Mexico", *Geophysics*, **66**(4), 1015-1025.

Ripley, B. D. (1981), *Spatial statistics*, Wiley, New York.

Rosenthal, J. S. (1994), "Minorization conditions and convergence rates for Markov Chain Monte Carlo, Technical Report 9321, Department of Statistics, University of Toronto.

Ross, C. P., G. B. Cunningham, and D. P. Weber (1996), "Inside the cross-equalization black box", *The Leading Edge*, **15**(11), 1233-1240.

Sambridge, M., (1999), "Geophysical inversion with a neighbourhood algorithm-I. Searching a parameter space", *Geophys. J. Int*., **138**, 479-494.

Sambridge, M., K. G., A. Jackson, and P. Rickwood, (2006), "Trans-dimensional inverse problems, model comparison and the evidence", *Geophys. J. Int.* doi: 10.1111/j.1365-246X.2006.03155.x: 1-15.

Schiozer, D. J., E. L. Ligero, and J. A. M. Santos, (2004), "Risk assessment for reservoir development under uncertainty", *J. Braz. Soc. Mech. Sci. & Eng.*, **26**(2), 213-217.

Sen, M. K., and I. G. Roy (2003), "Computation of differential seismograms and iteration adaptive regularization in pre-stack seismic waveform inversion", *Geophysics*, **68** (6), 2026-2039.

Sen, M. K., and P. L. Stoffa, (1991), "Nonlinear one-dimensional seismic waveform inversion using simulated annealing", *Geophysics*, **56**, 1624-1638.

Sen, M. K., and P. L. Stoffa, (1992), "Rapid sampling of model space using genetic algorithms: Examples from seismic waveform inversion", *Geophys. J. Int.*, **108**, 281-292.

Sen, M. K., and P. L. Stoffa, (1995), *Global Optimization Methods in Geophysical Inversion*, Elsevier Science Publications, Netherlands.

Sen, M. K., and P. L. Stoffa, (1996), "Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion", *Geophysical Prospecting*, **44**, 313-350.

Sheriff, R. E. (1984), *Encyclopedic dictionary of exploration geophysics*, Society of Exploration Geophysicists.

Singh, S. C., G. F. West, N. D. Bregman, and C. H. Chapman (1989), "Full waveform inversion of reflection data", *J. Geophys. Res*., **94**, 1777-1794.

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

Stoffa, P. L., and M. K. Sen, (1991), "Nonlinear multiparameter optimization using genetic algorithms: Inversion of plane wave seismograms", *Geophysics*, **56**(11): 1794-1810.

Stoffa, P. L., and M. K. Sen (1992), Seismic Waveform Inversion Using Global Optimization, *Journal of Seismic Exploration*, 1(1), 9-27.

Suzuki, J. (1998), A further result on the Markov chain model of Genetic Algorithms and its applications to a Simulated Annealing-like strategy, *IEEE Transactions on Systems, Man. and Cybernetics- Part B*: *Cybernetics*, **28**(1): 95-102.

Tarantola, A., (1987), *Inverse Problem Theory, Methods of Data Fitting and Model Parameter Estimation*, Elsevier Publishing Company.

Tatham, R. H., and M. D. McCormack, (1991), *Multicomponent Seismology in Petroleum Exploration*, Society of Exploration Geophysicists.

Tjelmeland, H. (1996), Stochastic models in reservoir characterization and Markov random fields for compact objects, *Ph.D. Dissertation*, Norwegian University of Science and Technology.

Tura, A., and D. Lumley (1999), "Estimating pressure and saturation changes from time-lapse AVO data", *1999 SEG Expanded Abstracts*.

Ulrych, T. J., M. D. Sacchi, and A. Woodbury (2001). "A Bayesian tour of inversion: A tutorial." *Geophysics*, **66**(1): 55-69.

Waggoner, J. R., A. Cominelli, and R. H. Seymour (2002), "Improved reservoir modeling with time-lapse seismic in a Gulf of Mexico gas condensate reservoir", SPE77956, in 2002 SPE Asia Pacific Oil and Gas Conference and Exhibition.

Wang, Z. (1997), "Feasibility of time-lapse seismic reservoir monitoring: The physical basis." *The Leading Edge*: 1327-1329.

Wang, Z. (2000), Seismic and acoustic velocities in reservoir rocks, 3: Recent development: Society of Exploration Geophysicists.

Wang, Z., and A. Nur (1992), Elastic wave velocities in porous media: A theoretical recipe in seismic and acoustic velocities in reservoir rocks, Theoretical and Model Studies: Society of Exploration Geophysicists, **2**, 1-35.

Wang, Z., and A. Nur (2000), The Gassmann equation revised: Comparing laboratory data with Gassmann's predictions: Seismic and acoustic velocities in reservoir rocks, recent developments: Society of Exploration Geophysicists, **3**, 1-35.

Wayland, J. R., and D. Lee (1986), Seismic mapping of EOR processes, *The Leading Edge*, **5**(12), 36-40.

Wheeler, M. F., and M. Peszynska, (2002), Computational Engineering and Science Methodologies for Modeling and Simulation of Subsurface Applications. *Advances in Water Resources*, **25**, 1147-1173.

Whitley, D., (1989), The Genetic Algorithms and selection pressure: Why rank-based allocation of reproductive trials is best, in J. D. Schaffer (Ed.), *Proceedings of the Third International Conference on Genetic Algorithms*, 116-123.

Wright, A.H. (1991), "Genetic Algorithms for Real Parameter Optimization", in *Foundations of Genetic Algorithms 1*, ed. G. J. E. Rawlins, San Mateo, CA: Morgan Kaufmann, pp. 205-218.

Yilmaz, O. (2000), *Seismic data analysis: Processing, inversion, and interpretation of seismic data*, Society of Exploration Geophysicists.

Yoon, S., A. H. Malallah, A. Datta-Gupta, D. W. Vasco, and R. A. Behrens (1999). "A multiscale approach to production data integration using streamline models." SPE 56653.

Ziolkowski, A., J. T. Fokkema, K. Koster, A. Confurius, and R. Vanboom (1989), Inversion of common mid-point seismic data, in P. L. Stoffa (Ed.), *Tau-p: A Plane Wave Approach to the Analysis of Seismic Data*, Kluwer Academic Publishers, 141-176.

Vita

Tiancong Hong was born at Xuzhou, Jiangsu Province, P. R. China, on March 17, 1973, to Chengyun Hong and Zhenying Lu. He completed both middle and high school at the Baji Middle School, Xuzhou. He entered the Department of Earth and Space Science at the University of Science and Technology of China in 1993. He completed his B.Sc. in Geophysics and a second B.Sc. in Computer Sciences in 1998. He then enrolled in the South China Sea Institute of Oceanology, Chinese Academy of Sciences to obtain a Master's degree. He obtained his M.Sc. in 2001 and then joined the MS program at the Department of Geosciences, Pennsylvania State University. He graduated from Penn State in 2004 and started working on his Ph.D. at the University of Texas at Austin in the Jackson School of Geosciences in the fall of 2005. During the summer of 2007, he worked for BP America, Houston. Upon graduation, he will join the Geophysics Quantitative Interpretation team at ExxonMobil Upstream Research Company, Houston, Texas. He is a member of SEG and AGU.

Permanent address: Baji Township, Xuzhou
Jiangsu Province, P. R. China 221361

This dissertation was typed by the author.