RESERVOIR PARAMETER ESTIMATION USING WAVELET ANALYSIS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PETROLEUM ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Pengbo Lu
May 2001

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Roland N. Horne
(Principal Adviser)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Khalid Aziz

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Jef Caers

Approved for the University Committee on Graduate Studies:

Abstract

This work focuses on several important issues of reservoir characterization and data integration using wavelet analysis. The goal of reservoir characterization is to estimate the spatial distribution of the reservoir properties, e.g., permeability and porosity, by proper integration of all types of data available (either static or dynamic). We considered the integration of production history data, seismic data, and well test data in this work.

One of the key issues in parameter estimation is to develop an efficient and reliable nonlinear regression procedure. We adapted wavelet analysis to describe the distribution of sensitivity coefficients, to gain advantage from the multiresolution properties of wavelets. Wavelet analysis can compress the model parameter space, stabilize the algorithm, and avoid local minima. This new approach also significantly improves the computational efficiency by varying the resolution of estimation at different regression stages.

Wavelet analysis also has the capability to integrate different types of data efficiently, using different levels of wavelets to incorporate different data types. We can account explicitly for the resolving power of different data and estimate reservoir properties with nonuniform resolution.

We have applied this newly developed procedure to some multiphase reservoir examples that demonstrate the reliability and flexibility of the approach. We also gained very good convergence rates and excellent computational efficiency compared to conventional methods. The most important conclusion of this work is that wavelet analysis is a very useful tool for reservoir parameter estimation, and can speed up the

computation and improve the performance. The nonlinear regression procedure with wavelet analysis has substantial advantages over the conventional algorithms.

Acknowledgments

I would like to thank my adviser Professor Roland N. Horne for his advice, guidance, and encouragement during the course of this research. My girl friend, Jing Wan, also my colleague, always gives me her wholehearted support and help. My gratitude to her is beyond description. I also wish to express my appreciation to my friends and colleagues, Jorge Landa, Yan Pan and Pengju Wang for generously providing good ideas and techniques for this research.

Financial support from Stanford University Petroleum Research Institute (SUPRI-D) and the Department of Petroleum Engineering are gratefully acknowledged. Financial support from the Stanford Graduate Fellowship is also gratefully acknowledged.

I would also like to thank my family for their love and support. My special gratitude is endless to my parents, Zhijian Lu and Huaiyun Wang, and to my brother Zhaohui.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Statement of the Problem

The most commonly encountered and probably the most challenging work in reservoir engineering is to describe the reservoir accurately and efficiently. An accurate description of a reservoir is crucial to the production management and oil recovery. The goal of reservoir characterization is to estimate the spatial distribution of the reservoir properties, e.g., permeability and porosity, by integrating all kinds of information available. The information is usually divided into two categories: static information and dynamic information. The static information includes any terminological information on the spatial distribution from geological knowledge (core, log, seismic and geologic information, etc., they need not come from actual reservoirs). The dynamic information includes production data, pressure transient data, tracer testing data, long-term pressure history, etc. Recently, the introduction of 4-D seismic surveys has provided another type of dynamic data.

Reservoir characterization is a process of inferring the parameters from indirect measurements (or observations). Thus the process is posed as an inverse problem, or parameter estimation problem, which is commonly encountered in the fields of optimization, operations research and economics systems. Although there are many mathematical techniques available for this kind of problem, the reservoir characterization is unique and extremely difficult for many reasons. First of all, permeability and

porosity are the parameters that have the largest influence on reservoir performance. Usually we can not obtain any direct measurements of these parameters (the formations are often several thousand feet deep underground). Even if we can make some direct measurements, for example, using well logs, the sampling locations are sparse and the amount of information is scarce. Secondly, direct and indirect measurements are obtained with different technologies and reflect different aspects of reservoirs at different scales. Furthermore, due to the complex nature of the reservoir, we usually need to use a numerical reservoir simulation as the mathematical model of the reservoir behavior. It is problematic to implement an efficient and robust inverse procedure based on the numerical simulator. One reason is that the computational load is extremely heavy, another reason is that the inverse problem is ill-posed due to the scarcity of information in comparison to the number of unknown parameters in the simulation grid.

In the context of reservoir engineering, the parameter estimation problem has several key issues to be addressed: parameterization of the system, data integration, inversion procedure, and preservation of geological structures. Although many studies have focused on these issues, the problems are far from being solved and are currently the subjects of major research effort in several places. The current work addresses several of these important issues by incorporating inverse problem theory and wavelet analysis. The forward mathematical model is a finite-difference reservoir simulator, and the inverse approach is a gradient-based Gauss-Newton algorithm.

1.2 Literature Review

1.2.1 Sensitivity Coefficients and Gradient-Based Methods

Reservoir characterization has been a major research subject in reservoir engineering since the 1960s. There are numerous papers available and a significant number of ongoing studies; a brief review is given here to summarize the recent related works. Reservoir characterization is a broad concept, for instance, well logging is one type of reservoir characterization used to infer reservoir properties. In this study, we

will restrict our scope to parameter estimation using automatic history matching procedures with numerical simulators.

Jacquard and Jain (1965) introduced an automated history matching procedure adapted from variational analysis in electric networking. Since then, there have been several developments of concepts and algorithms along similar lines. In 1974, Carter, Pierce, Kemp & Williams (1974) introduced the concept of sensitivity coefficients and derived a method to calculate them. Carter's method has high computational efficiency, although with the limitation that it is only applicable to linear problems.

In 1970s and early 1980s, several important works (Chen, Gavalas, Seinfeld & Wasserman 1973, Watson, Seinfeld, Gavalas & Woo 1979, Yang and Watson 1987) extended the scope of parameter estimation procedures using optimal control theory. These studies developed a more efficient method to compute the gradient of the objective function with respect to permeability and porosity without calculating the sensitivity coefficients. This method is computationally efficient, because it calculates the gradient of a form of the objective functions instead of sensitivity coefficients. However, this method is difficult to implement for numerical calculation. The parameter estimation algorithms using this method are also not as efficient as those using sensitivity coefficients, such as the Gauss-Newton algorithm.

In 1982, Tarantola and Valette (1982) published an important paper in the field of geophysical inverse theory. This paper provided a set of algorithms for generalized nonlinear inverse problems, using the least-square criterion. In 1987, Tarantola (1987) provided a comprehensive exposition for inverse problem theory for data fitting and model parameter estimation. Tarantola's work also provides a method to include *a priori* information into the definition of the objective function, under the assumption of multi-Gaussian probability distributions. Menke (1984) and Parker (1994) have also made contributions in this field.

Tang, Chen, Chen & Wasserman (1989) and Chen, Gavalas, Seinfeld & Wasserman (1973) presented the GPST (Generalized Pulse-Spectrum Technique) for history matching. The GPST does not compute sensitivity coefficients and is difficult to apply to certain parameter estimation algorithms, though Oliver (1994a) used the GPST approach to incorporate transient pressure data into reservoir characterization.

Starting in 1994, a series of papers (Reynolds, He, Chu & Oliver 1995, Chu, Reynolds & Oliver 1995, Chu & Komara 1996, He, Reynolds & Oliver 1996, Wu, Reynolds & Oliver 1998) used the modified GPST method to calculate the sensitivity coefficients. The efficient computation of sensitivity coefficients enabled the application of the Gauss-Newton algorithm for parameter estimation and the geophysical inverse theory to include *a priori* information. About the same time, Anterion, Eymard & Karcher (1989) also developed a similar procedure to compute sensitivity coefficients. This method was referred to as the gradient simulator. There have been several applications in the literature (Bissel, Sharma & Killough 1994, Bissel 1996, Tan & Kalogerakis 1991, Tan 1995) using this approach.

Streamline simulators offer substantial computational efficiency and minimize numerical diffusion compared to traditional finite-difference methods (Thiele, Batycky & Blunt 1997). Therefore, a streamline simulator will hold some advantages over a finite-difference simulator as forward modeling, in the sense of computational efficiency. Many studies have shown favorable results from integrating dynamic data into reservoir modeling using streamline simulators (e.g. Datta-Gupta, Vasco & Long 1995, Vasco, Yoon & Datta-Gupta 1998). Furthermore, a streamline simulator can provide semi-analytical sensitivity coefficients for watercuts at wells.

Recently, several students in SUPRI-D (Stanford University Petroleum Research Institute: Consortium on Innovation in Well Testing) have been working on reservoir characterization and data integration in different aspects. In 1997, Landa (Landa 1997, Landa & Horne 1997) investigated the impact of different data on reservoir characterization and uncertainty. Landa also studied the object-based approach, which can reduce the number of unknown parameters tremendously. In 1998, Phan (1998) looked into depth-dependent reservoir properties. In 1999, Wang (1999) presented an application integrating resistivity data into the parameter estimation problem.

1.2.2 Reparameterization Techniques

For any problem with a large number of model parameters, it is very difficult to apply gradient-based algorithms. First of all, the inverse problem will be under-determined

with too many parameters, while the available information does not have enough resolving power. Secondly, the computational overhead for gradient calculation is usually proportional to the number of parameters. As the number of parameters increases, the computational load will be too huge to be practical. Furthermore, it takes far more iterations for the nonlinear regression to reach convergence with a huge number of parameters.

These problems with gradient-based algorithms can be alleviated or even eliminated with reparameterization techniques. In reparameterization approaches, the model parameter space (usually the permeability and porosity fields) are written as linear combinations of a relatively small set of bases: $\vec{\alpha} = A \cdot \vec{c}$, where A is the transformation matrix (set of basis vectors), $\vec{\alpha}$ is the original model parameter vector, and $\vec{c}$ is the reduced parameter vector.

The simplest reparameterization is zonation, first introduced by Jacquard and Jain (1965), in which a reservoir is divided into blocks larger than simulation gridblocks. In other words, instead of using the simulation gridblock properties as model parameters, a zonation technique uses the properties of the large blocks (zones) as model parameters, which imposes a uniform distribution over each zone. Bissel, Sharma & Killough (1994) proposed a variation of the zonation technique, in which the reservoir is divided into high-sensitivity zones and a fast initial reduction in the objective function can often be achieved.

Another reparameterization technique is to use the eigenvectors with the largest eigenvalues of the covariance matrix as model parameters (e.g. Gavalas, Shah & Seinfeld 1976, Oliver 1996, Reynolds, He, Chu & Oliver 1995). This method is effective only if the eigenvalues decay rapidly. Similarly, singular value decomposition (SVD) has been used to choose the set of basis vectors dynamically (e.g. Shah, Gavalas & Seinfeld 1978, Dietrich 1990). This approach might give optimal set of basis vectors at each iteration. The subspace method (Oldenburg *et al.* 1993), which is related to SVD methods, can reduce computational overhead of sensitivity coefficient calculation and singular value decomposition.

de Marsily *et al.* (1984) introduced the pilot point method, which is another form of reparameterization (as stated by McLaughlin & Townley 1996). Pilot-point methods

(and related master-point methods) have been used extensively in groundwater field to reduce the number of parameters in history matching (e.g. LaVenue & Pickens 1992, LaVenue *et al.* 1995). An iterative geostatistically-based inverse technique, called the sequential self-calibrated (SSC) method, has been developed and previously shown to be quite efficient for constructing multiple equiprobable reservoir models that honor single-phase historical pressure data (Gómez-Hernández, Sahuquillo & Capilla 1998, Capilla, Gómez-Hernández & Sahuquillo 1998). The SSC utilizes the concept of master point to reduce the parameter space to be estimated in the optimization process. Wen, Deutsch & Cullick (1997, 1998) extended the SSC and proposed a fast streamline-base inverse method to integrate multiphase dynamic data. In this approach, the spatial correlation structure is honored through the kriging equations in the SSC. This leads to better-behaved objective functions with low final values and preserves the prior model spatial characteristics.

1.2.3 Wavelet Analysis

The development of wavelets is fairly recent in applied mathematics, although wavelets have already had a remarkable impact in the fields of signal processing, image analysis and data compression. There are numerous publications on this subject, we will only review some central works. Ogden (1997) presented an overview of some essential aspects of wavelets for statistical applications and data analysis. Mallat (1998) gave a quick tour of the use of wavelets in signal processing. Both publications provide substantial background on both theoretical and practical aspects. Foufoula-Georgiou & Kumar (1994) reviewed some important papers in the field of geophysics, most of which focus on seismic data interpretation and multiscale analysis of rock structures.

In the late 1990s, there were some applications of wavelets in reservoir engineering. Panda, Mosher & Chopra (1996) and Chu, Schatzinger & Tham (1996) presented some applications of wavelets for reservoir data analysis and property scaling, although their works focused mainly on reservoir property upscaling and multiresolution analysis. In 1997, Kikani & He (1998) investigated long-term pressure transient

data using wavelet methods. Later on, in 1999, Athichanagorn (1999) used wavelet methods to interpret long-term pressure data from permanent downhole gauges.

1.3 Overview

The methodologies and algorithms discussed so far addressed some most important aspects of reservoir characterization. However, early works were still problematic. For instance, gradient-based methods can give fast convergences, but are highly dependent of sensitivity calculation and may have huge computational overhead. Furthermore, most inverse problems are ill-posed with a large number of parameters. The reparameterization techniques can compress the model parameter space and improve the efficiency of gradient-based inverse algorithms to some extent. However, there are still some potential problems with previous approaches (Oliver *et al.* 2001):

- *Zonation technique* may result in discontinuous reservoir properties at zonation boundaries.
- *Spectrum and singular value decomposition* have strong dependency on eigenvalue decaying patterns. Furthermore, singular value decomposition needs the calculation of all sensitivity coefficients and a huge matrix decomposition. The subspace method can avoid the computational overhead, though with a trade-off of slower convergence.
- *Pilot-point and master-point methods* can result in unappealing visual artifacts at the pilot point locations and may cause errors in quantification of uncertainty (Omre *et al.* 1999).

In this study, we applied the wavelet analysis as a new reparameterization technique. The primary reason is that the wavelet analysis has some very attractive properties, such as multiresolution analysis, good time-space localization, and fast algorithms. Our study shows that wavelet analysis can compress the model parameter space, stabilize the algorithm, and avoid local minima. Our new approach also significantly improves the computational efficiency by varying the resolution of estimation at different regression stages.

Wavelet analysis also has the capability to integrate different types of data efficiently, using different levels of wavelets to incorporate different data types. We can account explicitly for the resolving power of different data and estimate reservoir properties with nonuniform resolution.

In the following chapters, we will discuss our new approach in more detail. The organization of this thesis is as follows:

- *Chapter 2* describes some mathematical preliminaries and background information, such as general inverse problems and basic concepts of wavelet analysis.
- *Chapter 3* shows the reservoir modeling and reparameterization techniques. More specifically, we will discuss how to parameterize reservoirs using wavelets.
- *Chapter 4* demonstrates how wavelets can be applied to reservoir characterization to yield a relatively small set of parameters to describe a reservoir. Also, we will look into the feasibility of integrating different types of data using wavelets.
- *Chapter 5* shows the methodology and implementation with applications on a simple synthetic reservoir.
- *Chapter 6* gives some more applications and test cases to demonstrate the efficiently and applicability of our new approach.
- *Chapter 7* concludes this thesis and addresses several open issues on this subject.

Chapter 2

Mathematical Preliminaries

2.1 General Discrete Inverse Problem

2.1.1 Model Space and Data Space

Let S be the *physical system* under study. In this work, S is the reservoir formation. As stated by Tarantola (1987), the scientific procedure for the study of a physical system can be divided into the following three steps:

1. *Parameterization of the system*: discovery of a minimal set of *model parameters* whose values completely characterize the system (from a given point of view).

2. *Forward modeling*: predictions that the *physical laws* governs, as to the results of measurements on some *observable parameters*, for given values of the model parameters.

3. *Inverse modeling*: use of the actual results of some measurements of the observable parameters to infer the actual values of the model parameters.

Because numerical reservoir simulators have been well defined and are widely used in the oil industry, in this study, a finite-difference reservoir simulator was used as the forward modeling tool, as well as the sensitivity-coefficient generator. Thus this study focuses on the first and the third steps: parameterization and inverse modeling.

Throughout this work, we denote the model parameters as $\vec{\alpha}$ in the parameter space **M**, and the observable parameters (or data) as $\vec{d}$ in the data space **D**.

2.1.2 Probability Distribution

The measurement experiment will give a certain amount of information as to the true values of the observable parameters. Let $f_D(\vec{d})$ be the probability density function describing this information. *A priori* information is the information obtained independently of the results of measurements. Let $f(\vec{d},\vec{\alpha})$ be a joint probability density function on the parameters $(\vec{d},\vec{\alpha})$. We can also define the marginal probability density functions

$$f_M(\vec{\alpha}) = \int_D d\vec{d}\, f(\vec{d},\vec{\alpha}), \tag{2.1}$$

$$f_D(\vec{d}) = \int_M d\vec{\alpha}\, f(\vec{d},\vec{\alpha}), \tag{2.2}$$

and the conditional probability density function

$$f_{M|D}(\vec{\alpha}|\vec{d}^{obs}) = f(\vec{d}^{obs},\vec{\alpha})/f_D(\vec{d}^{obs}), \tag{2.3}$$

$$f_{D|M}(\vec{d}|\vec{\alpha}) = f(\vec{d},\vec{\alpha})/f_M(\vec{\alpha}). \tag{2.4}$$

Following the Bayesian theorem, we also have

$$f_{M|D}(\vec{\alpha}|\vec{d}^{obs}) = f_{D|M}(\vec{d}^{obs}|\vec{\alpha})f_M(\vec{\alpha})/f_D(\vec{d}^{obs}) \tag{2.5}$$

Let $\theta(\vec{d}|\vec{\alpha})$ be the conditional probability density function describing the theoretical relationship between $\vec{d}$ and $\vec{\alpha}$ and let $\nu(\vec{d}^{obs}|\vec{d})$ be the probability density function for the output of a measuring instrument to be $\vec{d}^{obs}$ when the input is $\vec{d}$. Then the conditional *probability distribution function* defined in Eq. 2.4 is given by

$$f_{D|M}(\vec{d}^{obs}|\vec{\alpha}) = \int_D d\vec{d}\,\nu(\vec{d}^{obs}|\vec{d})\theta(\vec{d}|\vec{\alpha}). \tag{2.6}$$

Then Eq. 2.3 gives

$$f_{M|D}(\vec{\alpha}|\vec{d}^{obs}) = f_M(\vec{\alpha})\int_D d\vec{d}\,\nu(\vec{d}^{obs}|\vec{d})\theta(\vec{d}|\vec{\alpha})/f_D(\vec{d}^{obs}) \tag{2.7}$$

2.1.3 Some Special Cases

We can further simplify Eq. 2.7 in some special cases.

1. *Errors of the measurement are independent of the input, and errors in the theory are independent of the model value.* In this case, we have

$$\vec{d}^{obs} = \vec{d} + \vec{\epsilon}_d, \tag{2.8}$$
$$\vec{d} = \vec{d}^{cal} + \vec{\epsilon}_T, \tag{2.9}$$

and the theoretical model is

$$\vec{d}^{cal} = \vec{g}(\vec{\alpha}), \tag{2.10}$$

where $\vec{\epsilon}_d$ and $\vec{\epsilon}_T$ are unknown errors with known probability distribution functions $f_d(\vec{\epsilon}_d)$ and $f_T(\vec{\epsilon}_T)$. Then

$$\nu(\vec{d}^{obs}|\vec{d}) = f_d(\vec{d}^{obs} - \vec{d}), \tag{2.11}$$
$$\theta(\vec{d}|\vec{\alpha}) = f_T(\vec{d} - \vec{d}^{cal}). \tag{2.12}$$

Eq. 2.7 then gives

$$f_{M|D}(\vec{\alpha}|\vec{d}^{obs}) = f_M(\vec{\alpha}) f(\vec{d}^{obs} - \vec{d}^{cal}) / f_D(\vec{d}^{obs}), \tag{2.13}$$

where $f(\vec{\epsilon})$ is the convolution product of $f_d(\vec{\epsilon}_d)$ and $f_T(\vec{\epsilon}_t)$,

$$f(\vec{\epsilon}) = f_d * f_T(\vec{\epsilon}). \tag{2.14}$$

2. *Gaussian modeling and observation errors.* This case is a step further beyond the previous case. In this case, we have

$$\nu(\vec{d}^{obs}|\vec{d}) = \mathcal{N}_{\vec{d}^{obs}}(\vec{d}, C_d), \tag{2.15}$$
$$\theta(\vec{d}|\vec{\alpha}) = \mathcal{N}_{\vec{d}}(\vec{d}^{cal}, C_T), \tag{2.16}$$

and

$$f_{D|M}(\vec{d}^{obs}|\vec{\alpha}) = \mathcal{N}_{\vec{d}^{obs}}(\vec{d}^{cal}, C_D). \tag{2.17}$$

with

$$C_D = C_d + C_T. \tag{2.18}$$

$\mathcal{N}_{\vec{x}}(\vec{\mu}, C_{\vec{x}})$ is the Gaussian distribution defined as

$$\mathcal{N}_{\vec{x}}(\vec{\mu}, C_{\vec{x}}) = \frac{1}{\sqrt{(2\pi)^N |C_{\vec{x}}|}} \exp[-\frac{1}{2}(\vec{x}-\vec{\mu})^T C_{\vec{x}}^{-1}(\vec{x}-\vec{\mu})]. \tag{2.19}$$

Eq. 2.13 becomes

$$f_{M|D}(\vec{\alpha}|\vec{d}^{obs}) = f_M(\vec{\alpha})\mathcal{N}_{\vec{d}^{obs}}(\vec{d}^{cal}, C_D)/f_D(\vec{d}^{obs}). \tag{2.20}$$

If the model parameters also follow the Gaussian distribution, we would have

$$f_M(\vec{\alpha}) = \mathcal{N}_{\vec{\alpha}}(\vec{\alpha}^{pri}, C_M), \tag{2.21}$$

thus Eq. 2.20 becomes

$$f_{M|D}(\vec{\alpha}|\vec{d}^{obs}) = \text{const} \cdot \exp[-E(\vec{\alpha})], \tag{2.22}$$

where $E(\vec{\alpha})$ is the objective function defined as

$$E(\vec{\alpha}) = \frac{1}{2}[(\vec{d}^{cal} - \vec{d}^{obs})^T C_D^{-1}(\vec{d}^{cal} - \vec{d}^{obs}) + (\vec{\alpha} - \vec{\alpha}^{pri})^T C_M^{-1}(\vec{\alpha} - \vec{\alpha}^{pri})]. \tag{2.23}$$

The goal of the inverse problem is to obtain a maximum likelihood estimation, or to minimize the objective function, as demonstrated in Section 2.2.

2.2 Maximum Likelihood Methods (l_2 norm)

2.2.1 Basic Concepts

In this study, we focus on the least-square (l_2 norm) criterion, which we have justified in Section 2.1. The basic assumption is that all the probability distributions are Gaussian, including the measurement errors, the theory errors and the model parameters. Least-square methods are popular for solving inverse problems because they lead to the easiest computations. One drawback is the lack of robustness, i.e., the strong sensitivity of the least-square approach to a small number of large errors (outliers) in a data set.

If the forward problem is linear, we have

$$\vec{d}^{cal} = \vec{g}(\vec{\alpha}) = G\vec{\alpha}, \tag{2.24}$$

the *a posteriori* probability density function $f_M(\vec{\alpha})$ is then Gaussian:

$$f_M(\vec{\alpha}) = \mathcal{N}_{\vec{\alpha}}(< \vec{\alpha} >, C_M^{est}), \tag{2.25}$$

with center

$$< \vec{\alpha} >= \vec{\alpha}^{pri} + (G^T C_D^{-1} G + C_M^{-1})^{-1} G^T C_D^{-1} (\vec{d}^{obs} - G\vec{\alpha}^{pri}), \tag{2.26}$$

and covariance matrix

$$C_M^{est} = (G^T C_D^{-1} G + C_M^{-1})^{-1}. \tag{2.27}$$

The value $< \vec{\alpha} >$, center of the Gaussian, is both the mean value of $f_M(\vec{\alpha})$ and its maximum likelihood point. Therefore, for linear problems, we have explicit expressions for the solution and for the *a posteriori* covariance matrix.

The Gauss-Markov theorem (Rao 1973) shows that, for linear problems, the least-square estimator has minimum variance among all the estimators that are linear functions of $\vec{d}^{obs}$ and $\vec{\alpha}^{pri}$, irrespective of the particular form of the probability density functions of the random variables $\vec{d}$ and $\vec{\alpha}$. This is why the least-square criterion is sometimes used even if the form of the density functions is not Gaussian.

If the forward problem is nonlinear, the *a posteriori* probability density function is Gaussian, and the analysis of the solution is not so straightforward. If $f_M(\vec{\alpha})$ is reasonably well behaved, the *a posteriori* information in the model space may be well represented by a central estimator of $f_M(\vec{\alpha})$ and a properly defined covariance matrix. Among all the central estimators, the easiest to compute is generally the maximum likelihood point $\vec{\alpha}_{ML}$:

$$f_M(\vec{\alpha}) \quad \text{MAXIMUM for} \quad \vec{\alpha} = \vec{\alpha}_{ML}. \tag{2.28}$$

As defined in Eq. 2.22, this is equivalent to

$$E(\vec{\alpha}) \quad \text{MINIMUM for} \quad \vec{\alpha} = \vec{\alpha}_{ML}. \tag{2.29}$$

2.2.2 Gradient-Based Optimization Algorithms

There are many nonlinear regression algorithms available. In this study, we used gradient-based methods, mainly because algorithms in this category yield faster convergence. On the other hand, gradient-based algorithms require calculations of gradients, which introduce heavy computational overhead. Section 2.3 will outline the modified generalized pulse-spectrum technique (GPST), which makes the calculation of gradients (or sensitivity coefficients) much more efficient. The GPST technique also makes gradient-based algorithms feasible for this study.

In this section, several different algorithms will be outlined and compared. As mentioned earlier, our goal is to minimize the objective function $E(\vec{\alpha})$. Assuming the function is smooth enough, this objective function can be approximated in the neighborhood of $\vec{\alpha}_n$ with a Taylor's expansion

$$\begin{aligned} E(\vec{\alpha}) &= E(\vec{\alpha}_n + \delta\vec{\alpha}) \\ &= E(\vec{\alpha}_n) + \vec{F}_n^T \delta\vec{\alpha} + \frac{1}{2}\delta\vec{\alpha}^T H_n \delta\vec{\alpha} + O(\delta\vec{\alpha}^3) \\ &\approx E(\vec{\alpha}_n) + \vec{F}_n^T \delta\vec{\alpha} + \frac{1}{2}\delta\vec{\alpha}^T H_n \delta\vec{\alpha}, \end{aligned} \tag{2.30}$$

and

$$\nabla E(\vec{\alpha}) \approx \vec{F}_n + H_n \delta\vec{\alpha}, \tag{2.31}$$

where

$$\vec{F}_n = \nabla E(\vec{\alpha}_n) \tag{2.32}$$

is the gradient vector and

$$H_n = \nabla \vec{F}_n = \nabla\nabla E(\vec{\alpha}_n) \tag{2.33}$$

is the second-derivative matrix, or the Hessian matrix. The core concept of any gradient-based method is to find a step $\delta\vec{\alpha}$ to reduce the objective function $E(\vec{\alpha})$, on the basis of the gradient $\vec{F}$.

For least-square minimization, there are mainly three categories of methods: steepest descent and conjugate gradient methods, Newton and Gauss-Newton methods, and quasi-Newton (variable metric) methods. Oliver *et al.* (2001) gave an overview of different methods in the context of reservoir engineering.

Steepest Descent and Conjugate Gradient Methods

The most straightforward gradient-based method is the steepest descent method, in which the direction of $\delta\vec{\alpha}_n$ is given by

$$-\vec{F}_n = -\nabla E(\vec{\alpha}_n), \tag{2.34}$$

which is the local downhill gradient. Thus $\vec{\alpha}_{n+1} = \vec{\alpha}_n + \delta\vec{\alpha}_n$ minimizes $E(\vec{\alpha})$ on the line given by $-\vec{F}_n$. and we have

$$\nabla E(\vec{\alpha}_{n+1}) \cdot (-\vec{F}_n) = (-\vec{F}_{n+1}) \cdot (-\vec{F}_n) = 0. \tag{2.35}$$

Therefore, any two consecutive steps in the steepest descent method are perpendicular to each other. The problem with the steepest descent method is that the method will perform many small steps in going down a long, narrow valley, even if the valley is in a perfect quadratic form.

Conjugate gradient methods form a class of algorithms that generate directions of search directions without storing a matrix, in contrast to the Newton-type methods we will discuss in the next section (Gill, Murry & Wright 1981). Conjugate gradient methods are essential in circumstances when methods based on matrix factorization are not viable because the relevant matrix is too large or too dense.

The procedure of the conjugate gradient algorithm is outlined as follows. Starting with an arbitrary initial vector $\vec{g}_0 = -\nabla E(\vec{\alpha}_0)$ and letting $\vec{h}_0 = \vec{g}_0$, the algorithm constructs two sequences of vectors from the recurrence

$$\vec{g}_{n+1} = -\nabla E(\vec{\alpha}_{n+1}) \tag{2.36}$$

$$\vec{h}_{n+1} = \vec{g}_{n+1} + \gamma_n \vec{h}_n, \tag{2.37}$$

where $n = 0, 1, 2, ...$, and the iterative procedure is

$$\vec{\alpha}_{n+1} = \vec{\alpha}_n + \lambda_n \vec{h}_n. \tag{2.38}$$

The scalar λ_n can be solved by minimizing $E(\vec{\alpha}_{n+1})$. γ_n is given by

$$\gamma_{n+1} = \frac{\vec{g}_{n+1} \cdot \vec{g}_{n+1}}{\vec{g}_n \cdot \vec{g}_n}. \tag{2.39}$$

Therefore, this algorithm does not require knowledge of the Hessian matrix H. A sequence of directions $\vec{h}_n$ is constructed using only line minimizations, evaluations of the gradient vector, and an auxiliary vector to store the latest in the sequence of $\vec{g}$'s.

The algorithm described so far is the original Fletcher-Reeves version of the conjugate gradient algorithm. Polak (1971) introduced one small but also significant change, where γ_n is given by

$$\gamma_{n+1} = \frac{(\vec{g}_{n+1} - \vec{g}_n) \cdot \vec{g}_{n+1}}{\vec{g}_n \cdot \vec{g}_n}. \tag{2.40}$$

The conjugate gradient algorithm was derived on the basis that the objective function is in a quadratic form. If an objective function $E(\vec{\alpha})$ is not in a quadratic form, two modifications need to be introduced. First of all, the scalar λ_n can not be solved analytically and must be calculated by an iterative process. Secondly, a restart procedure is required to reset $\vec{h}_n$ to the steepest descent direction $\vec{g}_n$ after a certain number of iterations.Gill, Murry & Wright (1981) have shown that the conjugate gradient algorithm is theoretically $n-$step superlinearly convergent. However, the conjugate gradient method is nearly always linearly convergent, regardless of whether or not restarting takes place. Although conjugate-gradient-type algorithms are far from ideal, they are currently the only reasonable method available for a general problem in which the number of variables is extremely large.

Newton and Gauss-Newton Methods

We can use the second-order approximation to calculate a step to reduce E. Setting the gradient to zero in Eq. 2.31, we have

$$H_n \delta\vec{\alpha}_n = -\vec{F}_n, \tag{2.41}$$

and

$$\vec{\alpha}_{n+1} = \vec{\alpha}_n + \delta\vec{\alpha}_n \tag{2.42}$$

The method described by Eqs. 2.41 and 2.42 is known as Newton's method. It is well known that this method does not give a direction of descent unless the matrix H_n is positive-definite. Besides, Newton's method has a major drawback: second-order

derivatives are required to form the Hessian matrix H_n, which therefore can be very expensive or even impossible to calculate in practice. The Gauss-Newton algorithm as used in this work avoids the computation of second derivatives.

The objective function is shown in Eq. 2.23, therefore we can write the gradient vector $\vec{F}_n$ and the second-order matrix H_n of Newton's method as

$$\vec{F}_n = \nabla E_n = G_n^T C_D^{-1}(\vec{d}_n^{cal} - \vec{d}^{obs}) + C_M^{-1}(\vec{\alpha}_n - \vec{\alpha}^{pri}), \tag{2.43}$$

and

$$H_n = \nabla \vec{F}_n = G_n^T C_D^{-1} G_n + C_M^{-1} + \nabla G_n^T \cdot C_D^{-1}(\vec{d}^{obs} - \vec{d}_n^{cal}), \tag{2.44}$$

where

$$G_n = \frac{\partial \vec{d}_n^{cal}}{\partial \vec{\alpha}} \tag{2.45}$$

is the sensitivity coefficient matrix.

In the Gauss-Newton method, we can further simplify Eq. 2.44 by omitting the last right hand side term. This approximation holds as far as the second-order derivatives of the errors are negligible. Thus we have the Gauss-Newton Hessian matrix as

$$H_n^{GN} = G_n^T C_D^{-1} G_n + C_M^{-1}. \tag{2.46}$$

Finally, the regressive iteration is written as

$$\vec{\alpha}_{n+1} = \vec{\alpha}_n - \mu_n H_n^{-1} \nabla E_n, \tag{2.47}$$

where μ_n is the line-search coefficient or the damping factor.

A popular variation on the Gauss-Newton method is Levenberg-Marquardt method. The Levenberg-Marquardt search direction is given by

$$(H_n^{GN} + \nu_n I)\delta\vec{\alpha}_n = -\vec{F}_n, \tag{2.48}$$

where ν_n is a non-negative scalar. The Levenberg-Marquardt method improves the condition of the Hessian matrix and usually results in faster convergence. This variation of the Gauss-Newton approach was used in this study.

The major advantages of Gauss-Newton algorithm can be summarized as follows:

- *Positive-definite Hessian matrix.* The Hessian matrix is positive-definite, and the inclusion of the C_D further stabilizes the Hessian matrix. Therefore the direction $\delta\vec{\alpha}$ is always descending.

- *Quadratic convergence.* In the neighborhood of the true solution, we have $H^{GN} \rightarrow H$. Because the Newton method has quadratic convergence, the Gauss-Newton method will also have quadratic convergence near the true solution.

- *Only first derivatives.* We only need to calculate the first derivatives (or sensitivity coefficient matrix G).

The Gauss-Newton formulation also has some disadvantages. First of all, we still need to calculate sensitivity coefficients, which can be time-consuming in the case of reservoir characterization. The problem can be alleviated by using the modified generalized pulse-spectrum technique (GPST) to calculate sensitivity coefficients. Secondly, we need to construct the Hessian matrix, which is a dense matrix and can be huge if the number of parameters is large.

Quasi-Newton (Variable Metric) Methods

The goal of variable metric methods (also called quasi-Newton methods), is not much different from the goal of conjugate gradient methods: to accumulate information from successive line minimizations so that N such line minimizations lead to the exact minimum of a quadratic form in N dimensions. The variable metric methods differ from the conjugate gradient methods in the way that it stores and updates the information that is accumulated. The basic idea of the variable metric methods is to build up, iteratively, a good approximation to the inverse Hessian matrix H^{-1}, without storing or solving the Hessian matrix. The process is to construct a sequence of matrices A_i, with the property

$$\lim_{n\rightarrow\infty} A_n = H^{-1}. \tag{2.49}$$

Press, Teukolsky, Vetterling & Flannery (1992) showed that the variable metric methods do not hold any overwhelming advantage over the conjugate gradient

methods, except perhaps a historical one. Therefore, we will not get into details of this approach. Yang and Watson (1987) investigated the use of four variable metric methods: the BFGS method, Fletcher's switch algorithm, a self-scaling variable metric method, and an optimally conditioned self-scaling variable metric method.

In general, the hybrid methods combining steepest descent with Gauss-Newton are most efficient for relatively small problems and conjugate gradient or quasi-Newton methods would be required for large practical problems (Deschamps, Grussaute, Mayers & Bissel 1998).

One major objective of this study is to reduce the number of model parameters using reparameterization techniques (specifically, the wavelet analysis in this study). Thus the Gauss-Newton can still be used effectively for larger problems with a small parameter set. Furthermore, different set of model parameters are used at each iteration in our approach. Therefore, it is not very convenient to apply either conjugate gradient methods or quasi-Newton methods.

2.3 Generalized Pulse-Spectrum Technique

As mentioned in Section 2.2.2, we need to calculate sensitivity coefficients. In this section, a brief description is given of an efficient algorithm to calculate sensitivity coefficients.

Tang, Chen, Chen & Wasserman (1989) presented an approximate method for solving inverse problems by a method called the generalized pulse-spectrum technique (GPST) for history matching. The GPST does not calculate sensitivity coefficients and is difficult to apply to certain parameter estimation algorithms. Chu, Reynolds & Oliver (1995) proposed the modified GPST to calculate sensitivity coefficients for two-dimensional, single-phase reservoirs, which is slightly different from the original GPST. Landa, Kamal, Jenkins & Horne (1996) presented another version of modified GPST, which is suitable for two-dimensional, two-phase black-oil models. In this study, a similar approach was used with some slight modification for three-dimensional, multiblock-completion models.

A discrete reservoir system can be written as a set of partial differential equations

$$\vec{f}^{(k+1)} = \vec{f}(\vec{u}^{(k+1)}, \vec{u}^{(k)}, \vec{v}, \Delta t, \vec{\alpha}) = \vec{0}, \tag{2.50}$$

where $\vec{u}^{(k)}$ are reservoir response at timestep k (well pressure, watercut, saturation distribution, etc.), $\vec{v}$ are known reservoir properties, $\vec{\alpha}$ are unknown reservoir properties (permeability, porosity, etc.), or model parameters. If we impose a small perturbation $\delta\vec{\alpha}$ on the model parameters, we have

$$\vec{f}(\vec{u}^{(k+1)} + \delta\vec{u}^{(k+1)}, \vec{u}^{(k)} + \delta\vec{u}^{(k)}, \vec{v}, \Delta t, \vec{\alpha} + \delta\vec{\alpha}) = \vec{0}. \tag{2.51}$$

We can expand Eq. 2.51 into a Taylor's series as follows

$$\vec{f}^{(k+1)} + \frac{\partial \vec{f}^{(k+1)}}{\partial \vec{u}^{(k+1)}}\delta\vec{u}^{(k+1)} + \frac{\partial \vec{f}^{(k+1)}}{\partial \vec{u}^{(k)}}\delta\vec{u}^{(k)} + \frac{\partial \vec{f}^{(k+1)}}{\partial \vec{\alpha}}\delta\vec{\alpha} + O(\delta^2) = \vec{0}. \tag{2.52}$$

The first right hand side term is zero from Eq. 2.50. If we drop high-order terms $O(\delta^2)$, Eq. 2.52 is written as

$$J^{(k+1)}\delta\vec{u}^{(k+1)} = -D^{(k+1)}\delta\vec{u}^{(k)} - Y^{(k+1)}\delta\vec{\alpha}, \tag{2.53}$$

where

$$J^{(k+1)} = \frac{\partial \vec{f}^{(k+1)}}{\partial \vec{u}^{(k+1)}} \tag{2.54}$$

is the Jacobian matrix, which can be obtained from the simulator, and

$$D^{(k+1)} = \frac{\partial \vec{f}^{(k+1)}}{\partial \vec{u}^{(k)}} \tag{2.55}$$

is a block-diagonal matrix, which is very sparse and easy to calculate, and

$$Y^{(k+1)} = \frac{\partial \vec{f}^{(k+1)}}{\partial \vec{\alpha}} \tag{2.56}$$

is also a sparse matrix with a similar pattern as the Jacobian matrix, which is also very easy to calculate.

In a conventional approach (i.e., pixel modeling), Eq. 2.53 is expressed as a recursive relationship for sensitivity coefficients

$$J^{(k+1)}S^{(k+1)} = -D^{(k+1)}S^{(k)} - Y^{(k+1)}, \tag{2.57}$$

where

$$S^{(k)} = \frac{\partial \vec{u}^{(k)}}{\partial \vec{\alpha}} \tag{2.58}$$

is the sensitivity coefficient matrix. Therefore we can use the recursive relation to calculate all the sensitivity coefficients necessary for the nonlinear regression.

In a more general approach, Eq. 2.53 becomes

$$J^{(k+1)} \frac{\partial \vec{u}^{(k+1)}}{\partial \vec{c}} = -D^{(k+1)} \frac{\partial \vec{u}^{(k+1)}}{\partial \vec{c}} - Y^{(k+1)} \cdot \frac{\partial \vec{\alpha}}{\partial \vec{c}}. \tag{2.59}$$

Eq. 2.59 can be rewritten as

$$J^{(k+1)} S_{\vec{c}}^{(k+1)} = -D^{(k+1)} S_{\vec{c}}^{(k)} - Y^{(k+1)} \cdot T, \tag{2.60}$$

where

$$S_{\vec{c}}^{(k)} = \frac{\partial \vec{u}^{(k)}}{\partial \vec{c}} \tag{2.61}$$

is the generalized sensitivity coefficient matrix and

$$T = \frac{\partial \vec{\alpha}}{\partial \vec{c}} \tag{2.62}$$

is the transform matrix, $\vec{c}$ is a set of parameters used for nonlinear regression. It is obvious that Eq. 2.60 will be reduced to Eq. 2.58 if the set of regression parameters is chosen the same as the reservoir parameters, because the transform matrix T becomes an identity matrix. We should note that T should not necessarily be constant. For example, in the case of a dynamic-object approach, the object can be expressed as

$$\vec{\alpha} = \vec{f}(\vec{c}), \tag{2.63}$$

where the relationship defined by $\vec{f}$ is not constant. Therefore

$$T = \frac{\partial \vec{\alpha}}{\partial \vec{c}} = \frac{\partial \vec{f}}{\partial \vec{c}} \tag{2.64}$$

varies as the object floats around inside the reservoir in different iterations.

2.4 Basic Concepts of Wavelet Transform

2.4.1 Overview and Background

The development of wavelets is fairly recent in applied mathematics, but wavelets have already had a remarkable impact. A lot of people are now applying wavelets to all aspects of the real world, and many are reporting favorable results. Broadly defined, a wavelet is simply a wavy function carefully constructed so as to have certain mathematical properties. A wavelet usually has such properties as good time-frequency localization, fast algorithms, simplicity of form, and the most important in our study, the capability of multiresolution analysis.

The good time-frequency localization is perhaps the most important advantage that makes wavelets so unique. Wavelets have a built-in "spatial adaptivity" that allows efficient estimation of functions with sharp spikes, and discontinuities in the function itself. Thus wavelet methods are useful in nonparametric regression for a very broad class of functions. Wavelets are intrinsically connected to the notion of "multiresolution analysis." That is, objects (signals, functions, data) can be examined using widely varying levels of focus. This is also known as the "zoom-in, zoom-out" property.

Because of these useful properties, wavelets have gained popularity in several practical fields. The most common application of wavelets is in signal processing (Mallat 1998). Image analysis is actually a special case of signal processing, one that deals with two-dimensional signals representing digital pictures. The general aim here is to remove as much of the noise as possible, but not at the expense of fine-scale details. The aim in data compression is to transform an enormous data set, saving only the most important elements of the transformed data, so that it can be reconstructed later with only a minimal loss of information or accuracy. Wavelets have proven extremely useful in solving such problems. This turns out quite useful in parameter estimation as well. In multiple linear regression, for example, it is desired to choose the simplest model that represents the data adequately, to achieve a parsimonious representation. With wavelets, a large data set can often be summarized well with only a relatively small number of wavelet coefficients.

2.4.2 Wavelet Transform

The wavelet transform decomposes signals (functions) over dilated and translated wavelets. A wavelet is a function $\psi \in \mathbf{L}^2(\mathcal{R})$ with a zero average:

$$\int_{-\infty}^{+\infty} \psi(t)\, dt = 0. \tag{2.65}$$

The function is normalized as $\|\psi\| = 1$, and centered in the neighborhood of $t = 0$. A family of time-frequency atoms is obtained by scaling ψ by s and translating it by u:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-u}{s}). \tag{2.66}$$

The wavelet transform of $f \in \mathbf{L}^2(\mathcal{R})$ at the time u and scale s is

$$Wf(u,\, s) = \langle f,\, \psi_{u,s}\rangle = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{s}}\psi^*(\frac{t-u}{s})\, dt. \tag{2.67}$$

The wavelet transform can be rewritten as a convolution product:

$$Wf(u,\, s) = \langle f,\, \psi_{u,s}\rangle = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{s}}\psi^*(\frac{t-u}{s})\, dt = f * \bar{\psi}_s(u) \tag{2.68}$$

with

$$\bar{\psi}_s(u) = \frac{1}{\sqrt{s}}\psi^*(\frac{-t}{s}). \tag{2.69}$$

By applying the Fourier-Parseval formula, the transform can also be written as a frequency integration:

$$Wf(u,\, s) = \langle f,\, \psi_{u,s}\rangle = \int_{-\infty}^{+\infty} f(t)\psi^*_{u,s}(t)\, dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(\omega)\hat{\psi}^*_{u,s}(\omega)\, d\omega. \tag{2.70}$$

In practice, real wavelets are often used to detect sharp signal transitions. In this study, we also focused on real wavelets. Therefore, throughout this study, real wavelets are assumed unless stated otherwise explicitly.

The wavelet transform can detect and characterize transients with a zooming procedure across scales. Because ψ has a zero average, a wavelet coefficient $Wf(u,\, s)$ means the variation of f in a neighborhood of u whose size is proportional to s. Sharp signal transitions create large amplitude wavelet coefficients. Singularities are detected by following across scales the local maxima of the wavelet transform. In

images, high-amplitude wavelet coefficients indicate the position of edges, which are sharp variations of the image intensity.

In parameter estimation, we need to describe a reservoir with a relatively small set of parameters, without losing vital information, such as major geological structures (faults, channels, etc.). So we can take advantage of wavelet's ability to reparameterize the reservoir and reduce the number of parameters to be estimated. This is described later in Chapter 3.

Chapter 3

Reservoir Modeling and Reparameterization

3.1 Reservoir Modeling and Numerical Simulation

3.1.1 Physical Laws and Assumptions

Oil and gas reservoirs lie deep underground and have very complex configuration and properties. In practice, we can only obtain data at a few widely spread locations with very different resolutions. One of the most challenging tasks in reservoir engineering is to describe the reservoir accurately and efficiently. An accurate description of a reservoir is crucial to the management of production and the efficiency of oil recovery. However, a reservoir is mathematically an infinite-dimensional system (the dimension here refers to the number of parameters, not the dimension in space), i.e., we need an infinite number of parameters to describe a reservoir fully, because both the formation and the fluid in the porous media are continuous and have very complicated properties. Usually, the formation is heterogeneous, compressible, while the fluid is multiphase, multicomponent, and with interaction between phases and components. The physical system is governed by several fundamental laws:

- *Mass conservation law*: this is the major driving force behind the reservoir mechanism. In any physical system, the total mass is conservative, i.e., the

mass gain (or loss) within any given volume is equivalent to mass source (or sink) minus the flux across the bounding surface.

- *Energy conservation law*: in any physical system, the energy is conservative. However, except in some special situations (e.g., geothermal reservoirs, thermal recovery, etc.), we will not consider the energy conservation law in oil and gas recovery. In most cases, we always assume the reservoir is isothermal, i.e., with a uniform temperature distribution.

- *Darcy's law*: this is the dynamic mechanism governing the flow through porous media. Although the Darcy's law is only an approximation with some limitations, it suits our purpose very well in this study. Darcy's law is expressed as

$$\vec{q_p} = -\frac{k_p}{\mu}\nabla(p_p + \rho_p g z). \tag{3.1}$$

- *Equation of state*: an algebraic expression that can represent the volumetric behavior and phase behavior of the fluid, both in and on the two-phase envelope. For a two-phase, oil-water system, the equation of state for each phase can be expressed as

$$p_p V_p = \gamma n_p R T. \tag{3.2}$$

- *Relative permeability and capillary pressure relationships*: these are dynamic relationships in multiphase flows. In this work, we consider only the two-phase flow. The relative permeability of phase p is $k_{rp} = k_{rp}(S_w)$, a function of the water saturation S_w. The capillary pressure is $p_c = p_c(S_w)$, also a function of S_w.

Combining these fundamental laws and the initial and boundary conditions, the physical system under study (the reservoir) can be represented as a mathematical model: a set of partial differential equations with suitable initial and boundary conditions. Usually, it is impossible to obtain an analytical/explicit solution to the partial differential equations, except for some very special cases, e.g., some well testing solutions. Therefore, it is necessary to use numerical methods to solve the problems

in this category. Numerical reservoir simulation is one of the most important applications in petroleum reservoir engineering. In order for a numerical simulator to describe and simulate (or approximate, more precisely speaking) a reservoir, we have to make some further assumptions and approximations.

First of all, the continuous system has to be described as a discrete system, with sufficient parameters to characterize the physical system as well as possible (or necessary). The most commonly applied discrete solution method, the finite-difference method, is a very powerful and efficient method in solving initial and boundary value problems. In this study, a fully-implicit finite-difference numerical simulator was applied, using Cartesian grids.

Secondly, the flow of fluid through porous media is multiphase and multicomponent, with interaction between different phases and components. In order to simplify the implementation and emphasize the major features of the study, we used only an oil-water model for this work. Therefore, there are only two phases (oil and water) in the system and each phase is treated as a single component. Furthermore, there is no dissolved water in the oil phase, and also no dissolved oil in the water phase.

3.1.2 Numerical Simulator

Based on all the assumptions and approximations stated above, the mathematical model for the physical system can be described and solved by a numerical simulator. In this study, we always refer to the numerical simulator as the mathematical (forward) model and will not make any distinction, although the numerical approximation is only accurate to a certain extent. More specifically, the numerical simulator is a three-dimensional, oil-water, black-oil, finite-difference simulator. Thus the inverse problem (parameter estimation problem) becomes a discrete inverse problem, and the general discrete inverse theory is fully applicable for our purpose, because the forward mathematical model is a discrete numerical simulator.

In this section, we will outline the setup of a two-phase, black-oil reservoir simulation. More details can be found in Aziz (1979) and Peaceman (1977). In the

numerical simulator, the primary variables are the oil pressure p_o and water saturation S_w at each gridblock, because there are two phases in the system. Accordingly, there are two mass conservation equations in each gridblock. The discrete differential equations for phase p at gridblock i and timestep $k+1$ are

$$f_{p,i}^{(k+1)} = \sum_{j \in \text{connection}} T_{p,ij}(\Phi_{p,j} - \Phi_{p,i}) - q_{p,i} - \frac{\Delta V_i}{\Delta t}\left(\left(\frac{\phi S_p}{B_p}\right)_i^{(k+1)} - \left(\frac{\phi S_p}{B_p}\right)_i^{(k)}\right) = 0, \tag{3.3}$$

where $p = o, w$ stands for oil phase and water phase respectively. The capillary pressure equation is

$$p_w = p_o + p_c(S_w), \tag{3.4}$$

where $p_c(S_w)$ is the capillary pressure between the two phases and is a function of water saturation. $\Phi_{p,i} = p_{p,i} + \rho_p g \Delta h_i$ is the phase potential, $q_{p,i}$ is the phase production rate, $T_{p,ij}$ is the transmissibility defined as

$$T_{p,ij} = \left(\frac{\bar{k}_{p,ij}\Delta A}{B_o \mu_o \Delta L}\right)_{i,j}, \tag{3.5}$$

where $\bar{k}_{p,ij}$ is the average phase permeability between two connected gridblocks, ΔA and ΔL are the cross-section area and distance between the two gridblocks. Depending on different implementations, the average phase permeability can be expressed as either a first-order or second-order upstream average. Along each wellbore, we have one more constraint for the production rate:

$$f_{well,i}^{(k+1)} = Q_{p,i} - \sum_{j \in \text{wellbore}} q_{p,j} = 0, \tag{3.6}$$

where $Q_{p,i}$ is the production rate of Well i, and $p = o, w, T$, because the production rate constraint can be either water, oil, or total flow rate. The wellblock production rate can be written as

$$q_{p,j} = WI_{p,j}(p_{p,j} - p_{well,j}),\ p = o, w, \tag{3.7}$$

$$q_{T,j} = q_{w,j} + q_{o,j}, \tag{3.8}$$

where the well index $WI_{p,j}$ is given by Peaceman (1977) as

$$WI_{p,j} = \left(\frac{2\pi k k_{rp}\Delta z}{B_p \mu_p \ln(\frac{r_0}{r_w} + s)}\right)_j. \tag{3.9}$$

This equation only applies for isotropic formations, i.e., $k_x = k_y = k_z$.

Generally, for a reservoir with $nBlock$ gridblocks and $nWell$ wells, there are totally $2nBlock + nWell$ variables ($nBlock$ oil pressures, $nBlock$ water saturations, and $nWell$ downhole pressures) and $2nBlock + nWell$ equations. Therefore, the discrete system is expressed as

$$\vec{f}_o(\vec{p}_o^{(k+1)}, \vec{S}_w^{(k+1)}, \vec{p}_{well}^{(k+1)}, \vec{p}_o^{(k)}, \vec{S}_w^{(k)}, \vec{p}_{well}^{(k)}, \vec{k}, \vec{\phi}^0, \Delta t) = \vec{0}, \tag{3.10}$$

$$\vec{f}_w(\vec{p}_o^{(k+1)}, \vec{S}_w^{(k+1)}, \vec{p}_{well}^{(k+1)}, \vec{p}_o^{(k)}, \vec{S}_w^{(k)}, \vec{p}_{well}^{(k)}, \vec{k}, \vec{\phi}^0, \Delta t) = \vec{0}, \tag{3.11}$$

$$\vec{f}_{well}(\vec{p}_o^{(k+1)}, \vec{S}_w^{(k+1)}, \vec{p}_{well}^{(k+1)}, \vec{p}_o^{(k)}, \vec{S}_w^{(k)}, \vec{p}_{well}^{(k)}, \vec{k}, \vec{\phi}^0, \Delta t) = \vec{0}, \tag{3.12}$$

where $\vec{\phi}^0$ is the porosity at the initial reservoir conditions, and $\phi_i = \phi_i(\phi_i^0, p_{w,i})$. A more general form for the discrete system is

$$\vec{f}(\vec{u}^{(k+1)}, \vec{u}^{(k)}, \vec{v}, \Delta t; \vec{\alpha}) = \vec{0}, \tag{3.13}$$

where

$$\vec{f} = [f_{o,1}, f_{w,1}, ..., f_{o,nBlock}, f_{w,nBlock}; f_{well,1}, ..., f_{well,nWell}]^T, \tag{3.14}$$

$$\vec{u} = [p_{o,1}, S_{w,1}, ..., p_{o,nBlock}, S_{w,nBlock}; p_{well,1}, ..., p_{well,nWell}]^T, \tag{3.15}$$

and

$$\vec{\alpha} = \begin{bmatrix} \vec{k} \\ \vec{\phi}^0 \end{bmatrix}, \tag{3.16}$$

$\vec{v}$ is a vector of other (known) reservoir properties.

In order to make the numerical solution stable and consistent, we usually solve the discrete system Eq. 3.13 in a fully implicit form. Thus Eq. 3.13 is a nonlinear system and can not be solved directly. One of the most frequently-used methods for solving nonlinear equations is the Newton-Raphson method, an iterative process. Instead of solving the nonlinear system, a series of linear systems are solved iteratively, until the system reaches convergence. At each iteration, the linear system to solve is

$$J_\nu^{(k+1)} \cdot \delta\vec{u}_\nu^{(k+1)} = -\vec{f}_\nu^{(k+1)}, \tag{3.17}$$

with

$$u_{\nu+1}^{(k+1)} = u_\nu^{(k+1)} + \delta u_\nu^{(k+1)}, \tag{3.18}$$

$$f_\nu^{(k+1)} = \vec{f}(\vec{u}_\nu^{(k+1)}, \vec{u}^{(k)}, \vec{v}, \Delta t; \vec{\alpha}). \tag{3.19}$$

The Jacobian matrix is

$$J_\nu^{(k+1)} = \frac{\partial \vec{f}_\nu^{(k+1)}}{\partial \vec{u}_\nu^{(k+1)}}. \tag{3.20}$$

The iterative process stops when $\left\|\delta\vec{u}_\nu^{(k+1)}\right\| < \epsilon_u$ and $\left\|\vec{f}_\nu^{(k+1)}\right\| < \epsilon_f$.

The structure of this iteration process and the form of the Jacobian is important to the construction of the sensitivity coefficient matrix, as is described in the next section.

3.2 Calculation of Sensitivity Coefficients

Gradient-based parameter estimation algorithms rely heavily on the calculation of sensitivity coefficients. By definition, sensitivity coefficients are the partial derivatives of reservoir response variables with respect to reservoir parameters: $S = \partial\vec{u}/\partial\vec{\alpha}$. However, the discrete system described by Eq. 3.13 is nonlinear and does not have an analytical solution, therefore we can only obtain a numerical approximation of the partial derivatives. In practice, the first-order approximation has sufficient precision and was shown to be adequate by many numerical experiments. In this study, we also use the first-order approximation of partial derivatives, i.e., the finite difference:

$$\frac{\partial\vec{u}(\vec{\alpha})}{\partial\vec{\alpha}} = \frac{\vec{u}(\vec{\alpha} + \Delta\vec{\alpha}) - \vec{u}(\vec{\alpha})}{\Delta\vec{\alpha}}. \tag{3.21}$$

3.2.1 Substitution Method

There are several different approaches to calculate the finite differences. The most straightforward algorithm is the substitution method. The algorithm is as follows:

1. Solve Eq. 3.13 (a single simulation run) with $\vec{\alpha} = \vec{\alpha}_0$ to obtain $\vec{u}_0 = \vec{u}(\vec{\alpha}_0)$.

2. For each parameter α_i, $i = 1, 2, ..., nPar$, solve Eq. 3.13 with $\alpha_i = \alpha_{0,i} + \delta\alpha_i$ to obtain $\vec{u}_i^*$, where $\delta\alpha_i$ is a small perturbation.

3. The sensitivity is calculated by

$$\left.\frac{\partial \vec{u}}{\partial \alpha_i}\right|_{\vec{\alpha}=\vec{\alpha}_0} = \frac{\vec{u}_i^* - \vec{u}_0}{\delta \alpha_i}. \tag{3.22}$$

This method is very straightforward and does not require any modification to the numerical simulator. The substitution method is feasible when the calculation involves only a small number of parameters. However, this approach has a very serious drawback: it is computationally inefficient. The total calculation for $nPar$ parameters requires $1 + nPar$ simulation runs. In situations where many parameters are involved, the substitution method is no longer feasible.

3.2.2 Modified Generalized Pulse-Spectrum Method

As described briefly in Chapter 2, the modified generalized pulse-spectrum method is much more efficient than the substitution method. The algorithm is expressed as:

$$J^{(k+1)} S^{(k+1)} = -D^{(k+1)} S^{(k)} - Y^{(k+1)}, \tag{3.23}$$

with the sensitivity coefficient matrix $S^{(k)} = \partial \vec{u}^{(k)} / \partial \vec{\alpha}$. In Eq. 3.23, $J^{(k+1)}$ is the last Jacobian matrix computed in the Newton-Raphson method in the numerical simulator. $D^{(k+1)} = \partial \vec{f}^{(k+1)} / \partial \vec{u}^{(k)}$ is a block-diagonal matrix, where

$$D_{2i-1,2i-1}^{(k+1)} = \frac{\Delta V_i}{\Delta t} \left(\frac{\partial (\phi S_o / B_o)}{\partial p_o} \right)_i^{(k)}, \tag{3.24}$$

$$D_{2i-1,2i}^{(k+1)} = \frac{\Delta V_i}{\Delta t} \left(\frac{\partial (\phi S_o / B_o)}{\partial S_w} \right)_i^{(k)}, \tag{3.25}$$

$$D_{2i,2i-1}^{(k+1)} = \frac{\Delta V_i}{\Delta t} \left(\frac{\partial (\phi S_w / B_w)}{\partial p_o} \right)_i^{(k)}, \tag{3.26}$$

$$D_{2i,2i}^{(k+1)} = \frac{\Delta V_i}{\Delta t} \left(\frac{\partial (\phi S_w / B_w)}{\partial S_w} \right)_i^{(k)}, \tag{3.27}$$

for $i = 1, 2, ..., nBlock$. Both the Jacobian matrix $J^{(k)}$and the block-diagonal matrix $D^{(k)}$ are constant for all the parameters, therefore we only need to calculate them once

at each timestep and there is no computational overhead at all. The other matrix $Y^{(k+1)}$ is a little more complicated:

$$Y^{(k+1)} = \frac{\partial \vec{f}^{(k+1)}}{\partial \vec{\alpha}} = \begin{bmatrix} \partial \vec{f}^{(k+1)}/\partial \vec{k} \\ \partial \vec{f}^{(k+1)}/\partial \vec{\phi}^0 \end{bmatrix}. \tag{3.28}$$

$\vec{Y}^{(k+1)}_{\phi_i^0} = \partial \vec{f}^{(k+1)}/\partial \phi_i^0$ is calculated as follows:

$$\vec{Y}^{(k+1)}_{\phi_i^0} = \left[0 \cdots \frac{\partial f_{o,i}}{\partial \phi_i^0} \; \frac{\partial f_{w,i}}{\partial \phi_i^0} \cdots 0 \right]^T, \tag{3.29}$$

with

$$\frac{\partial f_{o,i}}{\partial \phi_i^0} = \frac{\Delta V_i}{\Delta t} \left(\left(\frac{S_o}{B_o} \frac{\partial \phi}{\partial \phi^0} \right)_i^{(k+1)} - \left(\frac{S_o}{B_o} \frac{\partial \phi}{\partial \phi^0} \right)_i^{(k)} \right), \tag{3.30}$$

$$\frac{\partial f_{w,i}}{\partial \phi_i^0} = \frac{\Delta V_i}{\Delta t} \left(\left(\frac{S_w}{B_w} \frac{\partial \phi}{\partial \phi^0} \right)_i^{(k+1)} - \left(\frac{S_w}{B_w} \frac{\partial \phi}{\partial \phi^0} \right)_i^{(k)} \right). \tag{3.31}$$

$\partial \vec{f}^{(k+1)}/\partial \vec{k}$ has a similar sparse structure to the Jacobian matrix, the details can be found in Landa (1997) and Phan (1998).

In practice, the permeability distribution varies over a very wide range, usually from 0.01 md to over 1000 md. It is advantageous to use the logarithm of the permeability ($\ln k$) as the model parameter, to remove extreme values. The sensitivity of variables with respect to $\ln k$ is easy to calculate as:

$$\frac{\partial u}{\partial \ln k} = \frac{\partial u}{\partial k} \frac{\partial k}{\partial \ln k} = k \frac{\partial u}{\partial k}. \tag{3.32}$$

However, sometimes we need to use object modeling (as described in Section 3.3.2) to parameterize the system, in which case the model parameters are functions of k or $\ln k$. It would be much more efficient to calculate the sensitivities directly inside the modified GPST algorithm. We only need to make a minor modification to the modified GPST method to fulfill this purpose. The model parameter becomes

$$\vec{\alpha} = \begin{bmatrix} \ln \vec{k} \\ \vec{\phi}^0 \end{bmatrix}, \tag{3.33}$$

while Eq. 3.28 becomes

$$Y^{(k+1)} = \frac{\partial \vec{f}^{(k+1)}}{\partial \vec{\alpha}} = \begin{bmatrix} \partial \vec{f}^{(k+1)}/\partial \ln \vec{k} \\ \partial \vec{f}^{(k+1)}/\partial \vec{\phi}^0 \end{bmatrix}. \tag{3.34}$$

The $\partial \vec{f}^{(k+1)}/\partial \ln \vec{k}$ can be calculated easily by

$$\frac{\partial f}{\partial \ln k} = k \frac{\partial f}{\partial k}. \tag{3.35}$$

3.2.3 Sensitivity of Dynamic Data

As required by the parameter estimation algorithm (in our case, the Gauss-Newton algorithm), we need to calculate the sensitivity of the observed dynamic data (reservoir responses) with respect to the model parameters (the permeability and porosity field, for the pixel approach). All these sensitivities can be calculated easily using Eq. 3.23. In this study, the dynamic data were the bottomhole pressure at each well, watercut in each well, and water saturation change in the reservoir during a given time interval (from 4-D seismic information).

The sensitivity of the bottomhole pressure with respect to the parameters is already calculated in Eq. 3.23, because $u^{(k)}_{2nBlock+i} = p^{(k)}_{well,i}$. Thus

$$\frac{\partial p^{(k)}_{well,i}}{\partial \vec{\alpha}} = \frac{\partial u^{(k)}_{2nBlock+i}}{\partial \vec{\alpha}} = S^{(k)}_{2nBlock+i}. \tag{3.36}$$

At each well, the watercut at standard conditions is defined as:

$$wc = \frac{Q_w}{Q_T}, \tag{3.37}$$

where $Q_T = Q_w + Q_o$ is the total production rate. To simplify the problem, we only considered two different well production constraints: fixed total production rate and fixed oil production rate. For a well with a fixed total production rate ($\partial Q_T/\partial \alpha = 0$), the sensitivity of watercut is

$$\frac{\partial wc}{\partial \alpha} = \frac{1}{Q_T} \frac{\partial Q_w}{\partial \alpha}. \tag{3.38}$$

On the other hand, for a well with a fixed oil production rate ($\partial Q_o/\partial \alpha = 0$), the sensitivity of watercut is

$$\frac{\partial wc}{\partial \alpha} = \frac{Q_o}{Q_T^2} \frac{\partial Q_w}{\partial \alpha} = (1 - wc) \frac{1}{Q_T} \frac{\partial Q_w}{\partial \alpha}. \tag{3.39}$$

Obviously, we only need $\partial Q_w/\partial\alpha$ to calculate the sensitivities of watercuts. Because $Q_w = \sum_{j\in\text{well}} q_{w,j}$, we have

$$\frac{\partial Q_w}{\partial\alpha} = \sum_{j\in\text{well}} \frac{\partial q_{w,j}}{\partial\alpha}, \tag{3.40}$$

with

$$q_{w,j} = WI_{w,j}(p_{w,j} - p_{well,j}), \tag{3.41}$$

and

$$WI_{w,j} = \left(\frac{2\pi k\lambda_w \Delta z}{B_w \ln(r_0/r_w + s)}\right)_j. \tag{3.42}$$

In general, we have

$$\frac{\partial q_{w,j}}{\partial\alpha} = \frac{\partial WI_{w,j}}{\partial\alpha}(p_{w,j} - p_{well,j}) - WI_{w,j}(\frac{\partial p_{w,j}}{\partial\alpha} - \frac{\partial p_{well,j}}{\partial\alpha}). \tag{3.43}$$

In Eq. 3.43, both $\partial p_{w,j}/\partial\alpha$ and $\partial p_{well,j}/\partial\alpha$ are already calculated, therefore, the only missing term is $\partial WI_{w,j}/\partial\alpha$. The calculation is as follows:

$$\frac{\partial WI_{w,j}}{\partial k_i} = a_j\left(k_j b_j \frac{\partial p_{o,j}}{\partial k_i} + k_j c_j \frac{\partial S_{w,j}}{\partial k_i}\right) + \frac{WI_{w,j}}{k_j}\delta_{ij}, \tag{3.44}$$

$$\frac{\partial WI_{w,j}}{\partial \ln k_i} = a_j\left(k_j b_j \frac{\partial p_{o,j}}{\partial \ln k_i} + k_j c_j \frac{\partial S_{w,j}}{\partial \ln k_i}\right) + WI_{w,j}\delta_{ij}, \tag{3.45}$$

$$\frac{\partial WI_{w,j}}{\partial \phi_i^0} = a_j\left(k_j b_j \frac{\partial p_{o,j}}{\partial \phi_i^0} + k_j c_j \frac{\partial S_{w,j}}{\partial \phi_i^0}\right), \tag{3.46}$$

with

$$a_j = \left(\frac{2\pi\Delta z}{\ln(r_0/r_w + s)}\right)_j, \tag{3.47}$$

$$b_j = \frac{\partial}{\partial p_{o.j}}\left(\frac{\lambda_{w,j}}{B_{w,j}}\right), \tag{3.48}$$

$$c_j = \frac{\partial}{\partial S_{w.j}}\left(\frac{\lambda_{w,j}}{B_{w,j}}\right). \tag{3.49}$$

The water saturation change ΔS_w at the i^{th} gridblock during the time interval $[t_1, t_2]$ is defined as:

$$\Delta S_{w,i} = S_{w,i}^{t_2} - S_{w,i}^{t_1}. \tag{3.50}$$

The sensitivity is computed as follows:

$$\frac{\partial \Delta S_{w,i}}{\partial k_j} = \frac{\partial S_{w,i}^{t_2}}{\partial k_j} - \frac{\partial S_{w,i}^{t_1}}{\partial k_j}, \tag{3.51}$$

$$\frac{\partial \Delta S_{w,i}}{\partial \ln k_j} = \frac{\partial S_{w,i}^{t_2}}{\partial \ln k_j} - \frac{\partial S_{w,i}^{t_1}}{\partial \ln k_j}, \tag{3.52}$$

$$\frac{\partial \Delta S_{w,i}}{\partial \phi_j^0} = \frac{\partial S_{w,i}^{t_2}}{\partial \phi_j^0} - \frac{\partial S_{w,i}^{t_1}}{\partial \phi_j^0}. \tag{3.53}$$

3.2.4 Validation of the Modified GPST Method

The accuracy of the modified GPST method has been verified in several published papers (e.g. Chu & Komara 1996, Landa, Kamal, Jenkins & Horne 1996). In this section, we will show a simple numerical experiment to demonstrate the application of this method.

The test reservoir is a two-dimensional, rectangular formation with two-phase flow (water/oil), as depicted in Fig. 3.1. There are four wells located one at each corner, among which Well #1 is the injecting well, the others are producing. The water injection rate is fixed at 10050 STB/D, and the total production rates (Q_w+Q_o) at the producing wells are fixed at 3000 STB/D, 3000 STB/D and 4000 STB/D respectively. The permeability distribution is as depicted in Fig. 3.1: a high-permeability strip with $k = 2500$ md is located in the middle of the reservoir, while the rest of the reservoir has $k = 500$ md. The initial porosity is uniform at $\phi^0 = 0.22$. The reservoir is discretized with a 32×32 grid (totally 1024 gridblocks), while each gridblock is $50 \times 50 \times 100$ ft^3. In order to verify the accuracy of the sensitivity calculation, the permeability and porosity at gridblock #101 (k_{101} and ϕ_{101}^0 respectively) were perturbed and the sensitivity of dynamic responses were calculated with the substitution method. Fig. 3.2 shows the water saturation change in the period between $t = 100$ days and $t = 200$ days. Fig. 3.3 shows the water injection pressure in Well #1, while Fig. 3.4 shows the bottomhole pressure and watercut in Well #2. Figs. 3.5 through 3.10 show the sensitivities of dynamic data calculated by both the substitution method and the modified GPST method. The two computations gave almost identical results for this particular case. If we take the substitution method as the benchmark, the modified

GPST method gives very good results.

However, if the sensitivity is very small, both of the methods will be less reliable, especially in the case of the substitution method.

In the substitution method, the sensitivity is approximated as a finite difference

$$\frac{\partial u(\alpha)}{\partial \alpha} = \frac{u(\alpha + \Delta\alpha) - u(\alpha)}{\Delta\alpha} + O(\Delta\alpha), \tag{3.54}$$

where the perturbation $\Delta\alpha$ has to been very small to make the finite difference approximation accurate. However, the calculation of the response u has a numerical error ϵ, which is independent of $\Delta\alpha$. Therefore, the numerical calculation is

$$\frac{\partial u(\alpha)}{\partial \alpha} = \frac{u(\alpha + \Delta\alpha) - u(\alpha)}{\Delta\alpha} + O(\Delta\alpha) + \frac{\epsilon}{\Delta\alpha}. \tag{3.55}$$

In the situation that the sensitivity is very small, the numerical error $\epsilon/\Delta\alpha$ will dominate the calculation.

The modified GPST method has a similar problem for the calculation of small sensitivities, though with a different mechanism. Eq. 3.23 introduces a numerical error while computing the Y term, of the order of $\Delta\alpha$ (if a finite-difference method is used). The total computational error is $O(\Delta\alpha) + \epsilon$, not $O(\Delta\alpha) + \epsilon/\Delta\alpha$ as in the substitution method. Therefore, the relative error of the modified GPST method is much smaller than that of the substitution method, in the case in which both the sensitivity and $\Delta\alpha$ are very small.

In summary, the modified GPST method is accurate enough for practical purposes. though resulting from the numerical errors, the calculation of small sensitivity is not very reliable and there is always some difference between the two different calculations. As shown here, the substitution method may not be a good benchmark to validate the accuracy of other methods, because its accuracy will be degraded severely in certain situations.

3.3 Parameterization of the System

As described in Section 2.1, an important step in the study of a physical system is to parameterize, i.e., to determine a minimum set of model parameters whose values

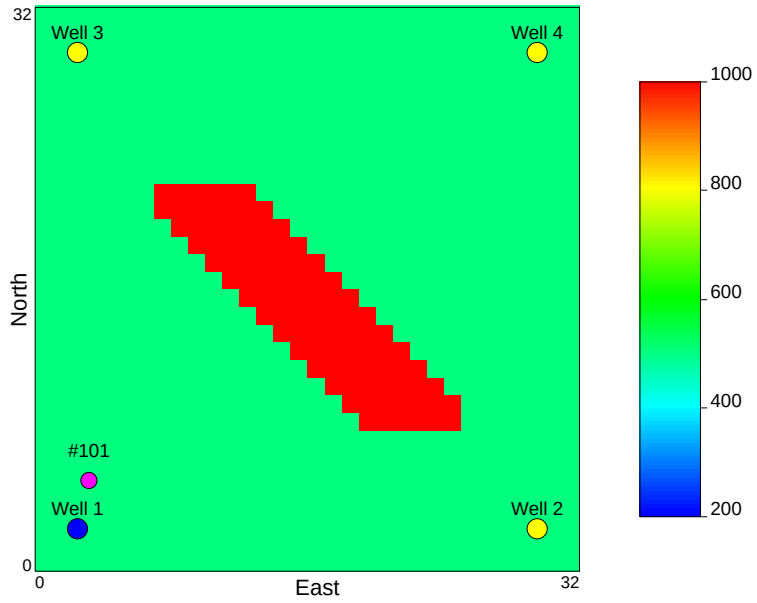

Figure 3.1: Reservoir permeability distribution

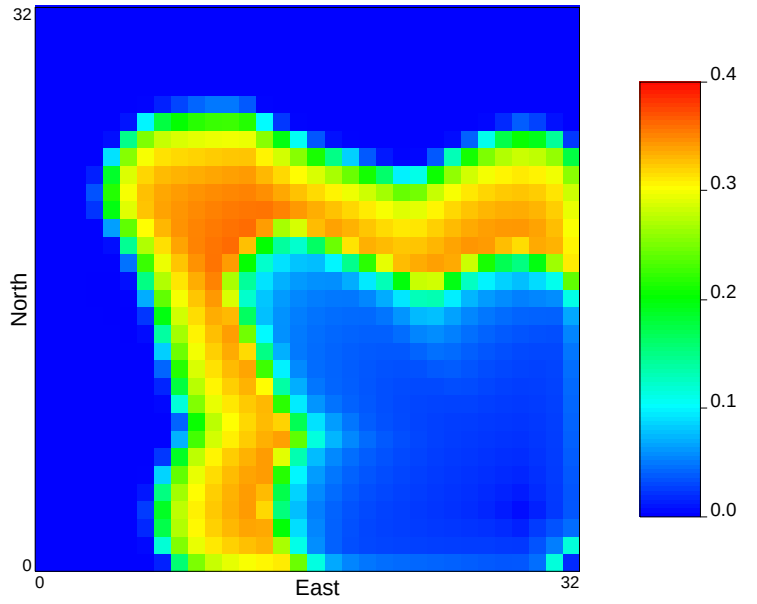

Figure 3.2: Water saturation change distribution during time t=100,200 days

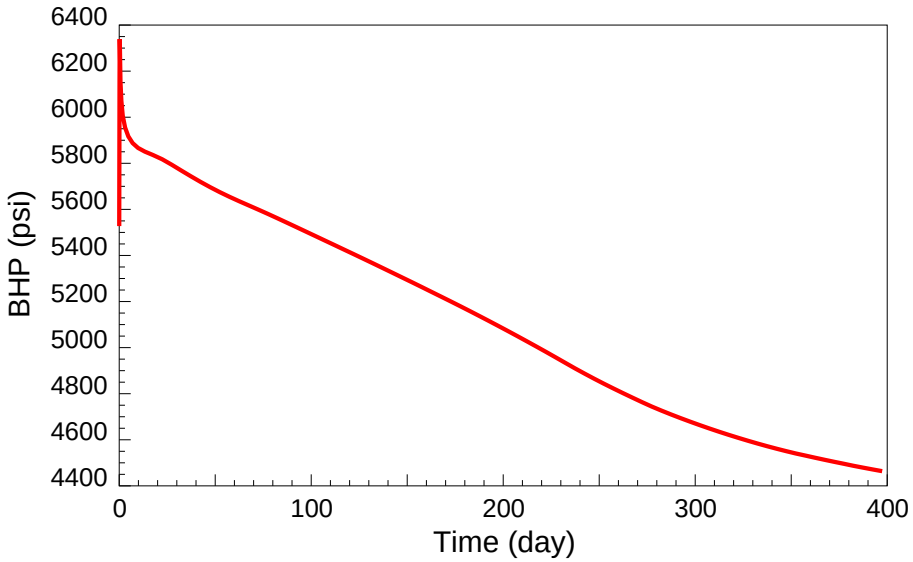

Figure 3.3: Bottomhole pressure at the injection well (Well #1)

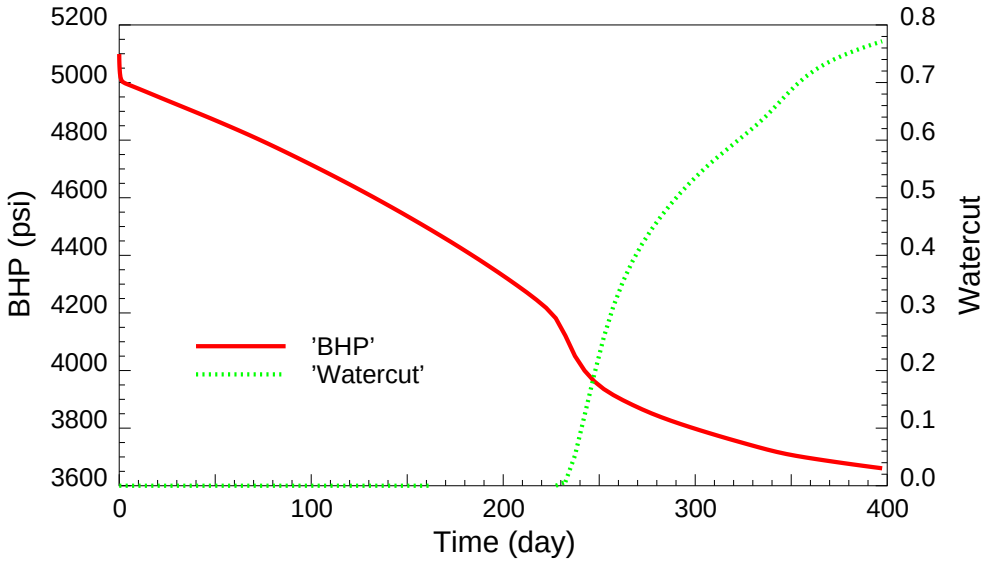

Figure 3.4: Bottomhole pressure and watercut at Well #2

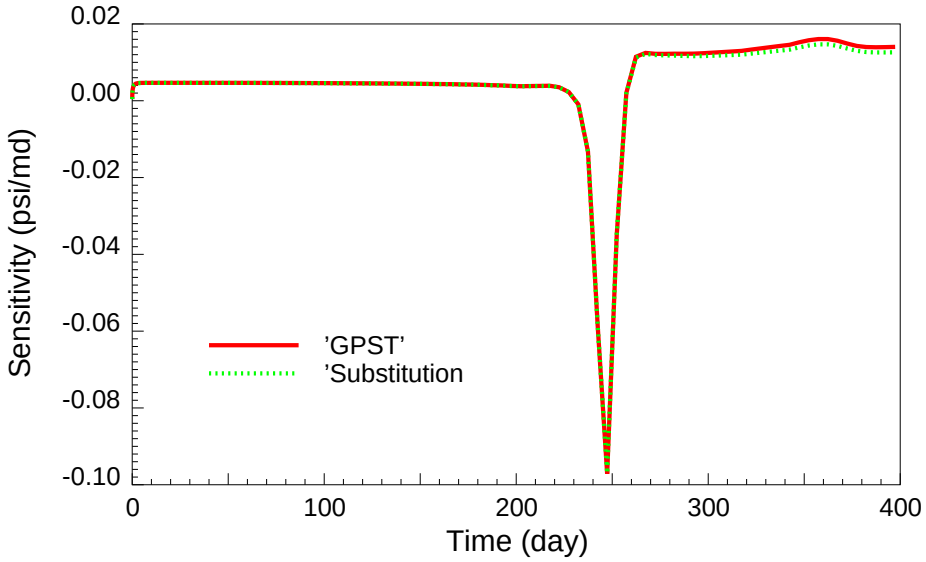

Figure 3.5: Sensitivity of bottomhole pressure in Well #2 with respect to k_{101}

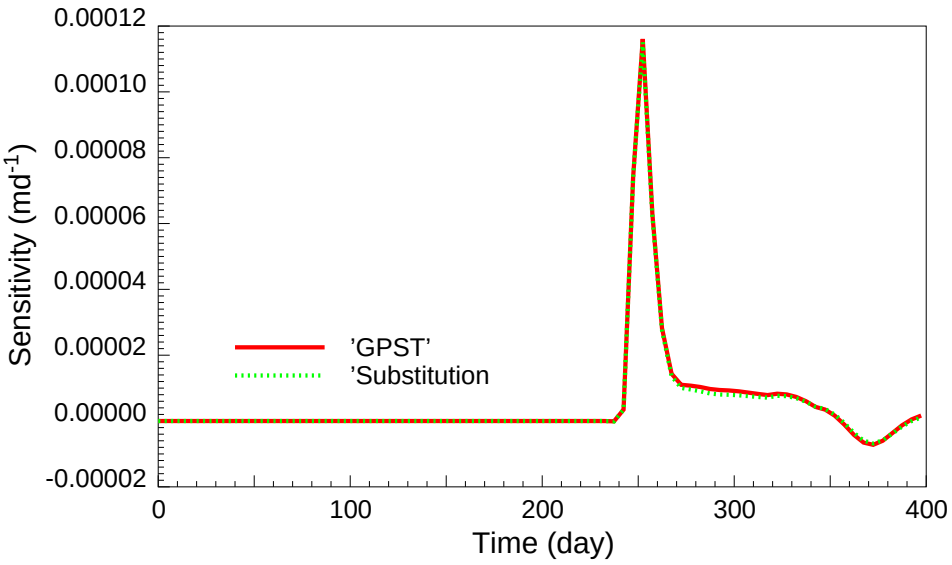

Figure 3.6: Sensitivity of watercut in Well #2 with respect to k_{101}

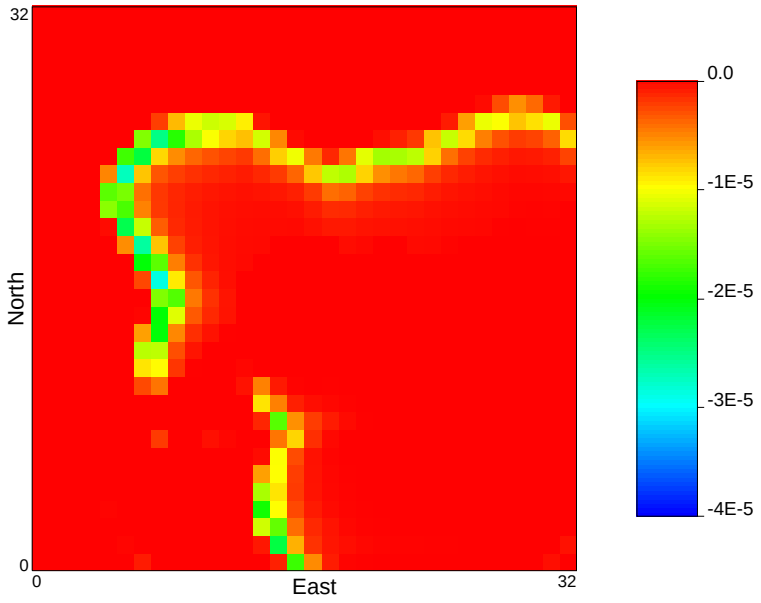

Figure 3.7: Sensitivity of water saturation change with respect to the permeability in gridblock #101, as calculated by the modified GPST.

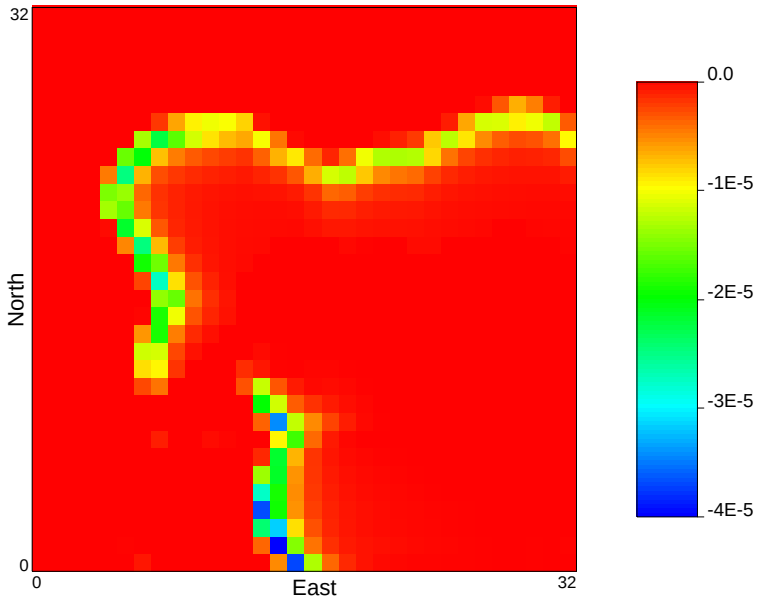

Figure 3.8: Sensitivity of water saturation change with respect to the permeability in gridblock #101, as calculated by the substitution method.

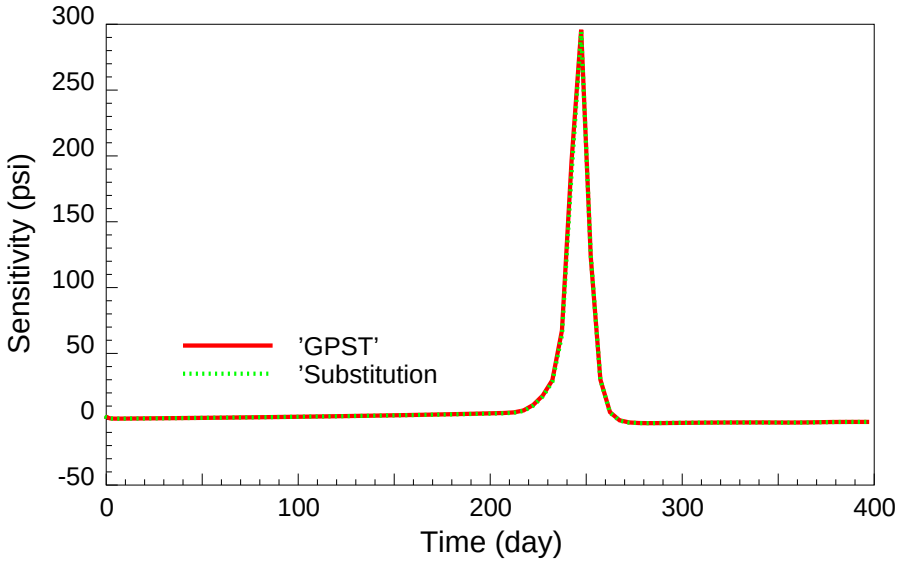

Figure 3.9: Sensitivity of bottomhole pressure in Well #2 with respect to ϕ^0_{101}

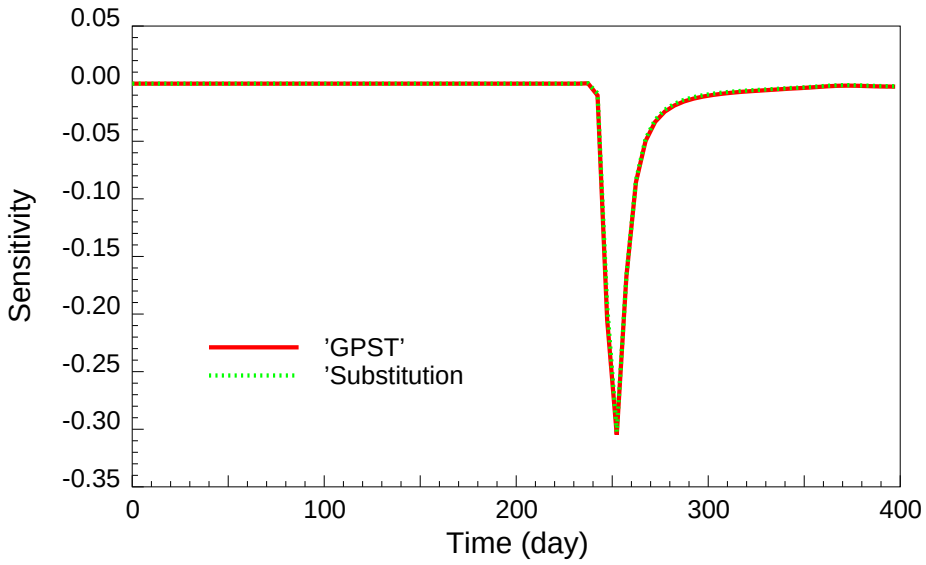

Figure 3.10: Sensitivity of watercut in Well #2 with respect to ϕ^0_{101}

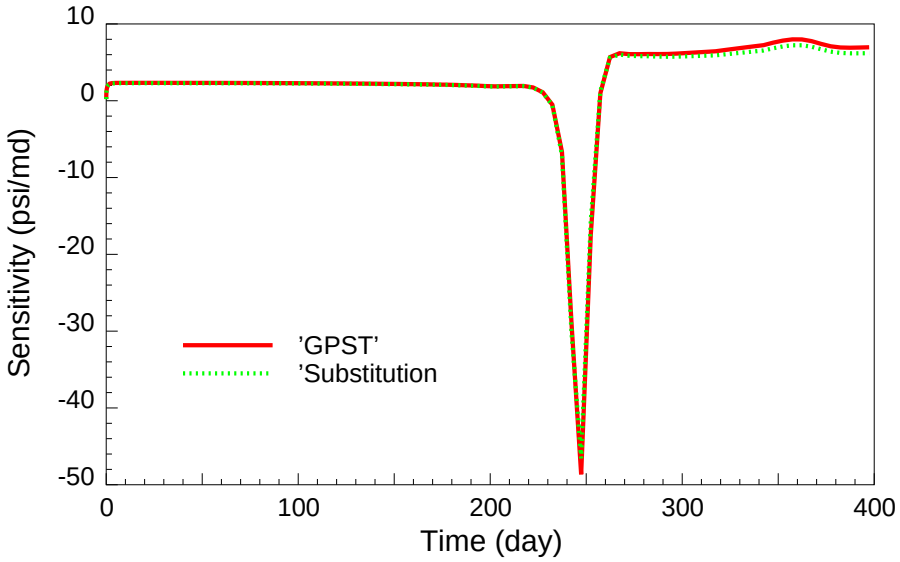

Figure 3.11: Sensitivity of bottomhole pressure in Well #2 with respect to $\ln k_{101}$

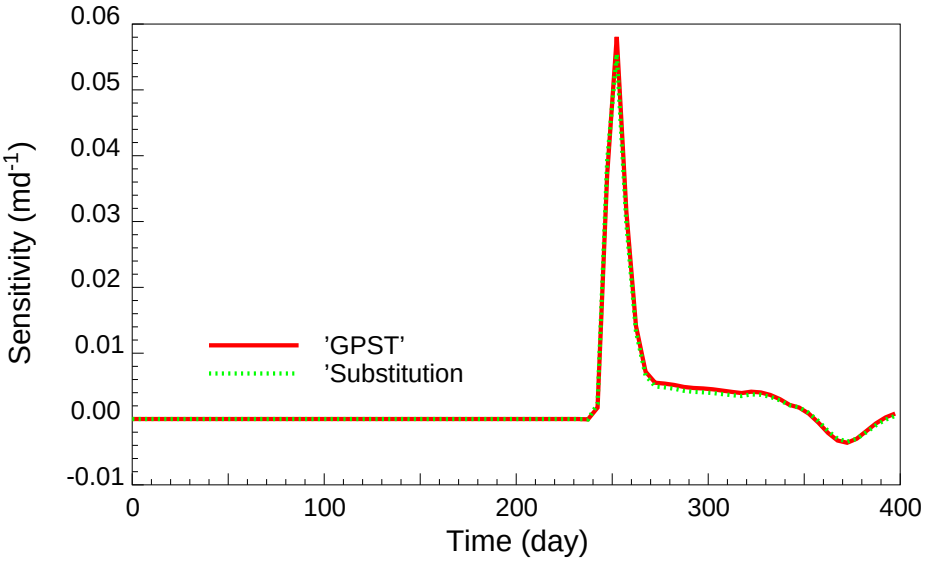

Figure 3.12: Sensitivity of watercut in Well #2 with respect to $\ln k_{101}$

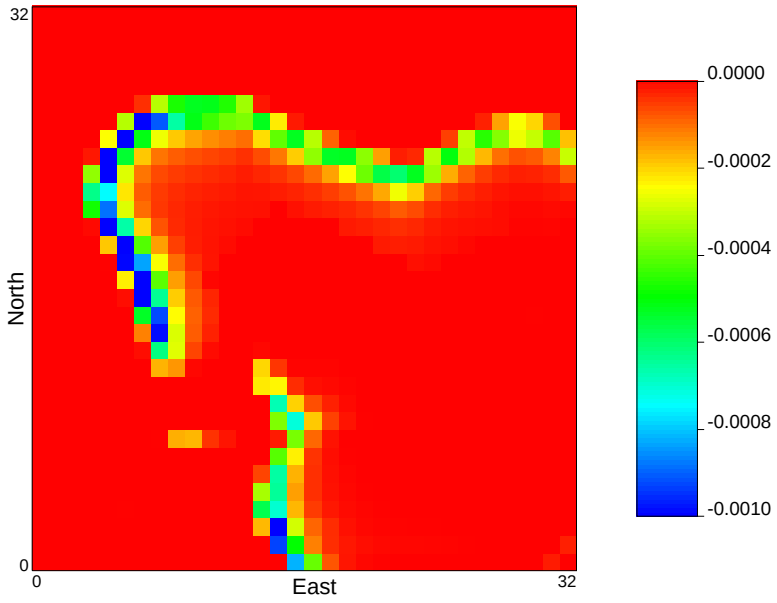

Figure 3.13: Sensitivity of water saturation change with respect to $\ln k_{101}$, as calculated by the modified GPST

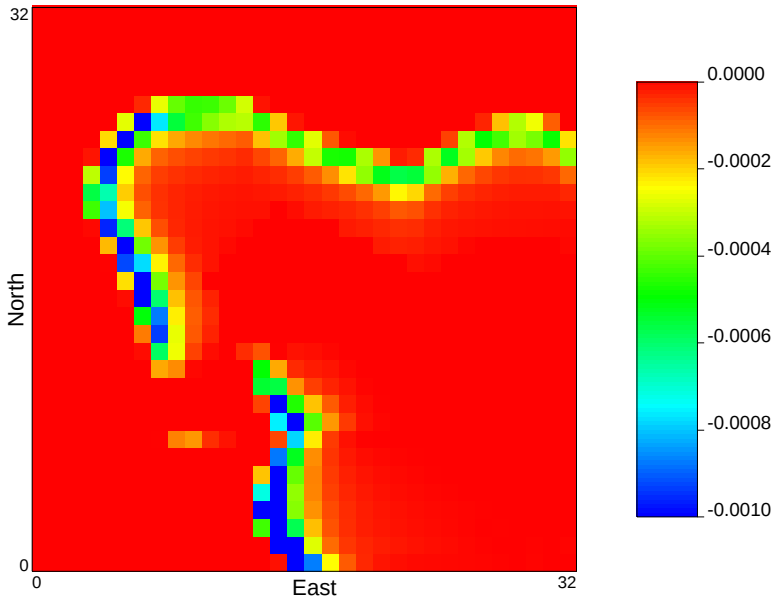

Figure 3.14: Sensitivity of water saturation change with respect to $\ln k_{101}$, as calculated by the substitution method

completely characterize the system. In this study, the physical system under investigation is a reservoir. The reservoir is represented as a discrete three-dimensional system with N gridblocks, in which each block is fully characterized by its permeability (k) and porosity (ϕ), the variables that have the largest influence in determining the performance of the reservoir. In a conventional approach (also referred to as pixel approach), the permeabilities and/or porosities of all blocks are the model parameters. A particular choice of model parameters is a *parameterization* of the system.

The choice of the model parameters to describe a system is generally not unique. For example, in order to describe a reservoir, we need to know the permeability and porosity of each gridblock. However, if there is a relationship between the permeability and the porosity, we can use either the permeability or the porosity as the model parameter. In this case, the two different parameterizations are equivalent because they are correlated.

This work explored several different approaches to define the model parameters, i.e., to determine the suitable parameterization. All the approaches have the same objective, to estimate the spatial distribution of reservoir properties (in this work, permeability and porosity).

3.3.1 Pixel Modeling

To describe the reservoir system, we need N parameters (if k and ϕ are correlated) or $2N$ parameters (if k and ϕ are uncorrelated) for a complete characterization. We need to infer these parameters from the observations such as production data, downhole pressure, etc.. There are many studies that have been done in this way. This approach (referred to as *pixel modeling*) has many advantages. Because the permeability and the porosity are used as the unknown parameters, we can obtain a direct description of the reservoir. This approach also computes the distributions of permeability and porosity at the smallest level of the simulation grid. Another advantage is that it is very easy for us to apply gradient-based methods in the parameter estimation problem, because all the sensitivity coefficients can be obtained easily using the modified GPST algorithm described in Section 3.2.

The problem with this approach is that it requires the computation of sensitivity coefficients for a large number of parameters, which in general is of the order of thousands (in a real simulation, the number of gridblocks may be up to the order of millions). The inverse problem is highly under-determined because the data are insufficient to resolve such a large number of parameters. As a result, the application of this approach may result in distributions of permeability and porosity that match the dynamic data but cannot be accepted from a geological point of view (Landa, Kamal, Jenkins & Horne 1996).

3.3.2 Object Modeling

In order to overcome the disadvantages of the pixel modeling, another approach has been developed as a way to reduce the number of parameters and also to preserve the large-scale information. In this approach (referred to as *object modeling*), the distribution of permeability and porosity can be described by a geometric or geostatistical model with a relatively small number of parameters (Landa 1997). Therefore, the model parameters are no longer the permeability and porosity. Instead, the model parameters are a set of parameters to define the geometry of the permeability and porosity distributions. Furthermore, because the number of parameters is much smaller than that in pixel modeling, the inverse problem is better determined. There is another important advantage in that the object modeling can preserve the large-scale geological information very well. In object modeling, the permeability and porosity distributions are expressed as:

$$\vec{k} = \vec{k}(\vec{\alpha}), \tag{3.56}$$

$$\vec{\phi}^0 = \vec{\phi}^0(\vec{\alpha}). \tag{3.57}$$

In practice, it is always desirable to choose objects such that the size of parameter set $\vec{\alpha}$ is much smaller than that of the original pixel parameter set. As an example, one of the simplest objects is a sinuous channel object describing a permeability distribution, which is defined as follows:

$$y_1 = \alpha_1 \sin(\alpha_2 x + \alpha_3) + \alpha_4, \tag{3.58}$$

$$y_2 = \alpha_1 \sin(\alpha_2 x + \alpha_3) + \alpha_5, \tag{3.59}$$

$$k(x,y) = \begin{cases} \alpha_6 & : \text{ if } y_2 < y < y_1 \\ \alpha_7 & : \text{ otherwise} \end{cases} \tag{3.60}$$

Each of the parameters has the following physical meaning:

- α_1 is the "amplitude" of the sinuous channel.
- α_2 is the "sinuosity" of the sinuous channel.
- α_3 is the displacement of the channel in the x direction.
- α_4 is the displacement of the higher channel bound in the y direction.
- α_5 is the displacement of the lower channel bound in the y direction.
- α_6 is the permeability inside the channel.
- α_7 is the permeability outside the channel.

Obviously, $\alpha_4 - \alpha_5$ is the "width" and $(\alpha_4 + \alpha_5)/2$ is the center of the channel in the y direction. The channel object can be rewritten as

$$k(x,y) = k(x,y,\vec{\alpha}) \tag{3.61}$$

with $\vec{\alpha} = [\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7]^T$. As we can see, there are only seven channel parameters to describe the spatial permeability distribution fully.

Just as in pixel modeling, we need to calculate the sensitivity of the pressure and watercut in the wells and the change of water saturation distribution in the reservoir with respect to the parameter set $\vec{\alpha}$. The sensitivity matrix is defined as:

$$S_{\vec{\alpha}} = \frac{\partial \vec{u}}{\partial \vec{\alpha}}. \tag{3.62}$$

The sensitivity of response variables with respect to a specific parameter α_i is a vector defined as:

$$S_{\alpha_i} = \frac{\partial \vec{u}}{\partial \alpha_i}. \tag{3.63}$$

The easiest method to calculate the sensitivity would be to use the chain rule. The sensitivity is calculated as the total differential with the partial derivatives and the basic sensitivities calculated with Eq. 3.23. The formulation is

$$\vec{S}_{\alpha_i} = \sum_{j=1}^{nBlock} \left(\frac{\partial \vec{u}}{\partial k_j} \frac{\partial k_j}{\partial \alpha_i} + \frac{\partial \vec{u}}{\partial \phi_j^0} \frac{\partial \phi_j^0}{\partial \alpha_i} \right) = \sum_{j=1}^{nBlock} \left(\vec{S}_{k_j} \frac{\partial k_j}{\partial \alpha_i} + \vec{S}_{\phi_j^0} \frac{\partial \phi_j^0}{\partial \alpha_i} \right), \tag{3.64}$$

where the $\partial k_j / \partial \alpha_i$ and $\partial \phi_j^0 / \partial \alpha_i$ terms can be calculated very easily. This approach shown in Eq. 3.64 is very easy to implement and there is no need to modify the original sensitivity calculator used in the pixel modeling. However there is one serious problem in that this approach would be very inefficient for the calculation of a small number of parameters, because the algorithm requires the calculation of all $\vec{S}_{k_j}$ and $\vec{S}_{\phi_j^0}$ for $j = 1, 2, ..., nBlock$. Therefore, we need to solve Eq. 3.23 $2 \times nBlock$ times. In most cases, the number of object parameters $nPar$ is much smaller than the number of the gridblocks $nBlock$, i.e., $nPar \ll nBlock$.

An alternative to calculate sensitivities with respect to object parameters would be to modify the mathematical formulation Eq. 3.23. By applying the chain rule on Eq. 3.23 directly, we have

$$J^{(k+1)} S_{\vec{\alpha}}^{(k+1)} = -D^{(k+1)} S_{\vec{\alpha}}^{(k)} - Y^{(k+1)} \cdot T, \tag{3.65}$$

where

$$T = \partial \begin{bmatrix} \vec{k} \\ \vec{\phi}^0 \end{bmatrix} / \partial \vec{\alpha} \tag{3.66}$$

is the object transformation matrix. We should note that T is not constant in the case of a dynamic-object approach, T varies as the object floats around inside the reservoir in different iterations. In this approach, we do not need to modify the numerical simulator and only need to solve Eq. 3.65 $nPar$ times.

The sensitivity of dynamic data (pressure, watercut in the wells and change of saturation distribution) can be calculated in a way similar to that of pixel modeling. The sensitivity of well pressures is

$$\frac{\partial p_{well,i}^{(k)}}{\partial \vec{\alpha}} = S_{2nBlock+i}^{(k)}. \tag{3.67}$$

We only need $\partial Q_w/\partial\alpha = \sum_j \partial q_{w,j}/\partial\alpha$ to calculate the sensitivities of watercuts. $\partial q_{w,j}/\partial\alpha$ is given by Eq. 3.43:

$$\frac{\partial q_{w,j}}{\partial\alpha} = \frac{\partial WI_{w,j}}{\partial\alpha}(p_{w,j} - p_{well,j}) - WI_{w,j}(\frac{\partial p_{w,j}}{\partial\alpha} - \frac{\partial p_{well,j}}{\partial\alpha}), \tag{3.68}$$

where both $\partial p_{w,j}/\partial\alpha$ and $\partial p_{well,j}/\partial\alpha$ are already calculated. $\partial WI_{w,j}/\partial\alpha$ is calculated as follows:

$$\frac{\partial WI_{w,j}}{\partial\alpha} = a_j\left(k_j b_j \frac{\partial p_{o,j}}{\partial\alpha} + k_j c_j \frac{\partial S_{w,j}}{\partial\alpha}\right) + \frac{WI_{w,j}}{k_j}\frac{\partial k_j}{\partial\alpha}, \tag{3.69}$$

or

$$\frac{\partial WI_{w,j}}{\partial\alpha} = a_j\left(k_j b_j \frac{\partial p_{o,j}}{\partial\alpha} + k_j c_j \frac{\partial S_{w,j}}{\partial\alpha}\right) + WI_{w,j}\frac{\partial \ln k_j}{\partial\alpha}, \tag{3.70}$$

where the definitions of a_j, b_j, c_j are the same as in Section 3.2.

The sensitivity of change of water saturation $\partial\Delta S_{w,i}/\partial\alpha$ is given by

$$\frac{\partial\Delta S_{w,i}}{\partial\alpha} = \frac{\partial S_{w,i}^{t_2}}{\partial\alpha} - \frac{\partial S_{w,i}^{t_1}}{\partial\alpha}. \tag{3.71}$$

We should notice that if we substitute α with either k_i or ϕ_i^0, the calculation is exactly the same for both the sensitivity of object parameters and those of the original gridblock parameters. In other words, the gridblock permeability and porosity are special objects, though in that case the transformation matrix is an identity matrix. Therefore the calculation of sensitivity of object parameters is a general formulation and can be used with different variations. As described in the following section, the formulation described in Eq. 3.65 is used to calculate sensitivities in both the wavelet analysis approach and the multicolor seismic data approach.

Although the object approach has many advantages, it may be difficult to choose a suitable model. Object modeling also may oversimplify the system, especially when the reservoir has complex configurations. Therefore, this approach may have very limited application in practice. Many different reparameterization techniques have been introduced to improve the efficiency and overcome the disadvantages of both pixel and object modeling. Oliver *et al.* (2001) gave a comprehensive review of different approaches. These approaches have different drawbacks and disadvantages in practice, as shown in Section 1.2. A more powerful and flexible modeling procedure is desirable for better performance. In the following section, we will explore the application of wavelet analysis in reservoir parameterization.

3.3.3 Parameterization with Wavelets

The central concept in this work was to use wavelet analysis to parameterize the parameter estimation problem. The object of this approach is still to obtain a reliable description of the reservoir, or in the case of a discrete system, the estimation of permeability and/or porosity of each gridblock. Unlike the pixel or object modeling, the wavelet analysis will treat the distribution of permeability and porosity as an image and perform a wavelet transform on this image. Section 2.4 has already described the basic concepts of the wavelet analysis. Although it is the three-dimensional distribution of permeability or porosity that is under investigation, we only considered the two-dimensional distribution of permeability or porosity within a single layer, at least at the current stage of the study. In this way, we do not have to perform the wavelet transform in the vertical direction and can stay with the two-dimensional wavelet transform. The reason is that there are many successful applications in the field of image processing, which is two-dimensional, therefore we can adapt many available theories and implementations of the techniques.

Nevertheless, we can still use the two-dimensional wavelet transform to describe the three-dimensional distribution of a reservoir. In most cases, a three-dimensional discrete reservoir can be treated as a summation of two-dimensional layers. Therefore, each layer can be parameterized using the two-dimensional wavelet transform. The drawback is that this approach can not take advantage of the multiresolution analysis in the vertical direction. However, in order to make problems practical and solvable, the number of layers in reservoirs is usually very small. Under such circumstances, the wavelet analysis will not help much in the vertical direction anyhow.

Wavelet Representation

Any function $f \in \mathbf{L}^2(\mathcal{R}^2)$ can be decomposed over a wavelet basis of $\mathbf{L}^2(\mathcal{R}^2)$:

$$f(x,y) = \sum_{l=-\infty}^{+\infty} \sum_{m,n=-\infty}^{+\infty} \sum_{k=1}^{3} d_l^k[m,n]\psi_{l,m,n}^k(x,y), \tag{3.72}$$

where

$$d_l^k[m,n] = \langle f, \psi_{l,m,n}^k \rangle \quad \text{for} \quad k = 1,2,3. \tag{3.73}$$

$\psi^k_{l,m,n}$ is the l^{th}-level wavelet coefficient on k direction. We always take the reservoir as a discrete system for the purpose of characterization, therefore we focus on the two-dimensional discrete wavelet transform. The original discrete image $a_0[i,j]$ characterizes uniquely f defined by $a_0[i,j] = \langle f, \phi_{0,i,j} \rangle$, and $\phi_{0,i,j}$ is the scaling function. At all scales $l \geq 0$, we denote

$$a_l[m,n] = \langle f, \phi_{l,m,n} \rangle. \tag{3.74}$$

The wavelet representation of a_0 at scale L is

$$a_0[i,j] = \sum_{m,n=-\infty}^{+\infty} a_L[m,n]\phi_{L,m,n}[i,j] + \sum_{l=1}^{L} \sum_{m,n=-\infty}^{+\infty} \sum_{k=1}^{3} d^k_l[m,n]\psi^k_{l,m,n}[i,j], \tag{3.75}$$

where

$$\phi_{L,m,n}[i,j] = \phi_{L,m,n}(x_i, y_j) \quad \text{and} \quad \psi^k_{l,m,n}[i,j] = \psi^k_{l,m,n}(x_i, y_j). \tag{3.76}$$

When a_0 is a finite image of N by N pixels, the wavelet decomposition gives the coefficients in a wavelet basis of $\mathbf{L}^2[0,N]^2$. Assuming $N = 2^J$, the wavelet representation of a_0 at scale L is

$$a_0[i,j] = \sum_{m,n=0}^{2^{J-L}-1} a_L[m,n]\phi_{L,m,n}[i,j] + \sum_{l=1}^{L} \sum_{m,n=0}^{2^{J-l}-1} \sum_{k=1}^{3} c^k_l[m,n]\psi^k_{l,m,n}[i,j]. \tag{3.77}$$

The number of wavelet coefficients

$$2^{2(J-L)} + 3 \cdot \sum_{l=1}^{L} 2^{2(J-l)} = 2^J \cdot 2^J = N \times N \tag{3.78}$$

is the same as the number of original pixels. Therefore this is a perfect-reconstruction filter. We can reconstruct the original image a_0 from the wavelet coefficients perfectly.

Linear Transform Representation

In a more concise way, the discrete wavelet transform can be expressed as a linear transformation

$$\vec{c} = W \cdot \vec{a}_0, \tag{3.79}$$

where $\vec{a}_0$ is the original image, $\vec{c}$ is the wavelet coefficients after the transformation, and W is the wavelet transformation matrix. $\vec{c}$ is a combination of the scaling coefficients $\vec{a}_L$ and the decomposition coefficients $\vec{d}_l^k$, for $l = 1, 2, ..., L$ and $k = 1, 2, 3$. Because W is an orthogonal matrix, the inverse wavelet transform is

$$\vec{a}_0 = W^{-1} \cdot \vec{c} = W^T \cdot \vec{c}. \tag{3.80}$$

In this work, the original finite image is the distribution of permeability k (or $\ln k$) or porosity ϕ^0, i.e., $\vec{a}_0 = \vec{k}$ ($\vec{a}_0 = \ln \vec{k}$) or $\vec{a}_0 = \vec{\phi}^0$. The wavelet transform becomes

$$\vec{c}_k = W \cdot \vec{k}, \tag{3.81}$$

$$\vec{c}_{\ln k} = W \cdot \ln \vec{k}, \tag{3.82}$$

$$\vec{c}_{\phi^0} = W \cdot \vec{\phi}^0. \tag{3.83}$$

The inverse wavelet transform becomes

$$\vec{k} = W^T \cdot \vec{c}_k, \tag{3.84}$$

$$\ln \vec{k} = W^T \cdot \vec{c}_{\ln k}, \tag{3.85}$$

$$\vec{\phi}^0 = W^T \cdot \vec{c}_{\phi^0}. \tag{3.86}$$

One of the major objectives of this study has been to reduce the number of parameters. Apparently the wavelet transform will not help in this sense, because the numbers of parameters are the same for the wavelet coefficients and the original pixels. However, we only need to keep the most important elements of the wavelet coefficients so that the original data can be reconstructed later with only a minimal loss of information. In the field of data compression (mostly audio signals or images), it is very common to obtain a compression ratio of $5 \sim 7$ without degrading the reconstructed signals noticeably.

In imaging applications, a typical gray-level image has 512 by 512 pixels, each pixel coded with 8 bits. Usually images include many types of structures that are difficult to model. Currently, the best image compression algorithms are transform codes, with cosine bases or wavelet bases, e.g., the JPEG and JPEG-2 standards. The efficiency of these bases comes from their ability to construct precise nonlinear

image approximations with few nonzero vectors. With fewer than 1 bit/pixel (a compression ratio of 8.0), visually perfect images can be reconstructed. At 0.25 bit/pixel (a compression ratio about 32.0), the image remains of good quality.

The horizontal distribution of permeability or porosity can be considered as a gray-level image, with the gray level representing the value of properties. Therefore, image compression algorithms could be adapted to reduce the number of parameters in reservoir characterization. In other words, we can use much fewer elements to represent the distribution image and yet still obtain a good reconstruction later on.

Sensitivity of Wavelets

We can still use the chain rule to calculate the sensitivity with respect to wavelet coefficients as follows:

$$S_{c_k} = \sum_{j=1}^{nBlock} \vec{S}_{k_j} \frac{\partial k_j}{\partial \vec{c}_k} = S_k \cdot W^T, \tag{3.87}$$

$$S_{c_{\ln k}} = \sum_{j=1}^{nBlock} \vec{S}_{\ln k_j} \frac{\partial \ln k_j}{\partial \vec{c}_{\ln k}} = S_{\ln k} \cdot W^T, \tag{3.88}$$

$$S_{c_{\phi^0}} = \sum_{j=1}^{nBlock} \vec{S}_{\phi_j^0} \frac{\partial \phi_j^0}{\partial \vec{c}_{\phi^0}} = S_{\phi^0} \cdot W^T. \tag{3.89}$$

In a more general way, Eq. 3.87-3.89 can be written as

$$S_{c_\alpha} = S_\alpha \cdot W^T, \tag{3.90}$$

where $c_\alpha = W(\alpha)$ is the wavelet transform of α. Therefore, in the case of the orthogonal wavelet transform, the wavelet transform of the sensitivity of parameters is equivalent to the sensitivity of the wavelet transform of parameters.

Apparently the chain rule approach is practical in this case, because the wavelet transform is a perfect-reconstruction filter, and the number of pixels in the original image equals that of wavelet coefficients. However, this will not be the case in our implementation because we do not use all the wavelet coefficients. Therefore, we use the same approach as in Section 3.3.2 with some minor modifications. First of all,

the transformation matrix T becomes W^T. Secondly, Eq. 3.69 becomes

$$\frac{\partial WI_{w,j}}{\partial \vec{c}_k} = a_j \left(k_j b_j \frac{\partial p_{o,j}}{\partial \vec{c}_k} + k_j c_j \frac{\partial S_{w,j}}{\partial \vec{c}_k} \right) + \frac{WI_{w,j}}{k_j} \vec{W}_j^T, \tag{3.91}$$

$$\frac{\partial WI_{w,j}}{\partial \vec{c}_{\ln k}} = a_j \left(k_j b_j \frac{\partial p_{o,j}}{\partial \vec{c}_{\ln k}} + k_j c_j \frac{\partial S_{w,j}}{\partial \vec{c}_{\ln k}} \right) + WI_{w,j} \vec{W}_j^T, \tag{3.92}$$

$$\frac{\partial WI_{w,j}}{\partial \vec{c}_{\phi^0}} = a_j \left(k_j b_j \frac{\partial p_{o,j}}{\partial \vec{c}_{\phi^0}} + k_j c_j \frac{\partial S_{w,j}}{\partial \vec{c}_{\phi^0}} \right), \tag{3.93}$$

where $\vec{W}_j^T$ is the j^{th} column of W^T. In order to calculate the sensitivity of a particular wavelet coefficient, we only need to solve Eq. 3.23 once.

3.3.4 The Haar Wavelets

In this work, we mainly used the Daubechies wavelet family (Daubechies 1988) for the wavelet analysis. Daubechies wavelets have several properties that are suitable for our purpose. First of all, Daubechies wavelets have a compact support of minimum size for any given number n_v of vanishing moments. Furthermore, Daubechies wavelets are also orthogonal transforms and it is much easier to analyze the variance and resolution of the parameter estimation procedures.

We denote the Daubechies wavelet with a vanishing moment of p as D_p, and the discrete filter $h[n]$ has $2p$ non-zero coefficients. In this work, we used only Daubechies wavelets with vanishing moments of 1, 2, and 3. When $p = 1$, the Daubechies wavelet becomes the Haar wavelet, which is very simple and easy to implement.

The Haar wavelet has the shortest support among all orthogonal wavelets. The Haar wavelet is not well-adapted to approximating smooth functions because it has only one vanishing moment. However, the Haar wavelet does have some attractive properties that suit our purpose very well. Above all, the formulation is very simple, therefore the transform algorithm is straightforward and fast, and it is also easy to implement. Actually in the case of a $2^J \times 2^J$ image, we do not even need to consider boundary wavelets. Secondly, reservoirs in the real world always have some singularities (or discontinuities in some sense), such as faults, channels and shales. Because wavelets that overlap the singularities always create high amplitude coefficients, it

would be reasonable to use wavelets with short supports. The Haar wavelet has the shortest support.

Definition of the Haar Wavelet

We will demonstrate the power and feasibility of wavelet analysis with the Haar wavelet basis. The Haar basis is obtained with a multiresolution approximation composed of piecewise constant functions. The scaling function is $\phi = \mathbf{1}[0,1]$, where the indicator function $\mathbf{1}[a,b]$ is 1 in $[a,b]$ and 0 outside. Because the filter $h[n] = \langle \frac{1}{\sqrt{2}}\phi\left(\frac{t}{2}\right), \phi(t-n)\rangle$, it follows that

$$h[n] = \begin{cases} \frac{1}{\sqrt{2}} & \text{if} \quad n = 0,1 \\ 0 & \text{otherwise} \end{cases} \tag{3.94}$$

and

$$g[n] = (-1)^{1-n} h[1-n] = \begin{cases} -\frac{1}{\sqrt{2}} & \text{if} \quad n = 0 \\ \frac{1}{\sqrt{2}} & \text{if} \quad n = 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.95}$$

Hence

$$\frac{1}{\sqrt{2}}\psi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} g[n]\phi(t-n) = \frac{1}{\sqrt{2}}(\phi(t-1) - \phi(t)), \tag{3.96}$$

so

$$\psi(t) = \begin{cases} -1 & \text{if} \quad 0 \le t < \frac{1}{2} \\ 1 & \text{if} \quad \frac{1}{2} \le t < 1 \\ 0 & \text{otherwise} \end{cases} . \tag{3.97}$$

Fig. 3.15 shows the scaling function $\phi(t)$ and the wavelet $\psi(t)$ for the Haar wavelet basis.

As defined in Eq. A.37 and Eq. A.38, the fast wavelet decomposition for the Haar wavelet is

$$\begin{aligned} a_{j+1}[p] &= \sum_{n=-\infty}^{+\infty} h[n-2p]a_j[n] = h[0]a_j[2p] + h[1]a_j[2p+1] \\ &= \frac{1}{\sqrt{2}}(a_j[2p] + a_j[2p+1]), \end{aligned} \tag{3.98}$$

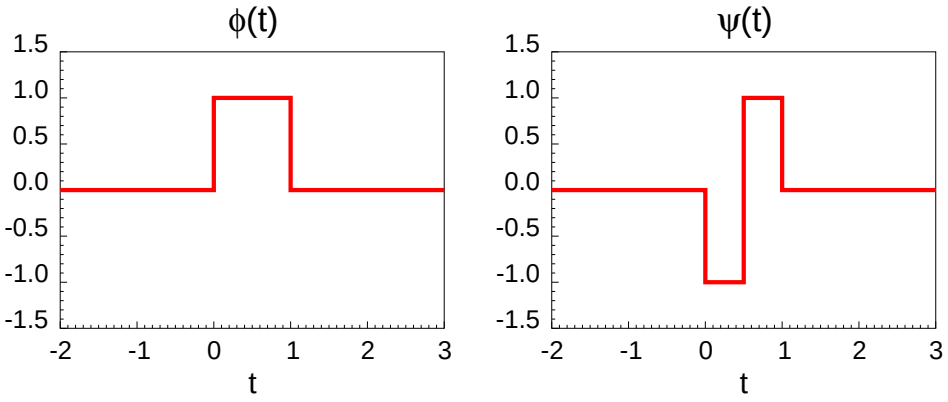

Figure 3.15: The scaling function ϕ and the wavelet ψ for the Haar wavelet basis.

and

$$\begin{aligned} d_{j+1}[p] &= \sum_{n=-\infty}^{+\infty} g[n-2p]a_j[n] = g[0]a_j[2p] + g[1]a_j[2p+1] \\ &= \frac{1}{\sqrt{2}}(-a_j[2p] + a_j[2p+1]). \end{aligned} \tag{3.99}$$

Eq. 3.98 and Eq. 3.99 have very obvious mathematical meanings: the scaling coefficient $a_{j+1}[p]$ is the normalized arithmetic average of two adjacent data points ($a_j[2p]$ and $a_j[2p+1]$) at lower decomposition level j (higher resolution level); the wavelet coefficient $d_{j+1}[p]$ is the arithmetic difference of those two coefficients. Eq. 3.98 can be rewritten as

$$a_j[p] = \left(\frac{1}{\sqrt{2}}\right)^j \sum_{n=0}^{2^j-1} a_0[2^j p + i], \tag{3.100}$$

which shows that any scaling coefficient at Level j is a normalized arithmetic average of 2^j signal points at Level 0, i.e., the original signal $a_0[n]$.

At the reconstruction, we have

$$a_j[2p] = h[0]a_{j+1}[p] + g[0]d_{j+1}[p] = \frac{1}{\sqrt{2}}(a_{j+1}[p] - d_{j+1}[p]), \tag{3.101}$$

$$a_j[2p+1] = h[1]a_{j+1}[p] + g[1]d_{j+1}[p] = \frac{1}{\sqrt{2}}(a_{j+1}[p] + d_{j+1}[p]). \tag{3.102}$$

As demonstrated through Eq. 3.98 to Eq. 3.102, the wavelet transform is a linear transform, therefore any wavelet coefficient $a_j[p]$ is a linear combination of the original signal $a_0[n]$.

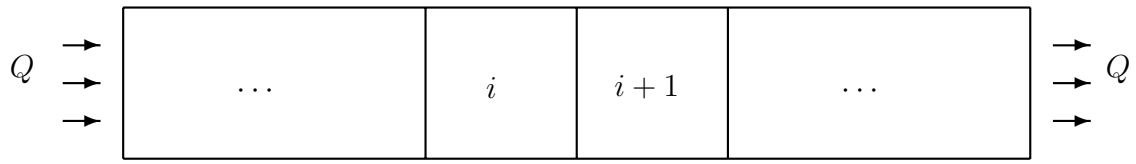

Figure 3.16: One-dimensional steady-state flow through a series of blocks with permeability of k_i

One-Dimensional Flow Example

Until this point, we have not explored any specific advantage of the wavelet analysis in parameter estimation problems. The following simple example shows some useful aspects of wavelets under the circumstance of a one-dimensional flow through a porous medium. Fig. 3.16 shows a series of cores connected together, each core with a different permeability k_i. At steady state, the pressure at the right end is fixed at 0, while the pressure at the left end is p. A constant flow rate Q is injected from the left end. The Darcy's law gives the pressure drop as

$$p = \frac{Q\mu L}{A}\sum_{j}^{N}\frac{1}{k_j}, \tag{3.103}$$

where A is the cross-section area, L is the length of each block, N is the total number of blocks. The sensitivity of pressure p with respect to permeability of Block $\#i$ (k_i) is given by

$$\frac{\partial p}{\partial k_i} = \frac{Q\mu L}{A}\frac{\partial}{\partial k_i}\sum_{j}^{N}\frac{1}{k_j} = -\frac{Q\mu L}{A}\frac{1}{k_i^2}. \tag{3.104}$$

The wavelet decomposition of the permeability distribution at Level 1 is

$$a_i = \frac{1}{\sqrt{2}}(k_{2i} + k_{2i+1}), \tag{3.105}$$

$$d_i = \frac{1}{\sqrt{2}}(-k_{2i} + k_{2i+1}). \tag{3.106}$$

The reconstruction is given by

$$k_{2i} = \frac{1}{\sqrt{2}}(a_i - d_i), \tag{3.107}$$

$$k_{2i+1} = \frac{1}{\sqrt{2}}(a_i + d_i). \tag{3.108}$$

Therefore the sensitivities of pressure p with respect to the scaling coefficient a_i and the wavelet coefficient d_i are

$$\frac{\partial p}{\partial a_i} = \frac{\partial p}{\partial k_{2i}}\frac{\partial k_{2i}}{\partial a_i} + \frac{\partial p}{\partial k_{2i+1}}\frac{\partial k_{2i+1}}{\partial a_i} = -\frac{1}{\sqrt{2}}\frac{Q\mu L}{A}(\frac{1}{k_{2i}^2} + \frac{1}{k_{2i+1}^2}), \tag{3.109}$$

$$\frac{\partial p}{\partial d_i} = \frac{\partial p}{\partial k_{2i}}\frac{\partial k_{2i}}{\partial d_i} + \frac{\partial p}{\partial k_{2i+1}}\frac{\partial k_{2i+1}}{\partial d_i} = -\frac{1}{\sqrt{2}}\frac{Q\mu L}{A}(-\frac{1}{k_{2i}^2} + \frac{1}{k_{2i+1}^2}). \tag{3.110}$$

In a case in which the core is almost homogeneous, i.e., $\{k_i\}_{i=1,\ldots,N}$ are are not very different from each other, $\partial p/\partial d_i$ would be very close to zero. In other words, the pressure response is not sensitive to the wavelet coefficient d_i. In the context of parameter estimation, the value of d_i will not be resolved. Therefore, it is not necessary to calculate the sensitivity of d_i, because we can not have a good estimation of it. In this particular case, only one parameter can be truly resolved, which is the average permeability

$$\bar{k} = \frac{N}{\sum_{j=1}^{N} 1/k_j}, \tag{3.111}$$

because the pressure response can be written as

$$p = \frac{Q\mu NL}{A\bar{k}}. \tag{3.112}$$

The average permeability $\bar{k}$ is the only parameter that can change the output p. This is a very extreme case: there is no transient for the steady-state flow, thus there is only one data point for the response: the pressure at the left end. Even so, we can make some very important observations regarding this simple case:

- Sensitivities with respect to certain wavelet coefficients are very small, i.e., the reservoir response is not sensitive to certain linear combinations of original parameters.

- In some cases, it is almost impossible to resolve individual parameters, if the response data do not have enough resolving power. However, certain combinations of parameters can be resolved much better, e.g., the average permeability $\bar{k}$ in the test case.

- Sensitivities with respect to high permeability are smaller than sensitivities with respect to low permeability, as shown in Eq. 3.104. Even in logarithm scales, this statement is still true, because

$$\frac{\partial p}{\partial \ln k_i} = -\frac{Q\mu L}{A}\frac{1}{k_i}. \tag{3.113}$$

These observations are also applicable to more general situations, e.g., parameter estimation with production history matching. The following section demonstrates the application of wavelets in a more general situation.

Verification of Sensitivity Calculation

The accuracy of the modified GPST method for the pixel modeling has been validated by the numerical experiment shown in Section 3.2.4. In this section, we will use the same example to verify the modified GPST method for wavelet coefficients, still using the substitution method as the benchmark.

In this work, we mostly used the logarithm of permeability ($\ln k$) as the parameters, instead of permeability (k). Therefore, we calculated the sensitivity of dynamic data with respect to the wavelet transform of $\ln k$. The wavelet representation of $\ln k$ at scale L is

$$\ln k_{i,j} = \sum_{m,n=-\infty}^{+\infty} a_L[m,n]\phi_{L,m,n}[i,j] + \sum_{l=1}^{L}\sum_{m,n=-\infty}^{+\infty}\sum_{k=1}^{3} d_l^k[m,n]\psi_{l,m,n}^k[i,j], \tag{3.114}$$

with $a_L[m,n]$ is the scaling coefficient at scale L, and $d_l^k[m,n]$ is the wavelet coefficient at scale l in the k^{th} direction. In this particular case, the decomposition level is set to $L = 2$. The sensitivity of dynamic data with respect to $a_2[0,1]$ is calculated using both the two methods.

For the Haar wavelet, $a_2[0,1]$ has a very simple form:

$$a_2[0,1] = \frac{1}{4}\sum_{i=0}^{4}\sum_{j=4}^{7}\ln k_{i,j}, \tag{3.115}$$

representing an arithmetic average of 16 gridblocks. Effectively, a small perturbation on $a_2[0,1]$ imposes perturbations on these 16 gridblocks, similar to a static object. We

should also notice that these 16 gridblocks are connected together to form a square region around Block #101, i.e., the gridblock located at (3,5).

Figs. 3.17 and 3.18 show the sensitivities of dynamic data with respect to $a_2[0,1]$ calculated by both the substitution method and the modified GPST method. The two computations give almost identical results for this particular case.

The same calculation can be performed for the wavelets of porosity ϕ^0. In this case,

$$a_2[0,1] = \frac{1}{4}\sum_{i=0}^{4}\sum_{j=4}^{7}\phi^0_{i,j}. \tag{3.116}$$

Figs. 3.19, 3.20 show the sensitivities of dynamic data with respect to $a_2[0,1]$, though $a_2[0,1]$ is the wavelet for ϕ^0 now.

For the logarithm permeability field, we can compare the impacts of different parameters: $k_{3,5}$, $a_2[0,1]$, and $d_1^3[1,2]$, with

$$d_1^3[1,2] = \frac{1}{2}(\ln k_{2,4} + \ln k_{3,5} - \ln k_{2,5} - \ln k_{3,4}). \tag{3.117}$$

Figs. 3.21 through 3.23 show the sensitivity of the bottomhole pressure in Well #2 with respect to these three parameters. As we can see, though these three parameters are localized in the same region around Block (3,5), the sensitivities are quite different from each other. The sensitivity of $a_2[0,1]$ is very close to that of $\ln k_{3,5}$, because $a_2[0,1]$ is a normalized arithmetic average of permeability of the region around Block (3,5). However, the sensitivity of $d_1^3[1,2]$ is about one magnitude smaller than that of $\ln k_{3,5}$, because $d_1^3[1,2]$ represents an interblock difference with several blocks. Obviously, the well response is not sensitive to the local variation of gridblock properties.

Figs. 3.24 through 3.26 show the sensitivity of the watercut in Well #2 with respect to these three parameters. The comparison between the sensitivities are similar to those of the bottomhole pressure. These comparisons show that well responses are sensitive to certain linear combinations of gridblock properties, as shown in this case, the arithmetic average. On the other hand, well responses are not sensitive to some other linear combinations, e.g., as shown above, the local variation represented by $d_1^3[1,2]$.

One issue of wavelet reparameterization is to reduce the number of parameters to be estimated. In Chapter 5, we will discuss how to compress the model space using wavelets on the basis of variance analysis and data integration. Wavelets are also important to uncertainty and variance analysis, which are discussed in Chapter 4. Although the wavelet transform itself is a linear transformation (perfect reconstruction filter), the approaches proposed in next few chapters give nonuniform representation of reservoir properties and use different resolutions at different iterations. This is usually referred to as *nonlinear estimation* (Mallat 1998), although we will use the terminology *nonuniform resolution* here to avoid confusion.

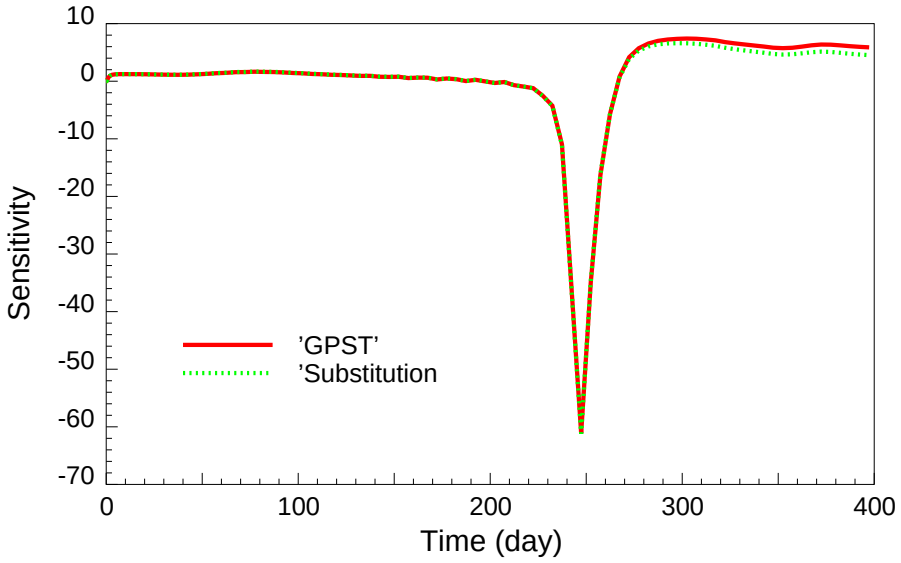

Figure 3.17: Sensitivity of bottomhole pressure in Well #2 with respect to $a_2[0.1]$, wavelet of $\ln k$

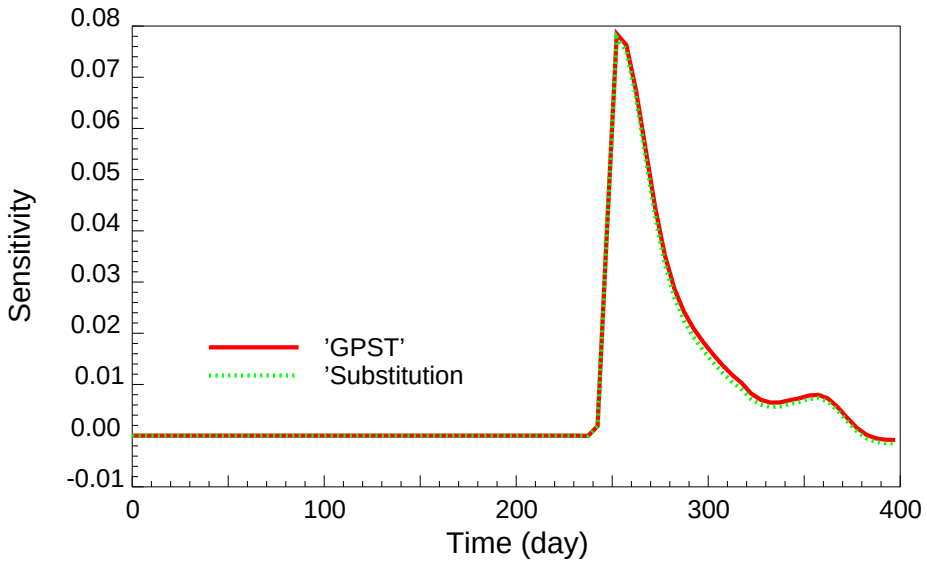

Figure 3.18: Sensitivity of watercut in Well #2 with respect to $a_2[0, 1]$, wavelet of $\ln k$

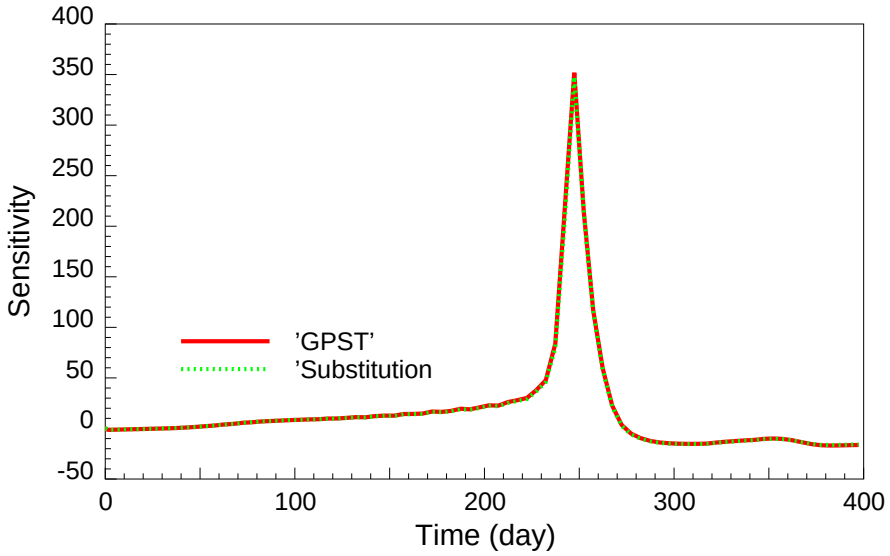

Figure 3.19: Sensitivity of bottomhole pressure in Well #2 with respect to $a_2[0,1]$, wavelet of ϕ^0

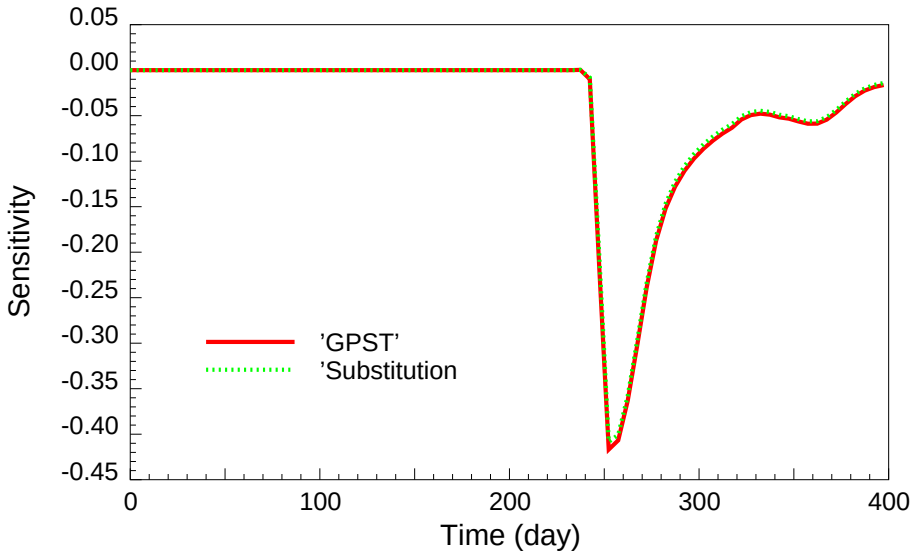

Figure 3.20: Sensitivity of watercut in Well #2 with respect to $a_2[0,1]$, wavelet of ϕ^0

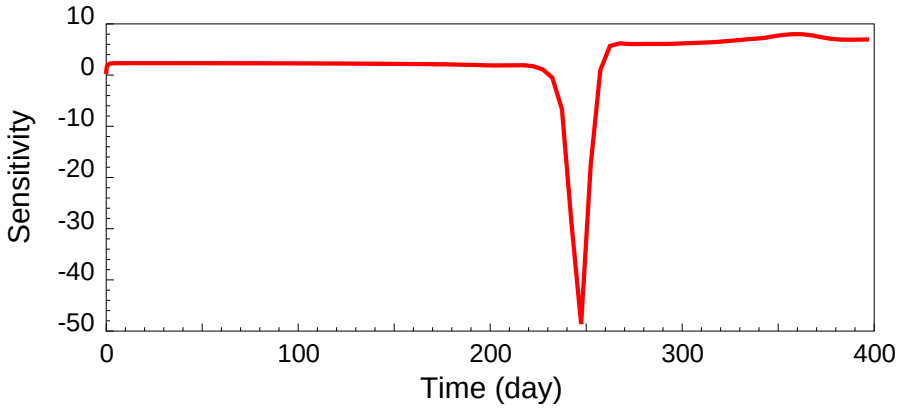

Figure 3.21: Sensitivity of bottomhole pressure in Well #2 with respect to $\ln k_{3,5}$

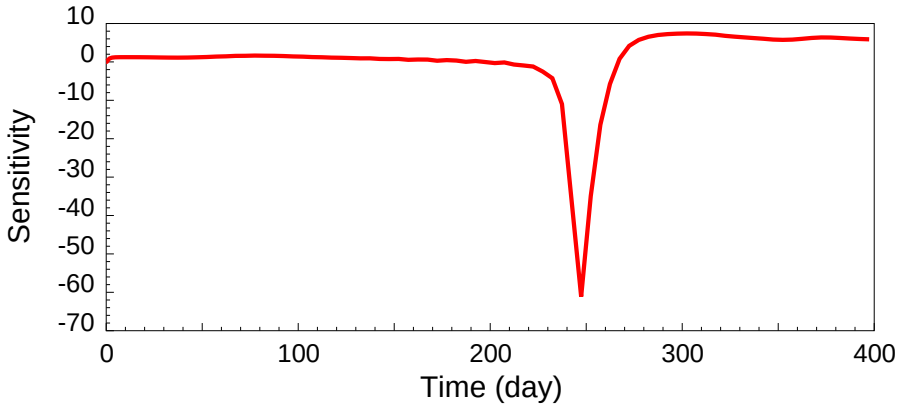

Figure 3.22: Sensitivity of bottomhole pressure in Well #2 with respect to $a_2[0,1]$, wavelet of $\ln k$

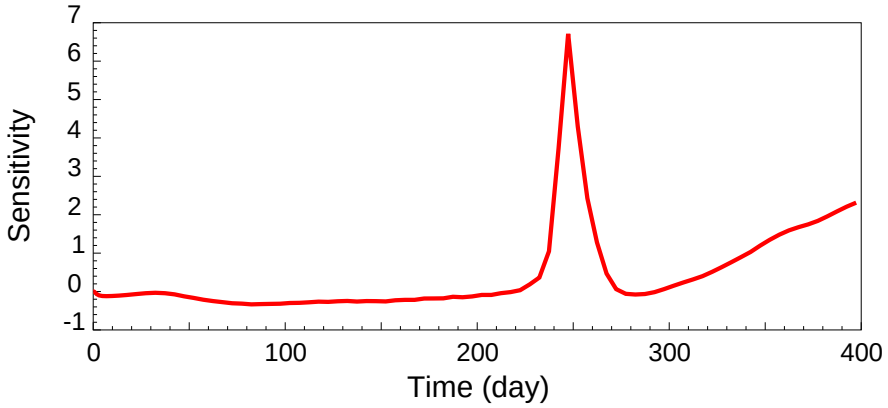

Figure 3.23: Sensitivity of bottomhole pressure in Well #2 with respect to $d_1^3[1,2]$, wavelet of $\ln k$

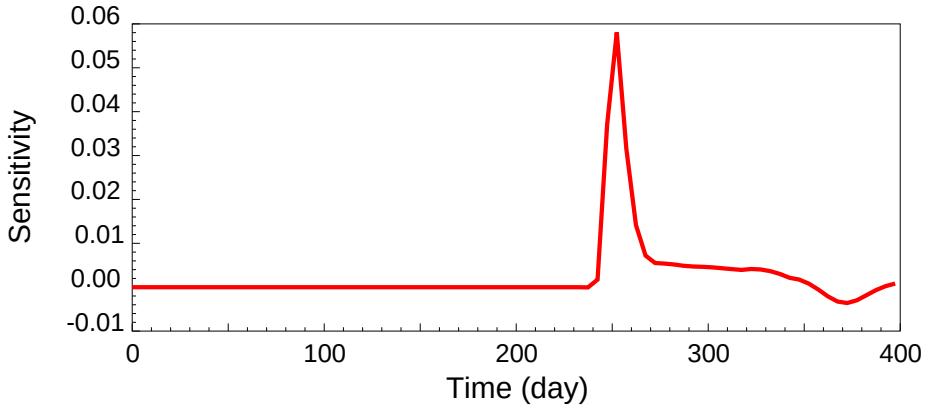

Figure 3.24: Sensitivity of watercut in Well #2 with respect to $\ln k_{3,5}$

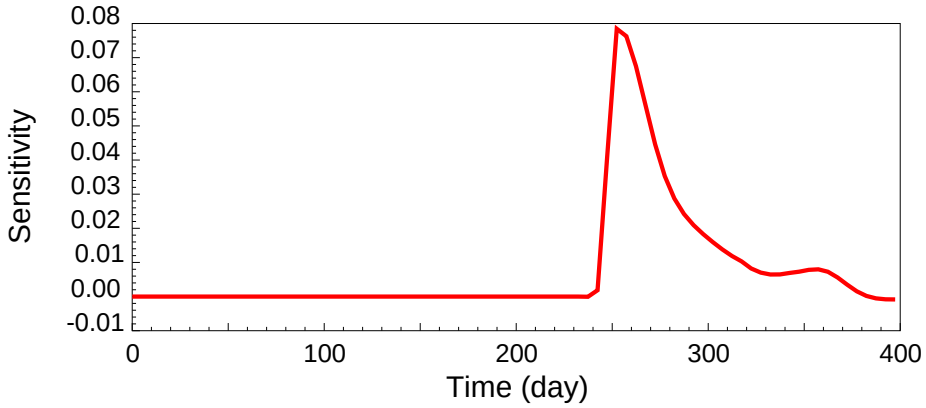

Figure 3.25: Sensitivity of watercut in Well #2 with respect to $a_2[0,1]$, wavelet of $\ln k$

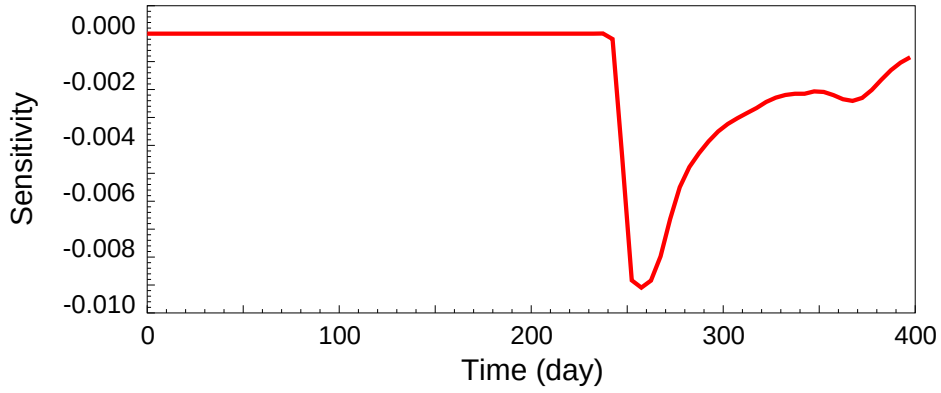

Figure 3.26: Sensitivity of watercut in Well #2 with respect to $d_1^3[1,2]$, wavelet of $\ln k$

Chapter 4

Analysis of Variance and Data Integration

4.1 Analysis of Variance and Resolution

We know that inverse problems are usually under-determined and frequently have nonunique solutions. One important question we should always ask ourselves: what does it mean to "solve" an inverse problem? The answer to this question depends on different applications. In reservoir parameter estimation, there are several important considerations:

- *The uncertainty of the estimated values*: the probability that the true values of the model parameters lie in a given range.
- *The resolution of the estimation*: how closely the values the model parameters are determined by the estimation procedure.

4.1.1 Computation of the Posterior Covariance Operator

If the probability density function f_M defined in Eq. 2.1 does not have a very complicated shape, i.e., it is very close to multi-Gaussian distribution, it is possible to describe it adequately by its central estimators and estimators of dispersion. Among

the central estimators, the most frequently used are the maximum likelihood value $\vec{\alpha}_{ML}$:

$$\vec{\alpha}_{ML} \text{ Maximizes } f_M, \tag{4.1}$$

and the mean value (the mathematical expectation) $\langle\vec{\alpha}\rangle$:

$$\langle\vec{\alpha}\rangle = \frac{1}{N}\int_M \vec{\alpha} f_M(\vec{\alpha}) d\vec{\alpha}. \tag{4.2}$$

We always use the posterior covariance operator as the estimator of dispersion:

$$C_{M|D} = \frac{1}{N}\int_M (\vec{\alpha} - \langle\vec{\alpha}\rangle)(\vec{\alpha} - \langle\vec{\alpha}\rangle)^T f_M(\vec{\alpha}) d\vec{\alpha} \tag{4.3}$$

$$= \frac{1}{N}\int_M \vec{\alpha}\vec{\alpha}^T f_M(\vec{\alpha}) d\vec{\alpha} - \langle\vec{\alpha}\rangle \langle\vec{\alpha}\rangle^T. \tag{4.4}$$

An important note that has to be taken is that the covariance operator can give useful information only in the case when $f_M(\vec{\alpha})$ can be fitted reasonably well by a Gaussian function.

If the forward mathematical model $\vec{g}(\vec{\alpha})$ is linear:

$$\vec{g}(\vec{\alpha}) = G\vec{\alpha}, \tag{4.5}$$

the posterior probability density is then Gaussian, and the covariance operator is given by

$$C_{M|D} = \left(G^T C_D^{-1} G + C_M^{-1}\right)^{-1}. \tag{4.6}$$

In most cases, $\vec{g}(\vec{\alpha})$ is nonlinear. If the probability distribution of the model parameters $f_M(\vec{\alpha})$ is reasonably well behaved, we can still approximate Eq. 4.6 with a linearized approximation around a reference point $\vec{\alpha}_{ref}$:

$$\vec{g}(\vec{\alpha}) \approx \vec{g}(\vec{\alpha}_{ref}) + G(\vec{\alpha} - \vec{\alpha}_{ref}). \tag{4.7}$$

For true nonlinear problems, the approximation Eq. 4.7 is no longer applicable. If the nonlinearity is not too strong, then $\vec{g}(\vec{\alpha})$ is linearizable in the region of significant posterior probability density, i.e., $\vec{g}(\vec{\alpha})$ is linearizable around the maximum likelihood point $\vec{\alpha}_{ML}$:

$$\vec{g}(\vec{\alpha}) \approx \vec{g}(\vec{\alpha}_{ML}) + G_{ML}(\vec{\alpha} - \vec{\alpha}_{ML}), \tag{4.8}$$

where $\vec{\alpha}_{ML}$ is obtained using a nonlinear algorithm, and

$$G_{ML} = \left(\frac{\partial \vec{g}}{\partial \vec{\alpha}}\right)_{ML}. \tag{4.9}$$

The covariance operator is

$$C_{M|D} = \left(G_{ML}^T C_D^{-1} G_{ML} + C_M^{-1}\right)^{-1}. \tag{4.10}$$

As shown in Eq. 2.46, the Hessian matrix at the n^{th} iteration for the Gauss-Newton method is defined as

$$H_n = G_n^T C_D^{-1} G_n + C_M^{-1}. \tag{4.11}$$

If the nonlinear regression is close enough to the maximum likelihood point and the assumption of Gaussian distribution still holds, the covariance operator can be written as the inverse of the Hessian matrix:

$$C_{M|D} = H_n^{-1}, \tag{4.12}$$

Using Eq. 4.12, we can obtain an estimation of the covariance operator from the Gauss-Newton regression engine very easily. In practice, the Gaussian assumption usually does not hold, therefore, the estimation given in Eq. 4.12 is only an approximation and not well defined. Even though, the inverse of the Hessian matrix is still an important indicator for uncertainty of estimation.

4.1.2 Analysis of Resolution

The diagonal elements of the posterior covariance matrix calculated in Section 4.1.1 are the variances of the estimation, with the square roots as the standard deviations. In most cases, the variances/standard deviations are used to describe the uncertainties of the posterior values of the model parameters.

The off-diagonal elements (covariances) describe the correlations between different parameters. A strong correlation in uncertainties means that the two parameters have not been resolved independently by the data set, and only some linear combinations of the parameters are resolved.

The concept of resolution defined by Tarantola (1987) has two meanings:

- A parameter is well resolved by the data set if the standard deviation of the estimation is much smaller than the prior one.

- Usually covariance matrices $C_{M|D}$ are not diagonal, but are "band diagonal" with covariances between neighboring parameters not zero. In this case, the neighboring parameters have correlated errors. The greater the correlation, the worse is the resolution.

Let $\vec{\alpha}^{true}$ represent the "true" model parameters. The observed data $\vec{d}^{obs}$ generally do not equal the computed values $\vec{g}(\vec{\alpha}^{true})$ due to measurement and modelization errors. Tarantola (1987) gave the following relationship:

$$\langle\vec{\alpha}\rangle - \vec{\alpha}^{pri} = R(\vec{\alpha}^{true} - \vec{\alpha}^{pri}), \tag{4.13}$$

where the linearized resolution operator R is defined as

$$\begin{aligned} R &= \left[G^T C_D^{-1} G + C_M^{-1}\right]^{-1} G^T C_D^{-1} G & (4.14)\\ &= I - C_{M|D} C_M^{-1}. & (4.15) \end{aligned}$$

Typically, the resolution operator is introduced only in approaches not directly based on probabilistic concepts. In this work, this is not the case. Nevertheless, we can use the resolution operator in some special cases to explore some basic concepts.

4.1.3 Some Special Cases

High-Certainty Prior Information

If the prior covariance operator C_M is with very small elements, Eq. 4.14 can be rewritten as

$$R = I - \left(C_M - C_M G^T \left(G C_M G^T + C_D\right)^{-1} G C_M\right) C_M^{-1} \approx I - C_M C_M^{-1} = 0. \tag{4.16}$$

In this case, the prior information is good enough to resolve the model parameters, the inverse procedure would not give much new resolving power.

No Prior Information

If there is almost no prior information or the prior information is highly uncertain, C_M would be very large and C_M^{-1} would be close to zero. Eq. 4.14 becomes

$$R = H^{-1} \cdot H, \tag{4.17}$$

and Eq. 4.6 becomes

$$C_{M|D} = H^{-1}, \tag{4.18}$$

where

$$H = G^T C_D^{-1} G \tag{4.19}$$

is the Hessian matrix without any prior information. Apparently, R should equal the identity matrix I in this case. However, in most cases, the Hessian matrix H is very close to singularity and is therefore unstable. Small perturbations on the data set will cause large changes in the model parameters. Menke (1984) described the singular value decomposition (SVD) method to reduce the uncertainty of estimation. The singular value decomposition of $C_D^{-1/2}G$ is

$$C_D^{-\frac{1}{2}} G = U \Lambda V^T, \tag{4.20}$$

where both U and V are orthogonal matrices with a dimension of $nPar \times nPar$ and $nData \times nData$ respectively. Λ is a $nPar \times nData$ diagonal matrix with diagonals λ_i, for $i = 1, 2, ..., nPar$ (assuming $nPar < nData$). Eq. 4.20 can be written as

$$C_D^{-\frac{1}{2}} G = [U_p \, U_0] \begin{bmatrix} \Lambda_p & 0 \\ 0 & \Lambda_0 \end{bmatrix} \begin{bmatrix} V_p^T \\ V_0^T \end{bmatrix}, \tag{4.21}$$

where Λ_0 is almost a zero matrix. For a particular p, we can take Λ_0 as a zero matrix and Eq. 4.21 becomes

$$C_D^{-\frac{1}{2}} G = U_p \Lambda_p V_p^T. \tag{4.22}$$

Originally the Hessian matrix is

$$\begin{aligned} H &= \left(C_D^{-\frac{1}{2}} G\right)^T \left(C_D^{-\frac{1}{2}} G\right) = \left(U \Lambda V^T\right)^T \left(U \Lambda V^T\right) \\ &= V \Lambda^T \Lambda V^T = V \tilde{\Lambda} V^T, \end{aligned} \tag{4.23}$$

where $\tilde{\Lambda}$ is a diagonal matrix with diagonal elements λ_i^2, for $i = 1, 2, ..., nPar$. The posterior covariance matrix $C_{M|D}$ is

$$C_{M|D} = H^{-1} = V\tilde{\Lambda}^{-1}V^T. \tag{4.24}$$

We can define the total uncertainty of estimation Σ as the summation of variances (the trace of the covariance matrix):

$$\begin{aligned} \Sigma &= Trace(C_{M|D}) = Trace(V\tilde{\Lambda}^{-1}V^T) && (4.25)\\ &= Trace(\tilde{\Lambda}^{-1}) = \sum_{i=1}^{nPar} \frac{1}{\lambda_i^2}. && (4.26) \end{aligned}$$

In this case, the resolution matrix R is an identity matrix, though Eq. 4.25 shows that the total uncertainty will be very large if any of λ_i is close to zero. Using the singular value decomposition defined in Eq. 4.22, the Hessian matrix H becomes

$$H = V_p\tilde{\Lambda}_pV_p^T, \tag{4.27}$$

the posterior covariance matrix $C_{M|D}$ becomes

$$C_{M|D} = H^{-g} = V_p\tilde{\Lambda}_p^{-1}V_p^T. \tag{4.28}$$

The total certainty is

$$\Sigma_p = Trace(C_{M|D}) = \sum_{i=1}^{p} \frac{1}{\lambda_i^2}, \tag{4.29}$$

and the resolution matrix is

$$R = V_pV_p^T. \tag{4.30}$$

The Gauss-Newton method becomes

$$\begin{aligned} \vec{\alpha}_{n+1} &= \vec{\alpha}_n + V_p\Lambda_p^{-1}U_p^TC_D^{-\frac{1}{2}}\left(\vec{g}(\vec{\alpha}_n) - \vec{d}^{obs}\right) \\ &= \vec{\alpha}_n + \left(C_D^{-\frac{1}{2}}G\right)^{-g} C_D^{-\frac{1}{2}}\left(\vec{g}(\vec{\alpha}_n) - \vec{d}^{obs}\right), && (4.31) \end{aligned}$$

where

$$\left(C_D^{-\frac{1}{2}}G\right)^{-g} = V_p\Lambda_p^{-1}U_p^T. \tag{4.32}$$

Uniform Independent Model Variance ($C_M = \sigma_M^2 I$)

When we have prior information, the SVD analysis is no longer applicable. However, we can derive a SVD formulation for a special case: the prior model covariance matrix is diagonal with the same diagonal element σ_M^2, i.e, $C_M = \sigma_M^2 I$. The Hessian matrix is

$$\begin{aligned} H &= G^T C_D^{-1} G + C_M^{-1} = V \tilde{\Lambda} V^T + \sigma_M^{-2} I \\ &= V \left(\tilde{\Lambda} + \sigma_M^{-2} I \right) V^T = V diag \left(\lambda_i^2 + 1/\sigma_M^2 \right) V^T. \end{aligned} \tag{4.33}$$

the covariance of estimation becomes

$$C_{M|D} = H^{-1} = V \cdot diag \left(\frac{1}{\lambda_i^2 + 1/\sigma_M^2} \right) \cdot V^T. \tag{4.34}$$

Therefore the total uncertainty is

$$\Sigma = Trace(C_{M|D}) = \sum_{i=1}^{nPar} \frac{1}{\lambda_i^2 + 1/\sigma_M^2}, \tag{4.35}$$

the resolution matrix is

$$R = V \cdot diag \left(\frac{\lambda_i^2}{\lambda_i^2 + 1/\sigma_M^2} \right) V^T. \tag{4.36}$$

There are several aspects of this case that we need to note:

- *The total certainty is bounded.* Even with some λ_i close to zero, the total uncertainty is bounded by $nPar \times \sigma_M^2$. In the case in which there is no prior information, any small λ_i will make the uncertainty very large.

- *The Hessian matrix is better-conditioned.* Eq. 4.33 shows that the condition number of the Hessian matrix is

  $$Cond(H) = \frac{\lambda_{max}^2 + 1/\sigma_M^2}{\lambda_{min}^2 + 1/\sigma_M^2}, \tag{4.37}$$

  which is much better than the condition number for the case with no prior information:

  $$Cond(H) = \frac{\lambda_{max}^2}{\lambda_{min}^2}. \tag{4.38}$$

  With a better condition number, it is much more stable to solve the Hessian matrix.

- *The resolution matrix is not identity in any case*, even with the full set of parameters.

These observations are applicable in the general situation, with some modifications. The total certainty is bounded by $Trace(C_M)$. The condition number of the Hessian matrix is bounded by

$$Cond(H) \leq \frac{\lambda_{max}^2 + 1/\sigma_{min}^2}{\lambda_{min}^2 + 1/\sigma_{max}^2}, \tag{4.39}$$

where σ_{min}^2 and σ_{max}^2 are the largest and smallest eigenvalues of C_M respectively. Therefore, the prior information always reduces the total uncertainty of estimation and stabilizes the inverse procedure. In practice, it is always desirable to obtain as much prior information as possible.

4.2 Analysis of Variances for Wavelets

In this work, we always use the wavelets of the original pixel parameters as the model parameters. As defined in Section 3.3.3, we have

$$\vec{c} = W \cdot \vec{\alpha}, \tag{4.40}$$

where W is the wavelet transformation matrix, an orthogonal matrix.

4.2.1 Gauss-Newton Method for Wavelets

Section 2.2.2 gives the formulation for the Gauss-Newton algorithm:

$$\vec{\alpha}_{n+1} = \vec{\alpha}_n - \mu_n \left(H^{-1} \nabla E \right)_n . \tag{4.41}$$

Multiplying both sides of Eq. 4.41 by W, we have

$$\begin{aligned} W \cdot \vec{\alpha}_{n+1} &= W \cdot \vec{\alpha}_n - \mu_n \left(W H^{-1} \nabla E \right)_n \\ &= W \cdot \vec{\alpha}_n - \mu_n \left(W H^{-1} W^T W \nabla E \right)_n . \end{aligned} \tag{4.42}$$

In Eq. 4.42, several terms can be simplified as follows:

$$\vec{c}_n = W \cdot \vec{\alpha}_n, \tag{4.43}$$

$$\begin{aligned} WHW^T &= W\left(G^T C_D^{-1} G + C_M^{-1}\right) W^T \\ &= \left(GW^T\right)^T C_D^{-1}\left(GW^T\right) + WC_M^{-1}W^T \\ &= G_W^T C_D^{-1} G_W + C_W^T = H_W, \end{aligned} \tag{4.44}$$

and

$$\begin{aligned} W\nabla E &= W\left(G^T C_D^{-1}(\vec{d}^{cal} - \vec{d}^{obs}) + C_M^{-1}(\vec{\alpha} - \vec{\alpha}^{pri})\right) \\ &= \left(GW^T\right)^T C_D^{-1}(\vec{d}^{cal} - \vec{d}^{obs}) + WC_M^{-1}W^T(W\vec{\alpha} - W\vec{\alpha}^{pri}) \\ &= G_W^T C_D^{-1}(\vec{d}^{cal} - \vec{d}^{obs}) + C_W^{-1}(\vec{c} - \vec{c}^{pri}) = \nabla E_W, \end{aligned} \tag{4.45}$$

where $C_W = WC_MW^T$, $G_W = GW^T$. Therefore, Eq. 4.42 becomes

$$\vec{c}_{n+1} = \vec{c}_n - \mu_n \left(H_W^{-1}\nabla E_W\right)_n, \tag{4.46}$$

which is exactly in the same form as Eq. 4.41. The sensitivity coefficient matrix G_W for wavelets is

$$G_W = G \cdot W^T = \frac{\partial \vec{d}}{\partial \vec{\alpha}} \cdot \frac{\partial \vec{\alpha}}{\partial \vec{c}} = \frac{\partial \vec{d}}{\partial \vec{c}}, \tag{4.47}$$

which can be calculated directly using the modified GPST method described in Section 3.3.3.

4.2.2 Covariance Operator of Wavelets

The posterior covariance operator for wavelet coefficients is

$$C_{W|D} = H_W^{-1} = \left(WHW^T\right)^{-1} = WH^{-1}W^T = WC_{M|D}W^T, \tag{4.48}$$

because W is an orthogonal matrix. In other words, $C_{W|D}$ is a similarity transformation of $C_{M|D}$, and a similarity transformation has several invariant variables, among which are the eigenvalues and the trace. Therefore, the total uncertainty of wavelets Σ_W equals the original uncertainty:

$$\Sigma_W = Trace(C_{W|D}) = Trace(WC_{M|D}W^T) = Trace(C_{W|D}) = \Sigma_M. \tag{4.49}$$

$C_{W|D}$ also has the same set of eigenvalues $\vec{\lambda}$ as $C_{M|D}$. Hence the wavelet transform does not change the overall uncertainty of the estimation. Apparently, the wavelet

transform does not do us any good if we use the full set of parameters for the inverse problem. However, in practice, we do not use the full set of parameters to model the spatial reservoir distribution. It is usually desirable to run the nonlinear regression with a reduced number of parameters, to have a faster procedure with higher stability. In this case, the wavelet analysis holds a huge advantage over the conventional approaches (pixel modeling and object modeling).

We can reorder the wavelet coefficients according to their variances, such that the p^{th} wavelet coefficient c_i has the p^{th} smallest variance, i.e., the p^{th} highest certainty. If we use the first p wavelet coefficients $\{c_i\}_{i=1,2...,p}$ to model the reservoir, the posterior covariance operator of wavelets is approximately

$$\tilde{C}_{W|D} = \tilde{H}_W^{-1} = \begin{bmatrix} H_{W,p}^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \tag{4.50}$$

where we assume H_W is in the following form:

$$H_W = \begin{bmatrix} H_{W,p} & B \\ B^T & A \end{bmatrix} \tag{4.51}$$

and $H_{W,p}$ is the $p \times p$ submatrix of H_W. We need to make one note here: $\tilde{C}_{W|D}$ is not the $p \times p$ submatrix of H_W^{-1}, because H_W is not a diagonal matrix. The inverse of H_W is

$$H_W^{-1} = \begin{bmatrix} \hat{H}_{W,p}^{-1} & \hat{B} \\ \hat{B}^T & \hat{A} \end{bmatrix}, \tag{4.52}$$

where

$$\hat{H}_{W,p}^{-1} = \left(H_{W,p} - B^T A^{-1} B\right)^{-1} \neq H_{W,p}^{-1}, \tag{4.53}$$

and $\hat{A}$ is a positive-definite matrix. The total uncertainty of the estimation with p parameters is

$$\begin{aligned} \tilde{\Sigma}_W(p) &= Trace(H_{W,p}^{-1}) \leq Trace(\left(H_{W,p} - B^T A^{-1} B\right)^{-1}) \\ &= Trace(\hat{H}_{W,p}^{-1}) = \Sigma_W - Trace(\hat{A}), \end{aligned} \tag{4.54}$$

where $Trace(\hat{A})$ is the summation of the variances for parameters $\{c_i\}_{i=p+1,N}$. Therefore, the total uncertainty of estimation with a smaller number of parameters is always

smaller than the original uncertainty:

$$\Sigma_W(p) < \Sigma_W, \text{ for } p < N. \tag{4.55}$$

Similarly, we can have the same conclusion for the estimation with the pixel parameters:

$$\tilde{\Sigma}_M(p) < \Sigma_M, \text{ for } p < N. \tag{4.56}$$

However, numerical experiments show that, for a small p, the following relationship holds:

$$\tilde{\Sigma}_W(p) \ll \tilde{\Sigma}_M(p). \tag{4.57}$$

This is not unexpected. Mathematically, the wavelet coefficients are linear combinations of the original model parameters $\vec{\alpha}$. Some of the wavelet coefficients are low-resolution spatial averages of the model parameters, the others are local variations at different resolution levels. In a situation in which the information does not have sufficient resolving power, especially where there is not much spatial information, the lower-resolution spatial average will be better determined. For example, it is very difficult to resolve the properties of an individual gridblock far away from any wells, because most of the information is concentrated around wellbores, e.g., bottomhole pressures, watercuts, etc. On the other hand, the certainty of estimation for the spatial property average of gridblocks around that particular gridblock would be much higher.

4.2.3 Comparison Between Pixel and Wavelet Modeling

In this section, we will use a simple numerical experiment to demonstrate the advantage of the wavelet approach over the pixel modeling. The test reservoir is similar to the one shown in Section 3.2.4, with some slight difference. We used a 16×16 grid (totally 256 gridblocks) instead of a 32×32 grid. It is easier for us to show the covariance matrix with a coarser grid. In this case, each gridblock is $100 \times 100 \times 100$ ft^3. There are four wells located at corners, among which Well #2 (the well in the left-lower corner) is the injecting well, the others are producing. The water injection

rate is fixed at 10000 STB/D, and the total production rates ($Q_w + Q_o$) at the producing wells are fixed at 2000 STB/D, 2000 STB/D and 6000 STB/D respectively. The permeability distribution is as depicted in Fig. 4.1: a high-permeability strip with $k = 2500$ md is located in the middle of the reservoir, while the rest of the reservoir has $k = 500$ md. The initial porosity is uniform at $\phi^0 = 0.22$.

In this example, we assumed no *a priori* information was available, i.e., the variance of logarithm permeability field is very large, $\sigma_{\ln k} = 10$. There are three different types of dynamic data available: the bottomhole pressure in each well, watercut at each producing well (shown in Figs. 4.3 through 4.6), and water saturation change in the time period between $t = 100$ days and $t = 200$ days (shown in Fig. 4.2). We also assumed that the standard deviations for the pressure, watercut and saturation change data are 1.0 psi, 0.003 and 0.06 respectively. There are no correlations between data points.

We used two different choices of reservoir model parameters: (a) gridblock log-permeability, $\ln k$, (b) Haar wavelet coefficients of log-permeability, $c_{\ln k}$. For both cases, the posterior covariance matrix $C_{M|D}$ ($C_{W|D}$ for wavelets) was calculated using the Hessian matrix at the maximum likelihood point, i.e.,

$$C_{M|D} = H^{-1}, \tag{4.58}$$

$$C_{W|D} = H_W^{-1}. \tag{4.59}$$

Figs. 4.7 and 4.8 show the covariance matrices for the pixel modeling and the wavelet modeling, respectively. There are many off-diagonal nonzero elements in both matrices, though with different patterns. In this particular case, there are no off-diagonal elements in the prior covariance matrix, therefore the posterior covariance matrix does not show strong correlations between parameters, except for those close by. The off-diagonal nonzeros in $C_{M|D}$ are not concentrated, except around the diagonal. $C_{W|D}$ has most of the significant off-diagonal elements clustered in small groups. In this particular case, there are two significant off-diagonal bands in $C_{W|D}$, showing correlation patterns at different resolution levels.

As shown in Section 4.2.2, $Trace(C_{M|D}) = Trace(C_{W|D})$, i.e., the summation of diagonal elements for the two covariance matrices are equivalent. However, the

distribution of the diagonal elements are different. In $C_{M|D}$, most of the diagonal elements are significant, with only a few small elements. In $C_{W|D}$, there are some clusters of small diagonal elements. In other words, the diagonal of $C_{M|D}$ is more uniform than that of $C_{W|D}$.

Figs. 4.9 and 4.10 show the variances of each parameter in the two approaches. Fig. 4.9 shows that the uncertainty of estimation is very high except for the regions nearby wellbores, i.e., we can not have a reliable estimation for most of the reservoir, at the gridblock level. Fig. 4.10 shows that, for those regions far away from any wellbores, the estimation of lower-resolution averages (linear combinations) is with much higher certainty than the individual gridblocks (the upper-left region in Fig. 4.10 represents the lower-resolution parameters). Therefore, in addition to wellbore neighborhoods, we can have a better estimation of the rest of the reservoir, though at a lower-resolution level. The concept is easy to understand: the well response is not sensitive to local variations in a region far away from the well. However, the well might be sensitive to overall property averages within such distant regions.

Fig. 4.11 shows the posterior variance of each parameter in order of magnitude. This figure shows clearly the difference in uncertainty distribution between the two approaches: the pixel approach yields very small variances (high certainty) for a small set of parameters (wellblocks, mostly); the wavelet approach has a higher resolving power for a much larger set of parameters (wellbore neighbors plus some lower-resolution averages). For the rest of the parameters, both approaches give an estimation variance close to the prior variance, that is, the inverse procedure does not improve the certainty of those parameters very much. Fig. 4.12 shows the inverse of parameter variances (certainty, in some sense), which demonstrates a similar trend.

Fig. 4.13 shows the total uncertainty $\tilde{\Sigma}(p)$ (trace of the covariance matrix) of estimation for the two approaches. Obviously, there is not much difference when there are many parameters in the inverse procedure. Furthermore, if we use the full set of parameters, the total uncertainty should be the same for the two approaches, as shown in Eq. 4.49. However, as shown in Figs. 4.11 and 4.12, we can only resolve a small set of parameters with high certainty. Fig. 4.14 shows that the total uncertainty of the wavelet approach is much lower than that of the pixel approach, when a small

number of parameters is used. For instance, for $p = 60$ (using 60 parameters for the inverse procedure), the total uncertainty of the wavelet approach is 7.15, which is much lower than that of the pixel approach (26.59).

In summary, the wavelet approach holds huge advantages over the pixel approach in the following aspects:

- The pixel approach estimates the parameters with a uniform grid, though each parameter might be with different variance. The wavelet approach estimates the parameters at multiple resolutions, and the different resolution level will reflect different scales of information provided by production data.

- Because the wavelet approach uses multiresolution analysis, it is possible for us to obtain a nonuniform resolution in the parameter estimation. That is, in regions without enough information (therefore, not much resolving power), we can have an estimation with sufficient certainty at a lower resolution. At the same time, we do not need to give up the high resolution at the other regions, e.g., wellbore neighborhoods, strong geological discontinuities, etc.

- The wavelet approach can reproduce the reservoir property distribution with a smaller set of parameters. Therefore, the inverse procedure would be much more efficient and stable. Because the calculation of sensitivity is a major part of the total computational load, the computational time will be dramatically reduced with a smaller set of parameters.

4.3 Impact of Additional Information

In practice, it is always desirable to have as much information as possible. More information usually means that there are more constraints on the model parameters, no matter if the information is dynamic or static data. More constraints can make an inverse problem better determined. In this section, we will explore the impact of different information on reservoir parameter estimation.

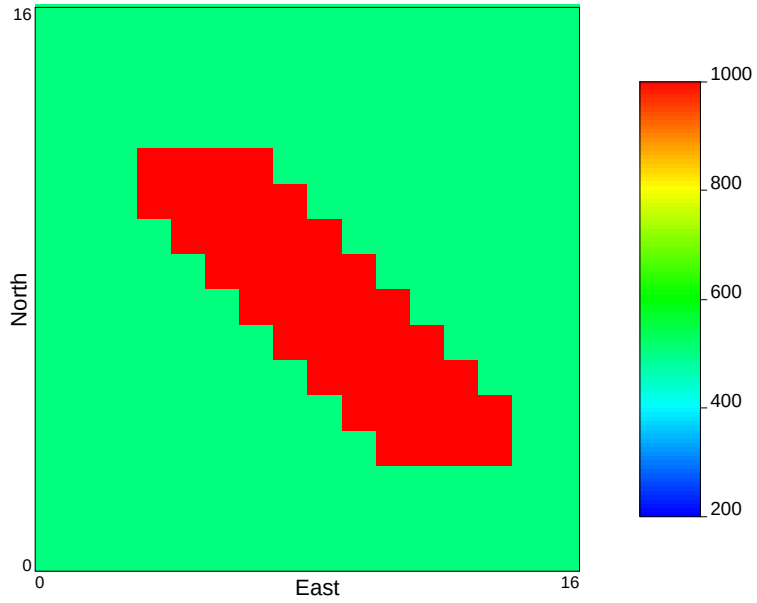

Figure 4.1: Reservoir permeability distribution

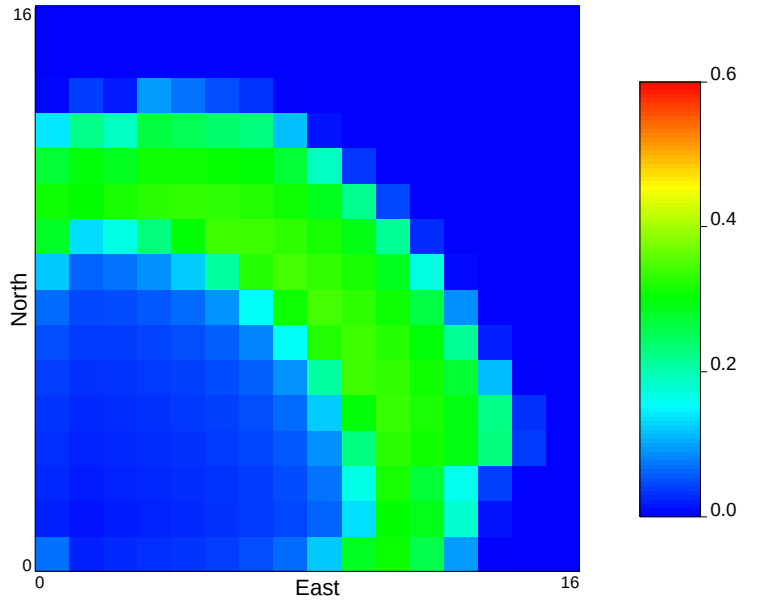

Figure 4.2: Water saturation change distribution during time t=100,200 days

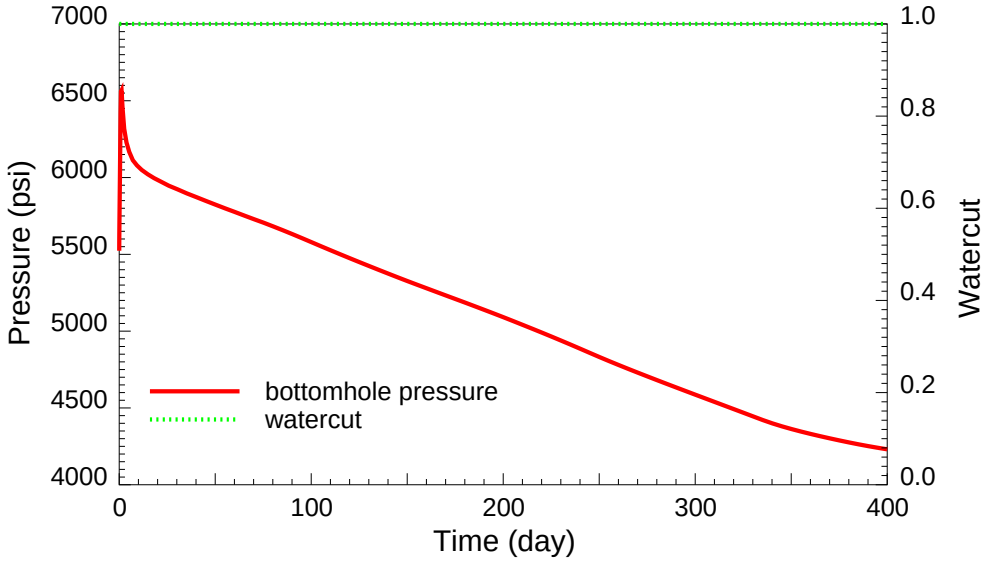

Figure 4.3: Bottomhole pressure and watercut at Well #1

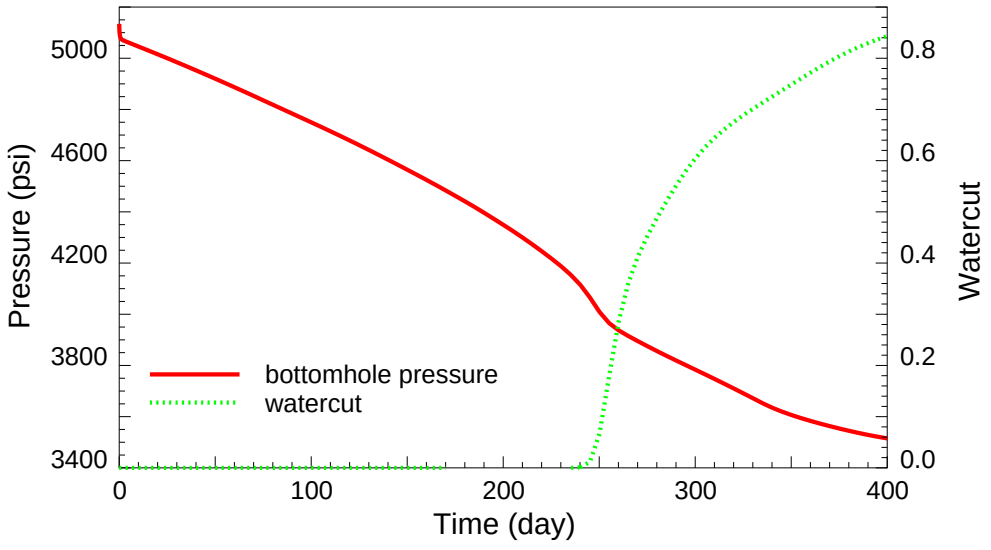

Figure 4.4: Bottomhole pressure and watercut at Well #2

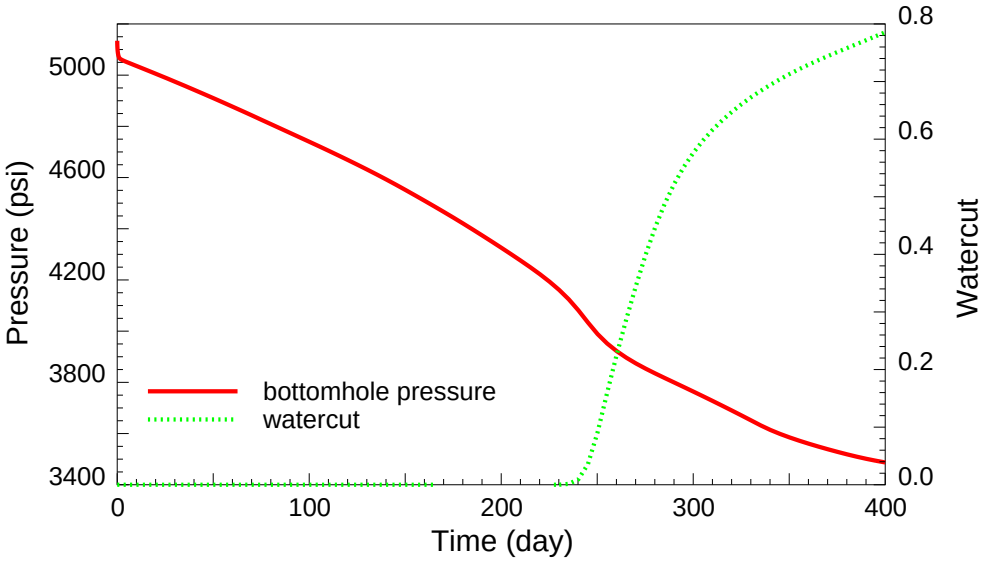

Figure 4.5: Bottomhole pressure and watercut at Well #3

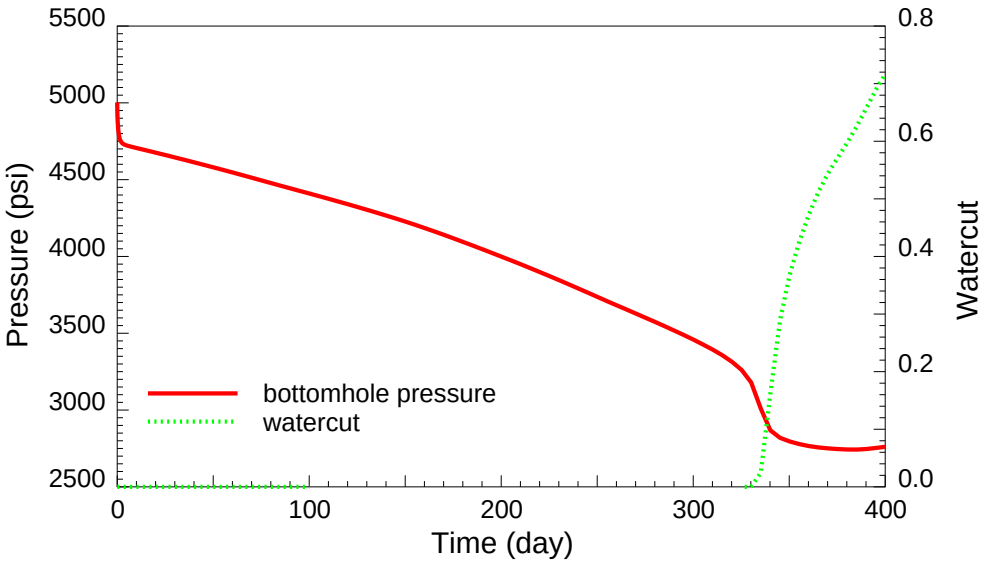

Figure 4.6: Bottomhole pressure and watercut at Well #4

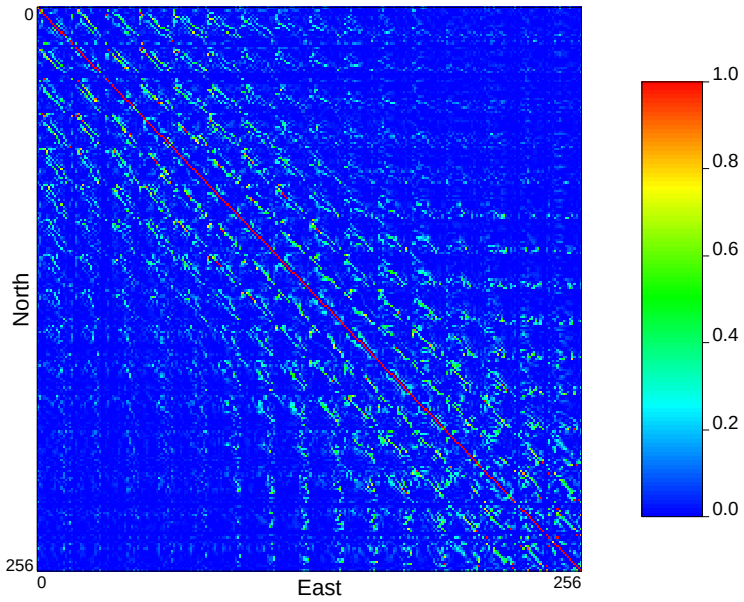

Figure 4.7: Covariance matrix for pixel modeling (256 parameters), $C_{M|D}$

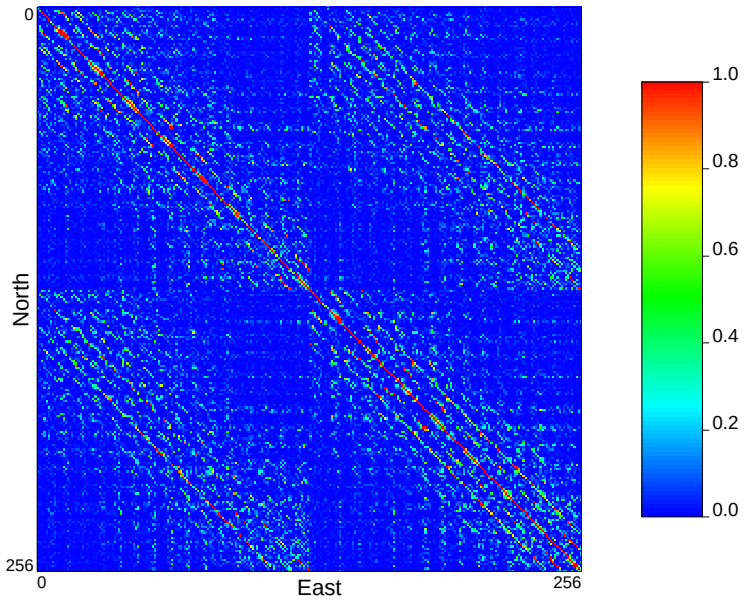

Figure 4.8: Covariance matrix for wavelet modeling (256 parameters), $C_{W|D}$

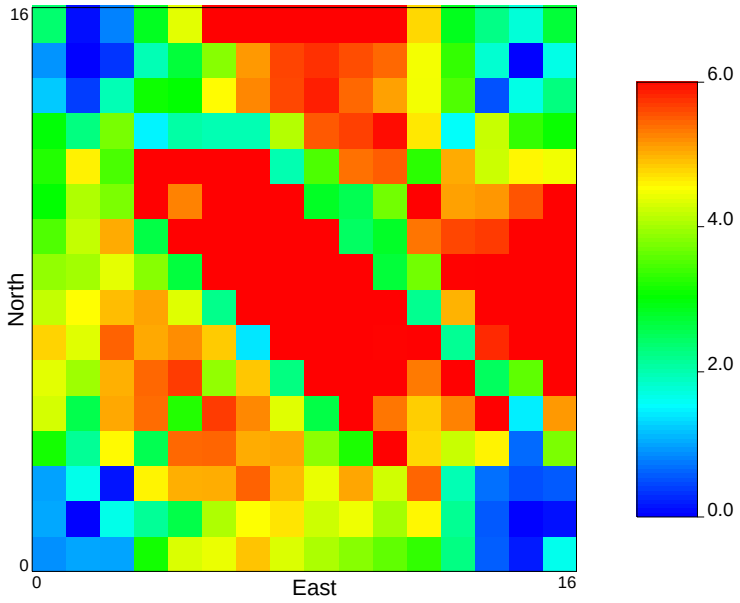

Figure 4.9: Variance distribution for pixel modeling (256 parameters)

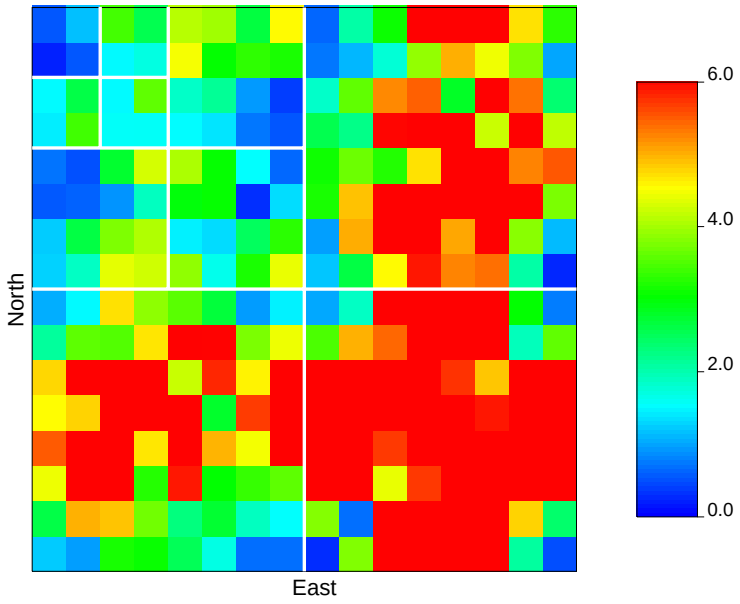

Figure 4.10: Variance distribution for wavelet modeling (256 parameters)

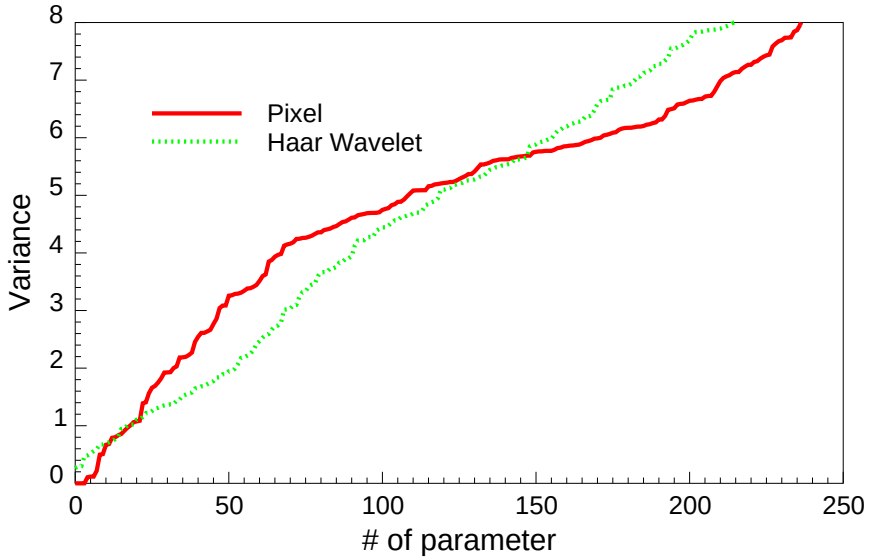

Figure 4.11: Variances of parameters for both pixel and wavelet modeling, ordered according to the magnitude

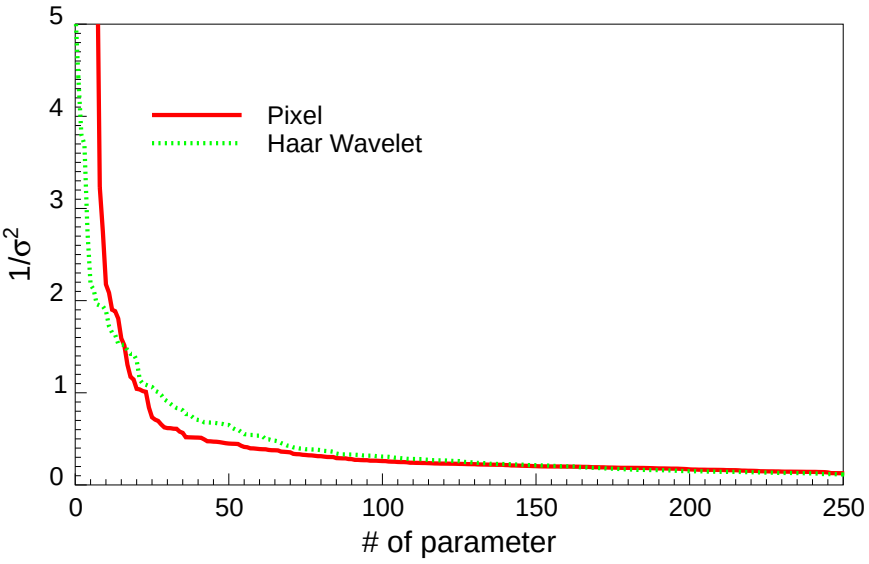

Figure 4.12: Estimation certainty for each parameter, ordered according to the magnitude

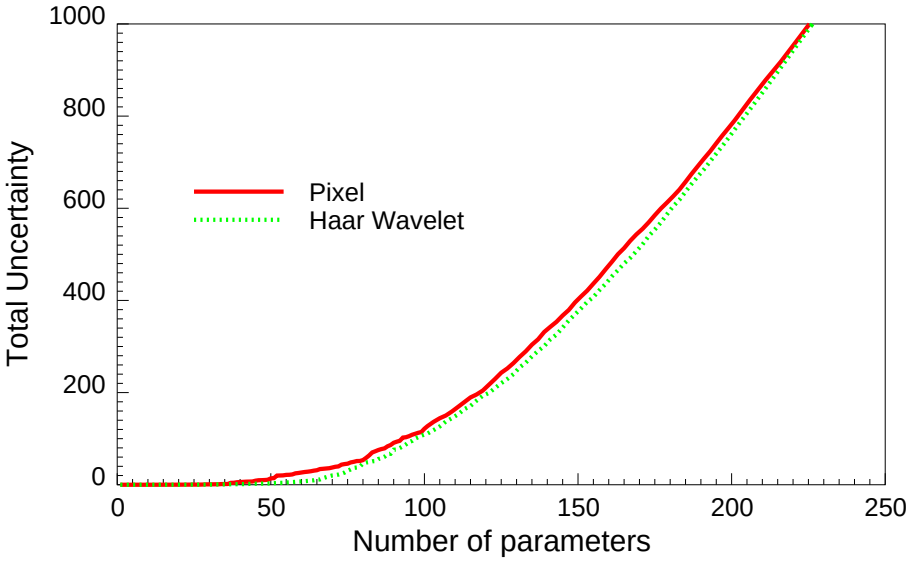

Figure 4.13: Trade-off curves for both pixel and wavelet modeling

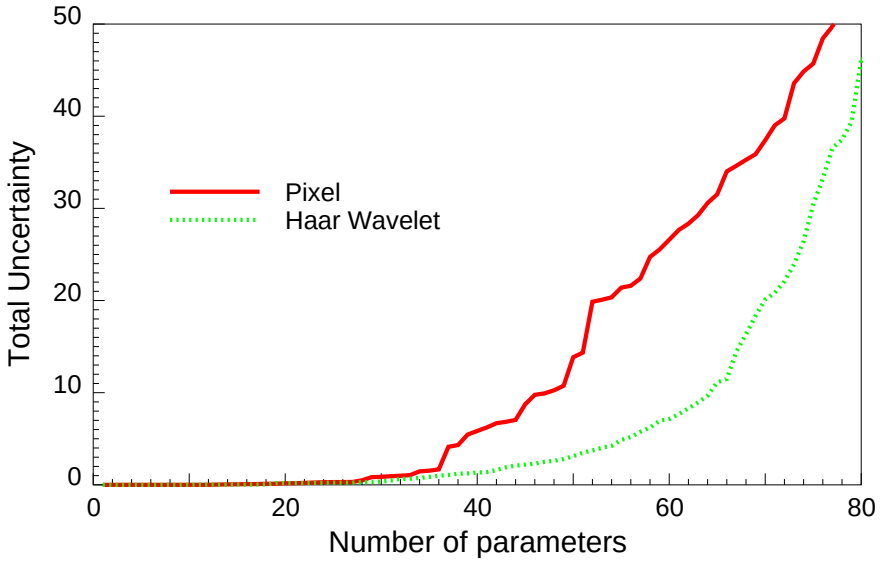

Figure 4.14: Trade-off curves for both pixel and wavelet modeling

4.3.1 Additional Dynamic Data

Mathematical Analysis

In the process of reservoir characterization, we may have additional dynamic data available. For example, a newly-drilled well can give us some production data, and a 4-D seismic survey can provide some spatial information. Let us denote the new dynamic data set as $\vec{d}^{new}$, the total reservoir response becomes

$$\vec{d^{obs}} = \begin{bmatrix} \vec{d}^{old} \\ \vec{d}^{new} \end{bmatrix}. \tag{4.60}$$

The sensitivity coefficient matrix is

$$\tilde{G} = \begin{bmatrix} \partial\vec{d}^{old}/\partial\vec{\alpha} \\ \partial\vec{d}^{new}/\partial\vec{\alpha} \end{bmatrix} = \begin{bmatrix} G \\ \hat{G} \end{bmatrix}. \tag{4.61}$$

In this section, we assume the data has been scaled by its covariance, i.e.,

$$\vec{d}^{scale} = C_D^{-1/2}\vec{d}, \tag{4.62}$$

and we will omit the superscript *scale* in this section. Thus the covariance matrix for the data is

$$\tilde{C}_D = \begin{bmatrix} I & P \\ P^T & I \end{bmatrix}, \tag{4.63}$$

where P is the covariance matrix between $\vec{d}^{old}$ and $\vec{d}^{new}$. It can be shown that

$$\begin{aligned} \tilde{C}_D^{-1} &= \begin{bmatrix} I & P \\ P^T & I \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \left(I - PP^T\right)^{-1} & -P\left(I - P^TP\right)^{-1} \\ -\left(I - P^TP\right)^{-1}P^T & \left(I - P^TP\right)^{-1} \end{bmatrix} \qquad (4.64) \\ &= \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} P\left(I - P^TP\right)^{-1}P^T & -P\left(I - P^TP\right)^{-1} \\ -\left(I - P^TP\right)^{-1}P^T & \left(I - P^TP\right)^{-1} \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -P \\ I \end{bmatrix} \cdot \left(I - P^TP\right)^{-1} \cdot \begin{bmatrix} -P^T & I \end{bmatrix}, \qquad (4.65) \end{aligned}$$

where a relationship holds

$$\left(I - PP^T\right)^{-1} = I + P\left(I - P^T P\right)^{-1} P^T. \tag{4.66}$$

We should also notice that $\left(I - P^T P\right)^{-1}$, the right-lower submatrix in Eq. 4.64, is positive-definite, because $\tilde{C}_D$ is positive-definite.

Eq. 4.65 yields a new Hessian matrix

$$\begin{aligned} \tilde{H} &= \tilde{G}^T \tilde{C}_D^{-1} \tilde{G} + C_M^{-1} \\ &= G^T G + \left(\hat{G}^T - G^T P\right)\left(I - P^T P\right)^{-1}\left(\hat{G} - P^T G\right) + C_M^{-1} \\ &= H + \check{G}^T \check{C}_D^{-1} \check{G}, \end{aligned} \tag{4.67}$$

where

$$\check{G} = \hat{G} - P^T G, \tag{4.68}$$

$$\check{C}_D = \left(I - P^T P\right), \tag{4.69}$$

which are the effective sensitivity coefficient matrix and the effective data covariance matrix for the additional data.

The posterior covariance matrix for the estimation becomes

$$\begin{aligned} \tilde{C}_{M|D} &= \tilde{H}^{-1} = \left(H + \check{G}^T \check{C}_D^{-1} \check{G}\right)^{-1} \\ &= H^{-1} - H^{-1}\check{G}^T\left(\check{G} H \check{G}^T + \check{C}_D^{-1}\right)^{-1} \check{G} H^{-1} \\ &= C_{M|D} - \check{C}_{M|D}, \end{aligned} \tag{4.70}$$

where

$$\check{C}_{M|D} = H^{-1}\check{G}^T\left(\check{G} H \check{G}^T + \check{C}_D^{-1}\right)^{-1} \check{G} H^{-1} \tag{4.71}$$

is a positive-definite matrix. All the diagonal elements of $\check{C}_{M|D}$ are positive, thus Eq. 4.70 yields

$$\tilde{C}_{M|D,ii} < C_{M|D,ii}, \text{ for } i = 1, 2, ..., nPar, \tag{4.72}$$

which means the variances of estimation for all the parameters are reduced with the new dynamic data. Therefore, as long as the data are properly weighted, additional

dynamic information (more constraints) always stabilizes the inverse problem and gives parameter estimation with a higher certainty.

Eqs. 4.70 and 4.71 show that the impact of the additional dynamic data depends on several aspects:

- *Effective sensitivity*: Eq. 4.68 gives the definition of the effective sensitivity of the new data: $\check{G} = \hat{G} - P^T G$. Because $G = \partial \vec{d}^{obs}/\partial \vec{\alpha}$, we can impose a small perturbation $\delta\vec{\alpha}$ on the parameter set $\vec{\alpha}$ and $\delta\vec{d} = G\delta\vec{\alpha}$. The effective response due to the new data is

$$\delta \vec{d}^{eff} = \check{G}\delta\vec{\alpha} = (\hat{G} - P^T G)\delta\vec{\alpha} = \delta\vec{d}^{new} - P^T \delta\vec{d}^{old}. \tag{4.73}$$

  This equation shows clearly that if there is strong correlation between $\vec{d}^{old}$ and $\vec{d}^{new}$, the effective response $\delta\vec{d}^{eff}$ would be much less significant than the real data response $\delta\vec{d}^{new}$.

- *Effective covariance*: Eq. 4.69 defines the effective covariance operator for the new data as $\check{C}_D = \left(I - P^T P\right)$. $\check{C}_D$ shows that if the correlation between the two data sets is high, the effective data covariance will be small. In other words, the data correlation has an opposite impact on the effective covariance.

- *The resolving power of the new data*: Even if there is no correlation between the two set of data, $P = 0$, the estimation still depends on the resolving power of the new data. In this case, Eq. 4.71 becomes

$$\check{C}_{M|D} = H^{-1}\hat{G}^T \left(\hat{G}H\hat{G}^T + I\right)^{-1} \hat{G}H^{-1}. \tag{4.74}$$

  In practice, we usually assume that there is no correlation between different sets of data. For example, the correlation can be neglected between different data types at the well (e.g., bottomhole pressure and watercut), between the same data types at different wells, or between different data types at different wells. In most cases, Eq. 4.74 is a very good approximation.

An Example with Two Data Points

As shown in the previous section, it is not very easy to define the impact of the data correlation, i.e., it is not very straightforward for us to determine whether the correlation reduces the estimation uncertainty. In this section, we will use a very simple example to show the impact of the data correlation. In this particular case, there is only one model parameter α without any prior information. We originally had only one data point d_1 with a variance of 1 and a sensitivity of 1. Therefore, the estimation covariance $C_{M|d_1} = 1$. Now we obtained one more data point d_2 also with a variance of 1, but a sensitivity of g. The correlation between d_1 and d_2 is ρ. The final estimation covariance is

$$C_{M|D=d_1+d_2} = \left(\begin{bmatrix} 1 & g \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ g \end{bmatrix} \right)^{-1} = \frac{1-\rho^2}{1-2\rho g+g^2}, \tag{4.75}$$

which is a function of ρ and g.

Fig. 4.15 shows the estimation variance as a function of ρ, with different g values. In this example, when $g = 1.0$, the variance increases monotonically as ρ increases. When $g = -1.0$, the variance decreases as ρ increases. With other g values, the variance does not change monotonically. Fig. 4.16 shows the estimation variance as a function of g at different ρ values. The estimation variance reaches a maximum when $\rho = g$. This example shows that the data correlation is not an isolated factor. It is not always true to state that strongly correlated data provide less new information. However, if there is strong correlation between two data sets, it would be safe to assume that the sensitivity would be in phase with the correlation. That is, the following relationship holds in linear cases

$$\hat{G} \approx P^T G. \tag{4.76}$$

Therefore, in most practical cases, strongly correlated data usually provide little new information.

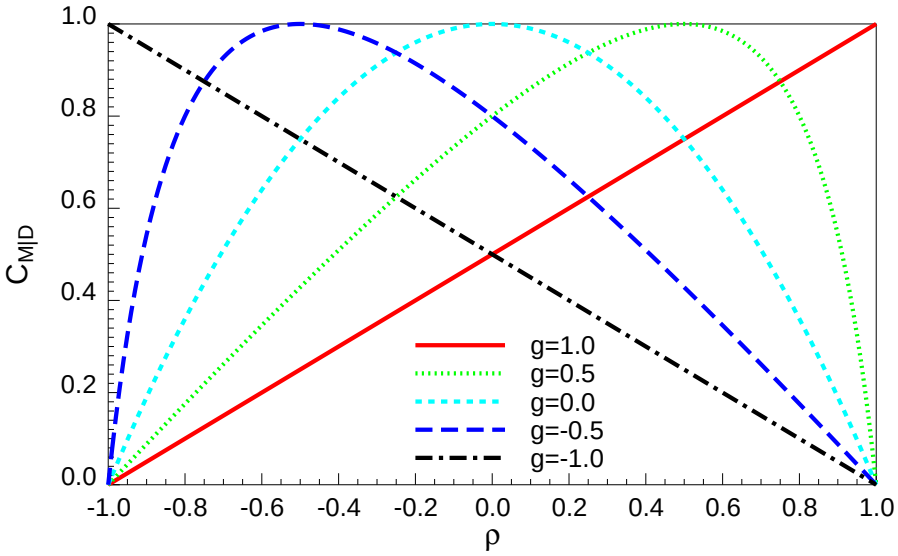

Figure 4.15: Relation between the data correlation and the estimation variance

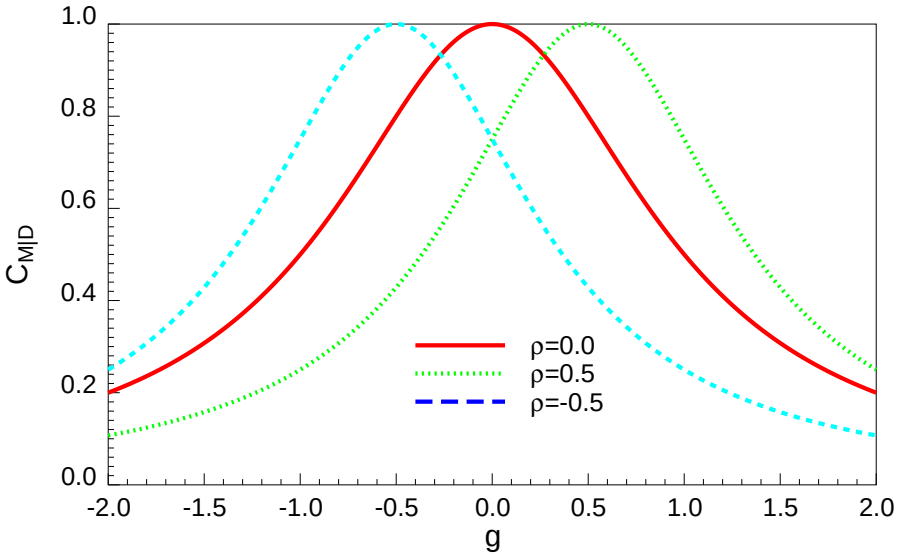

Figure 4.16: Relation between the sensitivity and the estimation variance

Numerical Experiment

We used the same test reservoir shown in Section 4.2.3 to explore the impact of different data on the parameter estimation. The reservoir is as depicted in Fig. 4.1. We will look into the variance distribution for four different data settings: 10-day bottomhole pressure history, 250-day bottomhole pressure history, 400-day production history with bottomhole pressures and watercuts, 400-day production history with some information for water saturation change. We assumed that there is no correlation between any two different data sets. In each setting, we used two choices of model parameters: gridblock logarithm permeability ($\ln k$) and wavelet coefficients ($c_{\ln k}$).

- Figs. 4.17 and 4.18 show the variance of estimation for the inverse procedure with only short-term data. In this case, the dynamic data are too scarce to resolve the model parameters, except for the parameters around wellbores. The wavelet analysis does not help much in this case.

- Figs. 4.19 and 4.20 show that long-term bottomhole pressure data can reduce the uncertainty of estimation. The pixel approach can resolve the model parameters around the injecting well (the well at the left-lower corner). Because water has not broken through yet in this case, the inverse procedure can resolve only a few parameters around the three producing wells. The middle of the reservoir is left unresolved. The wavelet approach performed much better than the pixel approach here. Instead of resolving the individual pixels, the wavelet approach was able to resolve the low-certainty regions with lower-resolution averages. The high-resolution local variations for these regions will be not resolved.

- Figs. 4.21 and 4.22 show that the watercut information has a large impact on the parameter estimation. Both the pixel approach and the wavelet approach can resolve most of the model parameters around the wellbore neighborhoods. Because both the pressure and watercut do not have enough spatial resolving power, the pixel approach can not resolve the region in middle of the reservoir. The wavelet approach can resolve almost all the low-resolution parameters with

very high certainty.

- Figs. 4.23 and 4.24 show the effect of water saturation change information. Though being interpreted from 4-D seismic surveys and with very high uncertainty, this information provides the spatial resolving power lacking in both the pressure and watercut data. Some parameters in middle of the reservoir are better determined.

Figs. 4.25 and 4.26 show the trade-off curves for different data settings, with pixel modeling and wavelet modeling respectively. As new data are added (longer production history, watercut, 4-D seismic data, etc.), the total uncertainty of the estimation is reduced. The impact of additional data is more dramatic for the wavelet approach. Instead of estimating the uniform-grid pixels with varying uncertainty, the wavelet approach can estimate a set of model parameters with lower uncertainty at different resolution levels.

Figs. 4.27 and 4.28 show the average uncertainty as a function of the number of parameters. The average uncertainty is defined as the total uncertainty divided by the number of parameters. The average uncertainty is an indicator of how well an inverse procedure can resolve the model parameters. As shown in Figs. 4.27 and 4.28, the average uncertainty for the wavelet approach increases much slower than that for the pixel approach, especially in the case where there is not much spatial information.

4.3.2 More Accurate Measurement

If we can reduce the measurement errors of the observed data, the variances of estimation will be reduced. In other words, the uncertainty of the estimation is reduced. However, there are two factors that limit the impact of the measurements:

- As long as the assumption of Gaussian distribution holds, the covariance operator of the data C_D is the summation of the measurement errors C_d and the model errors C_T:

$$C_D = C_d + C_T. \tag{4.77}$$

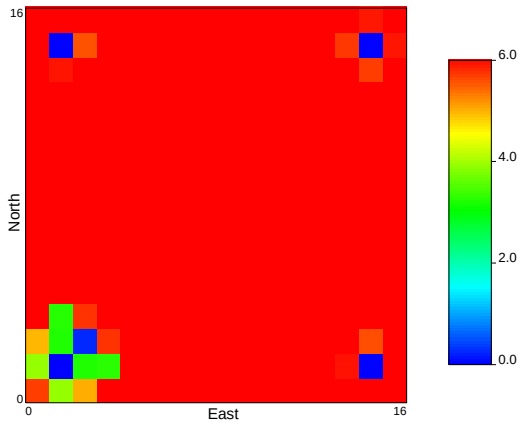

Figure 4.17: Variances of pixels ($t = 10$ days)

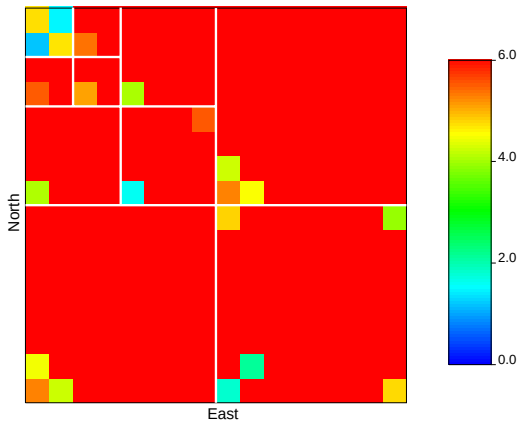

Figure 4.18: Variances of wavelets ($t = 10$ days)

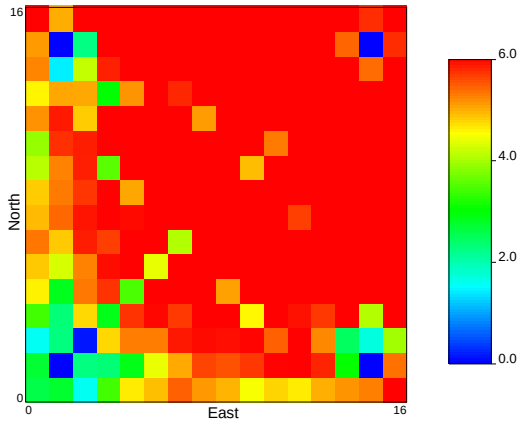

Figure 4.19: Variances of pixels ($t = 250$ days)

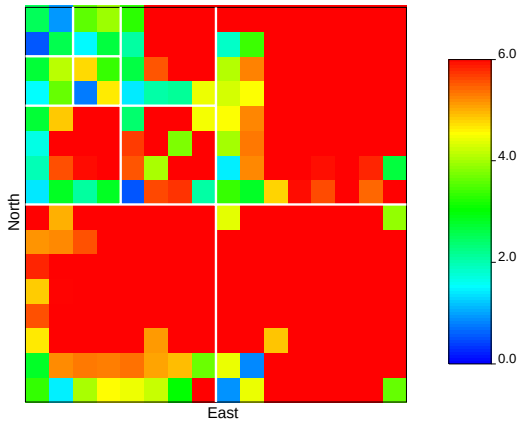

Figure 4.20: Variances of wavelets ($t = 250$ days)

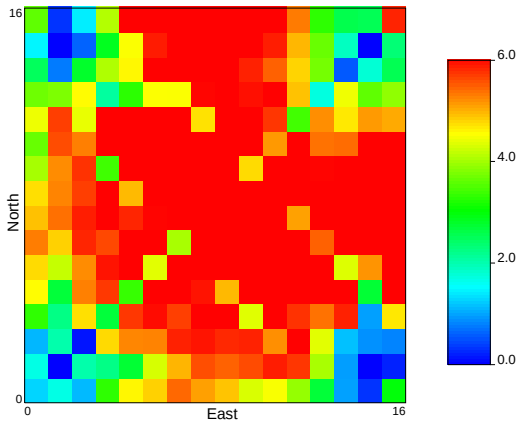

Figure 4.21: Variances of pixels ($t = 400$ days)

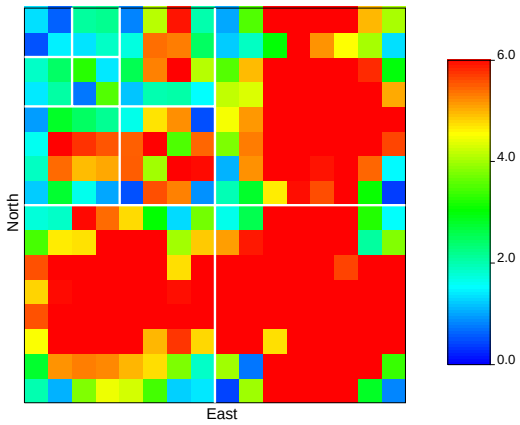

Figure 4.22: Variances of wavelets ($t = 400$ days)

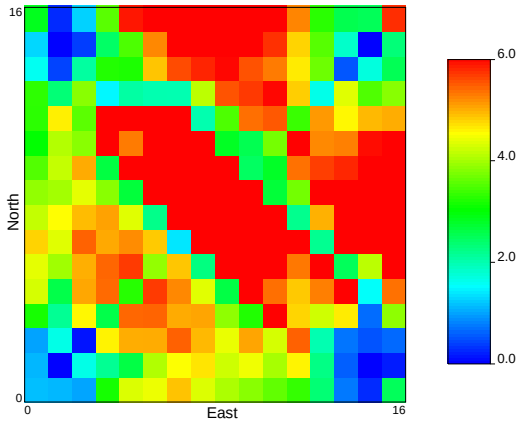

Figure 4.23: Variances of pixels ($t = 400$ days), with 4-D seismic data

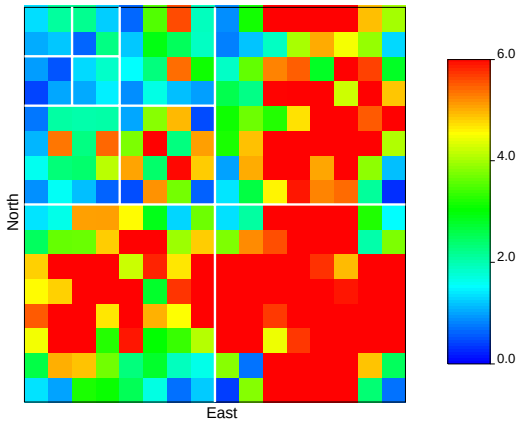

Figure 4.24: Variances of wavelets ($t = 400$ days), with 4-D seismic data

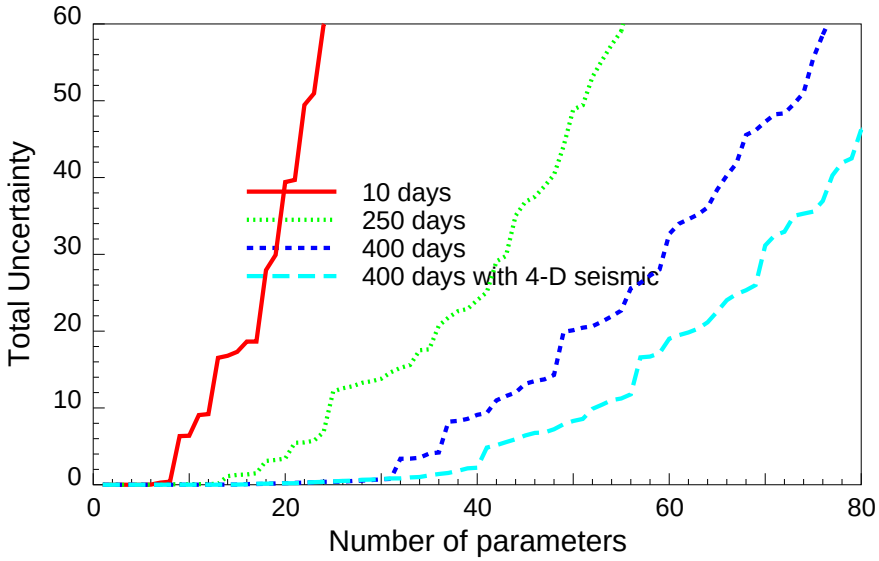

Figure 4.25: Trade-off curves for pixel modeling, with different data sets

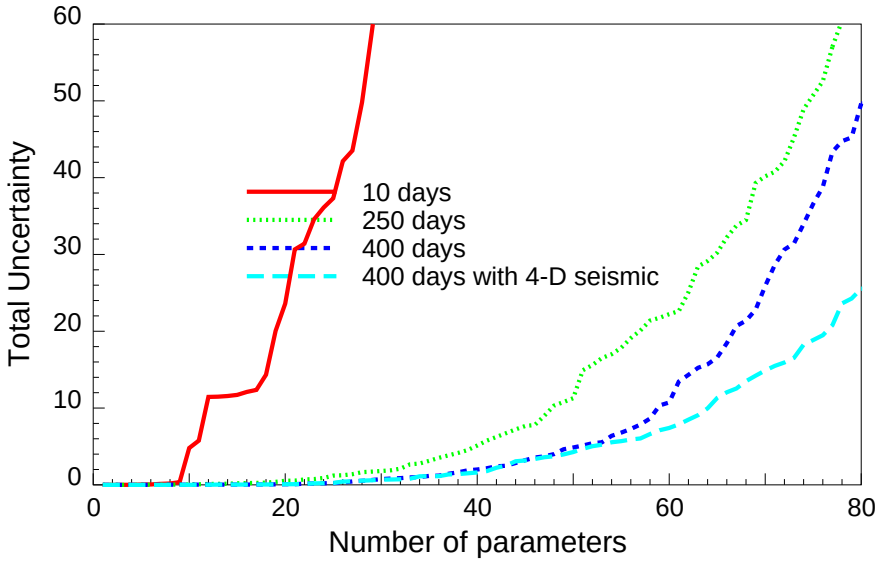

Figure 4.26: Trade-off curves for wavelet modeling, with different data sets

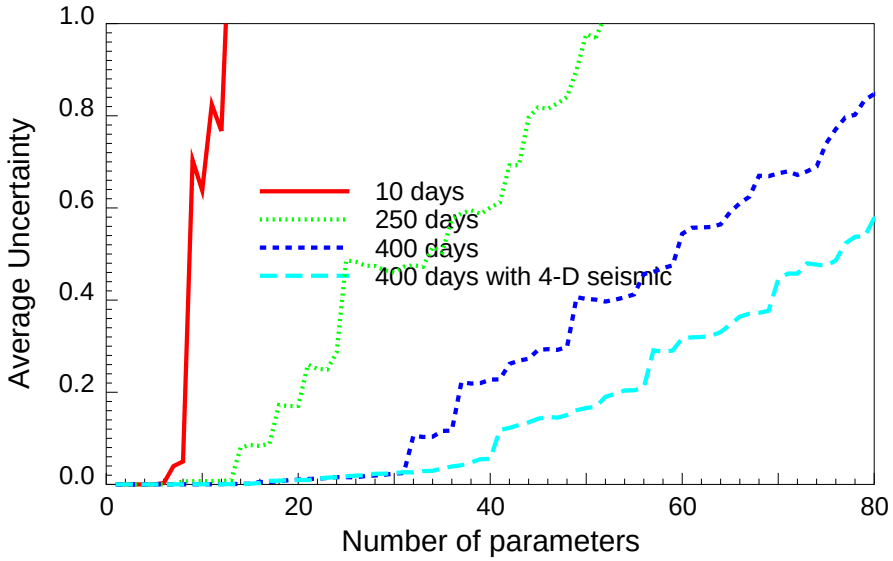

Figure 4.27: Trade-off curves (average uncertainty) for pixel modeling, with different data sets

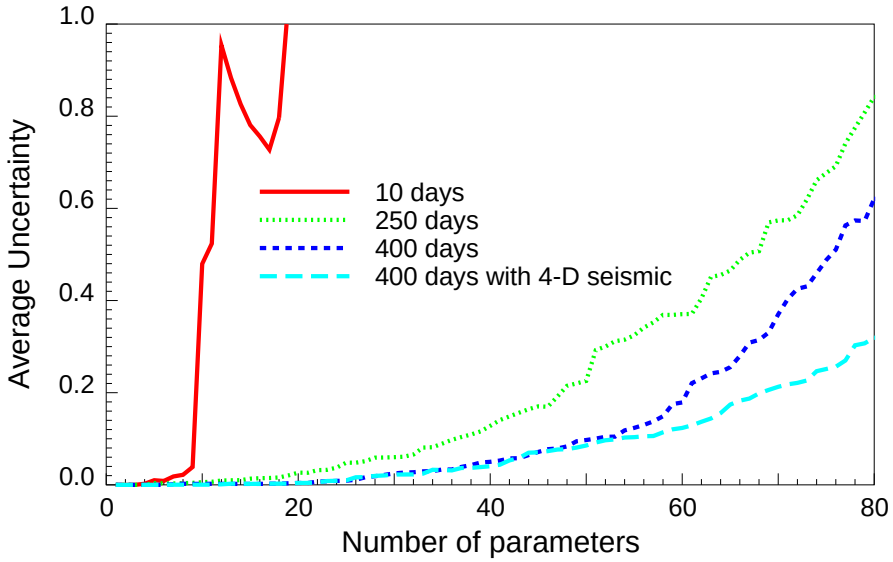

Figure 4.28: Trade-off curves (average uncertainty) for wavelet modeling, with different data sets

Therefore, even if we can reduce the measurement errors to some extent, C_D is still bounded by the model errors C_T, which can not be reduced unless we use different mathematical and numerical models. Furthermore, the accuracy of measurements has its own limit and can not be reduced infinitely.

- With a smaller C_D, the unresolved parameters will remain unresolved. We will address this issue in this section.

We can use a special case to investigate the impact of the accuracy of measurements. As shown in Section 4.1.3, if $C_M = \sigma_M^2 I$, the covariance of estimation is

$$C_{M|D} = H^{-1} = V \cdot diag\left(\frac{1}{\lambda_i^2 + 1/\sigma_M^2}\right) \cdot V^T, \tag{4.78}$$

where

$$H = G^T C_D^{-1} G + C_M^{-1}. \tag{4.79}$$

If C_D is reduced by a factor of θ, i.e., $\tilde{C}_D = \theta C_D$, the covariance of estimation becomes

$$\tilde{C}_{M|D} = \tilde{H}^{-1} = V \cdot diag\left(\frac{1}{\lambda_i^2/\theta + 1/\sigma_M^2}\right) \cdot V^T. \tag{4.80}$$

Let us consider three different cases:

- *The parameter is well resolved*: In this case, $\lambda_i \gg 1/\sigma_M$, and

$$\frac{1}{\lambda_i^2/\theta + 1/\sigma_M^2} \approx \theta \cdot \frac{1}{\lambda_i^2 + 1/\sigma_M^2}. \tag{4.81}$$

  Therefore, if a parameter is well resolved, then the more accurate measurement will make the uncertainty even lower. However, this does not help much, because we already have high-certainty estimation for that parameter.

- *The parameter is partially resolved*: In this case, λ_i should be roughly of the same order as $1/\sigma_M$. We have

$$\frac{1}{\lambda_i^2/\theta + 1/\sigma_M^2} < \frac{1}{\lambda_i^2 + 1/\sigma_M^2}. \tag{4.82}$$

  If a parameter is partially resolved, then the more accurate measurement will make this parameter better resolved.

- *The parameter is not resolved at all*: In this case, $\lambda_i \ll 1/\sigma_M$, and

$$\frac{1}{\lambda_i^2/\theta + 1/\sigma_M^2} \approx \frac{1}{\lambda_i^2 + 1/\sigma_M^2}. \tag{4.83}$$

  Therefore, if a parameter is not resolved, then the more accurate measurement will not make the estimation any better. In reservoir parameter estimation problems, the observed data usually do not have enough resolving power to resolve all the parameters. Many parameters are left unresolved, even with more accurate observed data.

In summary, the higher accuracy in observed data only affects the parameters that are already partially resolved. The partially-resolved parameters are only a portion of all the parameters, as shown in the following numerical experiment.

We used the same test case as in Section 4.3.1 to demonstrate the impact of measurement accuracy. The original case is with variances of 1.0 psi^2, 1×10^{-5} and 0.004 for pressure, watercut and saturation change. In this case, we assumed smaller variances for the pressure and watercut data, at 0.5 psi^2 and 5×10^{-6} respectively. Fig. 4.29 shows the variance distribution for the original case. Fig. 4.30 shows the variance distribution with reduced data variances. Obviously, there is not much difference between these two cases. With reduced data variances, the variance of estimation is reduced for only a small portion of the parameters.

Fig. 4.31 shows the singular values for the posterior covariance matrices, both for the original case and the reduced-variance case. The singular values are defined in

$$C_{M|D} = V \cdot diag\,(\nu_i) \cdot V^T, \tag{4.84}$$

where ν_i are ordered according to the magnitude. As we can see, the higher accuracy reduces most of the singular values. However, for very small singular values, any further reduction in magnitude does not make any difference, because the values are already small. For large singular values, the reduction in magnitude is too small. The effect of the higher accuracy in observed data is only visible for a small number of parameters.

Fig. 4.32 shows the trade-off curves for the original case and the reduced-variance case. The total uncertainty with the p smallest singular values is

$$\Sigma(p) = \sum_{i=1}^{p} \nu_i. \tag{4.85}$$

With higher-accuracy data, the reduction in the total uncertainty is negligible for most of the p values.

4.3.3 Additional Static Information

Parameter estimation based purely on the dynamic data may suffer from instability and nonuniqueness. Furthermore, the resulting distributions may not capture the major geological structures and yields unrealistic predictions. In this work we used two different strategies to deal with static information. The first choice is to treat static information as prior information, i.e., we can obtain a realization of model parameters prior to the parameter estimation procedure. The other choice is to integrate static information directly into the parameter estimation procedure. We will discuss the second strategy in Chapter 5. In this section, we focus on the prior information.

The posterior covariance operator $C_{M|D}$ is

$$C_{M|D} = \left(G^T C_D^{-1} G + C_M^{-1}\right)^{-1}, \tag{4.86}$$

where C_M is the covariance operator of model parameters based on prior information. Let us denote $\tilde{C}_{M|D}$ as the posterior covariance operator without any prior information:

$$\tilde{C}_{M|D} = \left(G^T C_D^{-1} G\right), \tag{4.87}$$

it can be shown that

$$C_{M|D} = \left(\tilde{G}_{M|D}^{-1} + C_M^{-1}\right)^{-1} = \tilde{C}_{M|D} - \tilde{C}_{M|D}\left(C_M + \tilde{C}_{M|D}\right)^{-1}\tilde{C}_{M|D}. \tag{4.88}$$

Because the second term on the right-hand side of Eq. 4.88 is a positive-definite matrix, we have

$$Trace(C_{M|D}) < Trace(\tilde{C}_{M|D}), \tag{4.89}$$

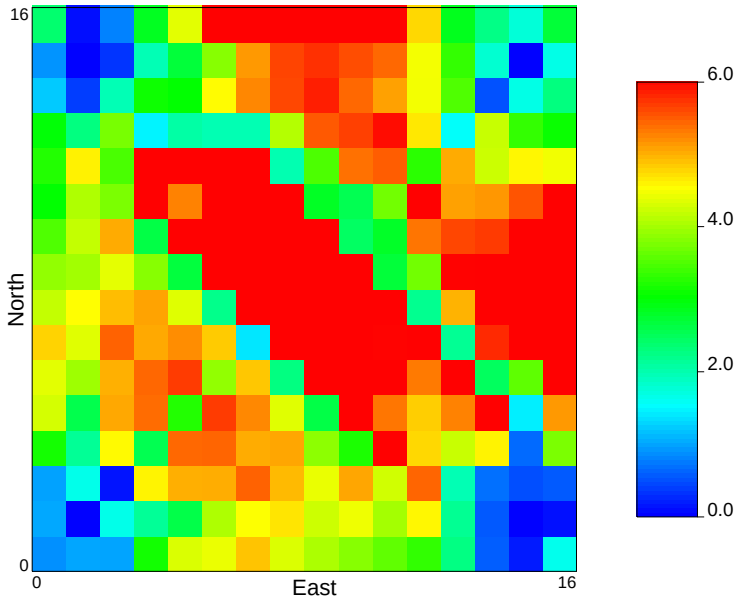

Figure 4.29: Variance distribution for $\sigma_p^2 = 1.0$ psi^2 and $\sigma_{wc}^2 = 1 \times 10^{-5}$

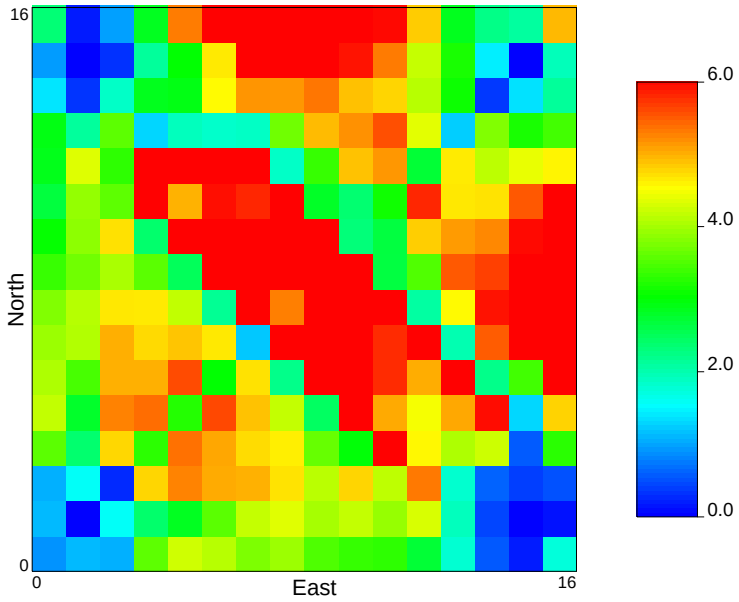

Figure 4.30: Variance distribution for $\sigma_p^2 = 0.5$ psi^2 and $\sigma_{wc}^2 = 5 \times 10^{-6}$

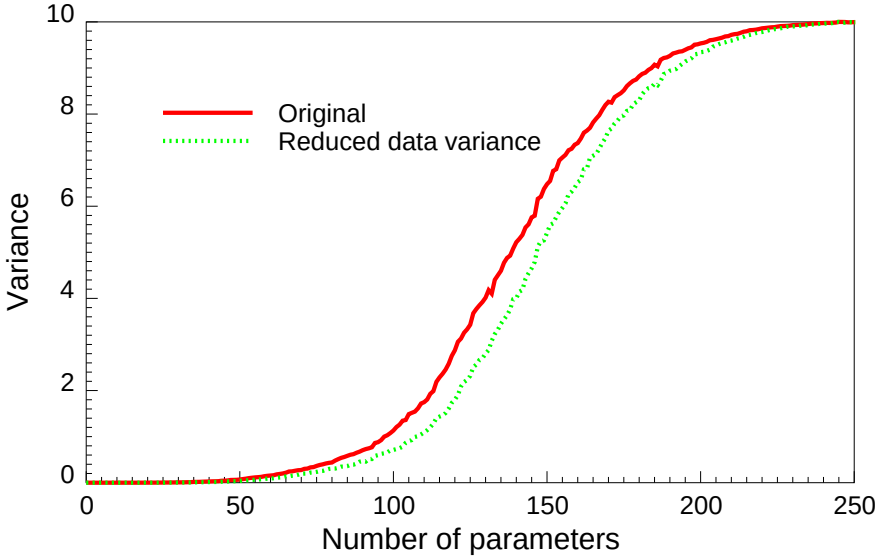

Figure 4.31: Singular values for the Hessian matrix (original and reduced data variance)

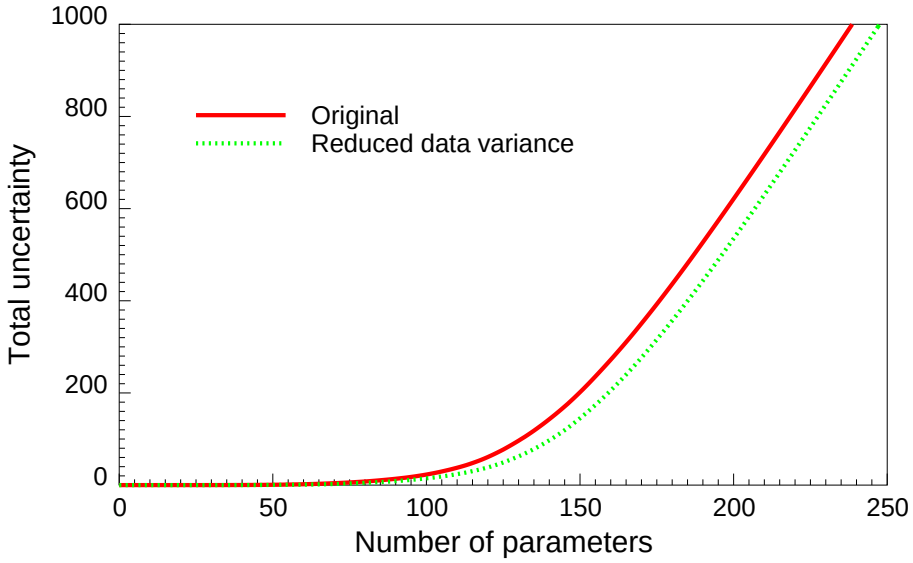

Figure 4.32: Trade-off curves (original and reduced data variance)

which means that prior information always reduces the uncertainty of estimation. Also as shown in Section 4.1.1, the Hessian matrix is better conditioned with prior information.

In this chapter, we demonstrated how to apply wavelets to reservoir characterization and to describe a reservoir with a relatively small set of parameters. We also looked into the feasibility of integrating different types of data using wavelets. In Chapter 5, we will shows the methodology and implementation with applications on a simple synthetic reservoir.

Chapter 5

Methodology and Implementation

In Chapter 3, we presented the methodology for a new parameterization technique using the wavelet analysis. As shown in Chapter 4, analysis of variance shows that the wavelet approach holds huge advantages over the conventional approaches (pixel or object modeling). However, in the context of reservoir parameter estimation, we still have several issues to address:

- *How to select the parameter set:* The original model parameters are the properties at each gridblock. Using the wavelet analysis, the model parameters become the wavelet coefficients of the original parameters. However, as shown in Chapter 4, the transformed model space is equivalent to the original model space. Our primary goal is to use a reduced parameter set to describe (or reproduce) the reservoir as well as possible. Several strategies are outlined for the parameter selection in this chapter.

- *Impact of different wavelet bases:* As we know, there are many different wavelet families, and in each family, there are different wavelet bases. Different wavelets will have different effects on the reservoir parameter estimation procedure.

- *Different reservoir types:* One of the major concerns about an inverse problem is that the procedure is usually under-determined and has nonunique solutions. The procedure is also problem-dependent. It is a primary consideration in this

study to investigate the viability of the inverse procedure for different reservoir configurations.

- *Feasibility on field-scale problems:* In practice, numerical reservoir simulations usually have discretizations with 100,000 to 1,000,000 gridblocks. With such a huge number of gridblocks, the computation is very intensive. Hence reservoir parameter estimation becomes very difficult, in the sense of computational loads and feasibility. Wavelets show promise as a means to reduce the size of the parameter space, however it remains to be seen whether the process of selection is computationally feasible with such large starting parameters - this issue will also be addressed here.

5.1 Image Compression

As described in Section 3.3.3, the discrete system (in this study, the discretized reservoir) can be reparameterized using the wavelet analysis. The new model space using wavelets is equivalent to the original model space, and the number of wavelet coefficients is the same as the number of the original parameters. One major objective of this study is to use as few parameters as possible to represent the discrete system. In other words, we want to compress the dimension of the model space. Wavelet analysis has proven to be very efficient in this category. By reducing the dimension of the model space, the inverse problem is better determined, and the computational load will be also decreased. The wavelet analysis is also a powerful tool for data integration, because it allows for the constraint of different resolutions to different types of data.

Wavelet analysis has gained popularity in image processing and compression. For instance, the JPEG-2 standard used the wavelet transform to replace the local cosine transform used in the JPEG standard. As demonstrated by the JPEG-2 standard, the wavelet analysis can compress images by a very large compression rate without losing much resolution.

The basic idea behind the data compression using wavelets is very straightforward: the wavelet transform can focus on localized signal structures with a zooming procedure that progressively reduces the scale parameter. Singularities and irregular structures often carry essential information in a signal. An image $\vec{\alpha}$ is also a two-dimensional signal (we use one-dimensional vectors to represent two-dimensional arrays throughout this study, in order to simplify the notation). The two-dimensional discrete wavelet transform can be represented as

$$\vec{c} = W \cdot \vec{\alpha}, \tag{5.1}$$

where W is the transformation matrix. The structure of W varies with responding to different wavelet bases and does not have to be orthogonal, though in this study we only considered orthogonal wavelet bases. Typically the number of significant wavelet coefficients is much smaller than the number of pixels in the original image. By eliminating the insignificant wavelet coefficients, we can reduce effectively the size of the transformed image. The image can be reconstructed by

$$\vec{\alpha}^w = W^{-1} \cdot \vec{c}^w, \tag{5.2}$$

where $\vec{c}^w$ is the wavelet coefficient vector with insignificant coefficients set to zero. The compression rate is defined as $R_c = N/N_w$, with N_w is the number of significant wavelet coefficients and N is the total number. We need to notice that this definition of compression rate is a little different from the one used in image compression, which is the ratio between the bit storage for an image before and after the transformation.

There are many techniques to optimize this data compression procedure, e.g., optimal wavelet basis, utilization of modulus maxima, etc. Nevertheless, the basis concept remains the same: reduce the number of coefficients without sacrificing resolution. The criterion for coefficient selection is based purely on the magnitudes.

The wavelet image representation of the original image a_0 is composed of $3L + 1$ subimages:

$$\left\{ a_L, \left\{ d_j^1, d_j^2, d_j^3 \right\}_{1 \le j \le L} \right\}. \tag{5.3}$$

Fig. 5.1 shows the disposition of wavelet coefficients $d_j^k[m, n]$ for decomposition level $L = 3$. More details can be found in (Mallat 1998).

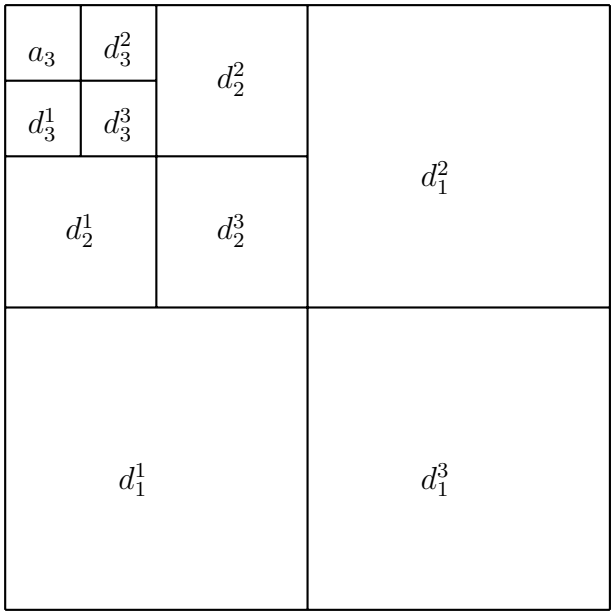

Figure 5.1: The disposition of wavelet coefficients $d_j^k[m,n] = \left\langle f, \psi_{j,m,n}^k \right\rangle$ for decomposition level $L = 3$

We demonstrated the feasibility of the image compression with a reservoir property description. Fig. 5.2 shows the porosity distribution (64×64 pixels) of a reservoir, which was generated by SGSIM in GSLIB (Deutsch & Journel 1997). The Haar wavelet was used as the orthogonal wavelet basis. Though the Haar wavelet only has a vanishing moment of $\nu = 1$ and has a tendency to yield blocky reconstructions, this wavelet still gave very good compression rates and low-distortion reconstructions.

In order to demonstrate the structure of the wavelet coefficients, Figs. 5.3 through 5.5 show the wavelet coefficients after discrete wavelet transforms with different decomposition levels. For instance, Fig. 5.3 shows the wavelet image representation of the original porosity distribution shown in Fig. 5.2 with $L = 1$. The wavelet image representation is composed of 4 subimages: $\{a_1, d_1^1, d_1^2, d_1^3\}$, each with a size of 32×32.

a_1 is the left-top subimage, d_1^1 is in the left-bottom, d_2^1 is in the right-top, and d_3^1 is in the right-bottom. a_1 is the scaling function at Level 1, while d_1^k are the wavelet decomposition coefficients at Level 1. In other words, a_1 is the lower-resolution average at Level 1, d_1^k are the local variations in three different directions. As we can notice from Fig. 5.3, most of the decomposition coefficients are very close to zero. There are only a small number of significant coefficients.

As shown in Appendix A, the discrete wavelet transform is a cascade procedure. The wavelet representation at Level $L+1$ is composed of $3 \times (L+1) + 1$ subimages:

$$\left\{ a_{L+1}, \left\{ d_j^1, d_j^2, d_j^3 \right\}_{1 \le j \le L+1} \right\},$$

which can be rewritten as

$$\left\{ \left\{ a_{L+1}, d_{L+1}^1, d_{L+1}^2, d_{L+1}^3 \right\}, \left\{ d_j^1, d_j^2, d_j^3 \right\}_{1 \le j \le L} \right\}. \tag{5.4}$$

The only difference in Eq. 5.4 from Eq. 5.3 is that a_L in Eq. 5.3 is replaced by $\left\{ a_{L+1}, d_{L+1}^1, d_{L+1}^2, d_{L+1}^3 \right\}$. Therefore, once we have the wavelet representation at Level L, we can obtain the wavelet representation at Level $L+1$ by decomposing a_L, the scaling functions at Level L, without changing any other decomposition coefficients. Figs. 5.4 and 5.5 show the wavelet representations of the original porosity distribution with $L = 2$ and $L = 3$. In each of these figures, the subimage in the left-top corner is the scaling function a_L at Level L, though with different sizes. As L increases, the size of a_L decreases.

What interests us most in the wavelet decomposition is that most of the decomposition coefficients are not significant. Therefore, we can threshold the coefficients and keep only the most significant coefficients. If we keep all the coefficients without thresholding, the inverse wavelet transform is a perfect reconstructor, which means the original image is reproduced perfectly. Fig. 5.6 shows the perfectly reconstructed porosity distribution, which is exactly the same as the distribution shown in Fig. 5.2. As we know, the redundancy in the image is very high, and we can reduce the redundancy by eliminating the insignificant coefficients.

Fig. 5.7 shows the reconstructed porosity distribution with 2390 wavelet coefficients. There is no visible difference between Fig. 5.6 and Fig. 5.7. In fact, the total

Table 5.1: Comparison between different reconstructed images

parameter#	E_r	$D_r(10^{-3})$	$\epsilon_r(\%)$	R_c
4096	0.000	0.00	0.000	1.000
2390	0.519	0.18	0.067	1.714
982	4.460	1.09	0.413	4.171
267	12.75	3.11	1.181	15.34

reconstruction error E_r is only 0.519, i.e., the standard deviation D_r is 1.76×10^{-4}. Because the average porosity $\bar{\phi}$ is 0.26346, the relative error ϵ_r is only 0.0667%. Figs. 5.8 and 5.9 show the reconstructed porosity distributions with 982 and 267 wavelet coefficients respectively. Even with such small numbers of coefficients, the reconstructed images are not much different from the original image. There is little visible difference in the reconstruction for both the figures. Table 5.1 shows the quality of reconstruction with different compression rates. Even with only 267 wavelet coefficients, the relative error of the reconstruction is only 1.181%.

We should also note that we used the Haar wavelet in this case. The Haar wavelet has a very compact support and a vanishing moment of 1, therefore it is not the best choice for image compression, especially when the compression rate is very large. For instance, Fig. 5.9 shows some blocky features in the reconstructed image. Even so, the reconstruction is still close enough to the original porosity distribution, for the purposes of reservoir performance prediction.

5.2 Different Choices for Parameter Selection

In Section 5.1, we explored the capability and efficiency of the wavelet analysis for image compression. In other words, we can use the wavelet analysis to reduce the number of parameters in images. In the case of a discrete inverse problem, we can effectively reduce the dimension of the model space this way. However, there is still an important question mark remaining: how to select the parameters? In inverse problems, the objective is to obtain an estimation, in our case, to obtain an image.

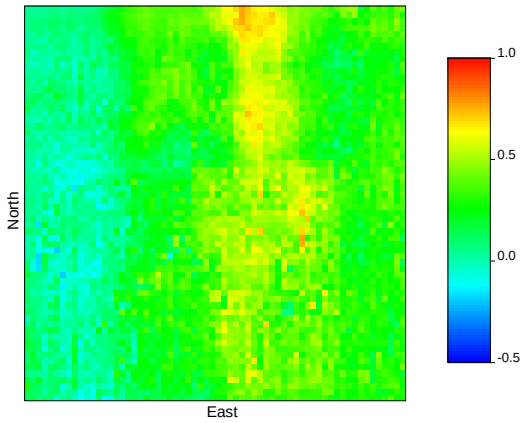

Figure 5.2: Porosity distribution

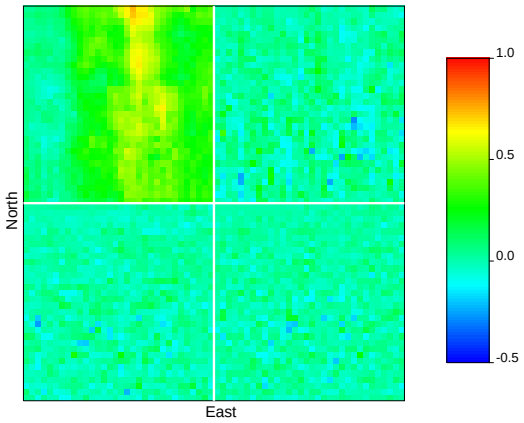

Figure 5.3: Wavelet decomposition of porosity (Level 1)

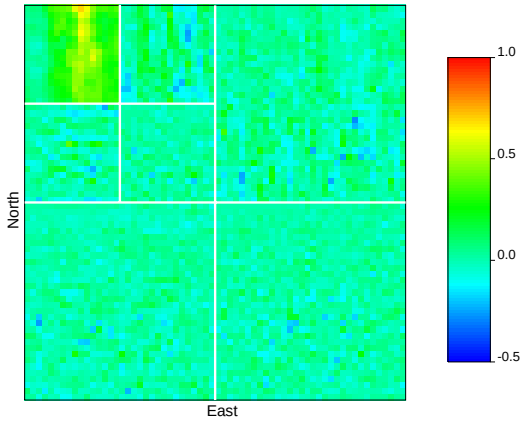

Figure 5.4: Wavelet decomposition of porosity (Level 2)

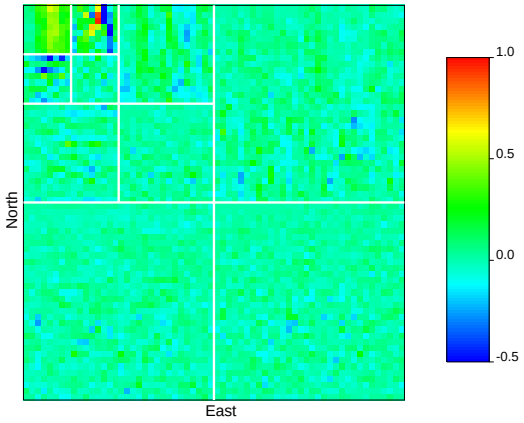

Figure 5.5: Wavelet decomposition of porosity (Level 3)

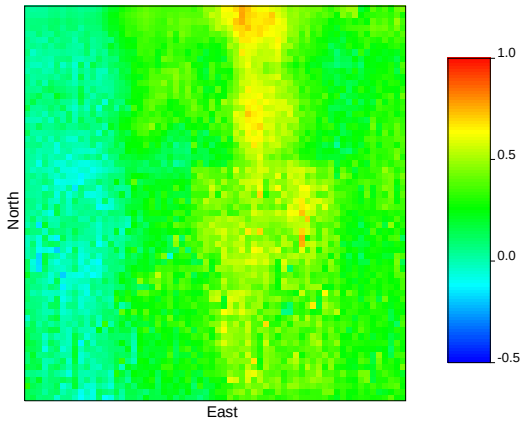

Figure 5.6: Original porosity distribution with 4096 pixels

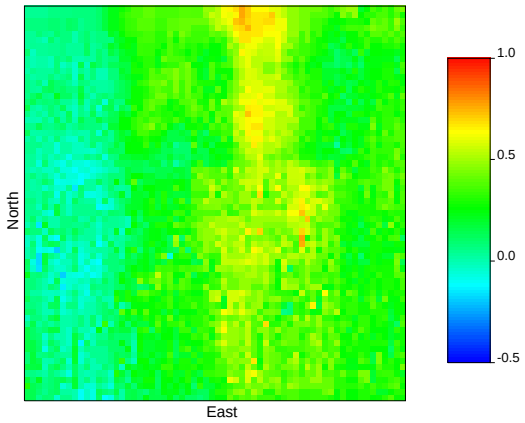

Figure 5.7: Reconstructed porosity distribution with 2390 wavelet coefficients

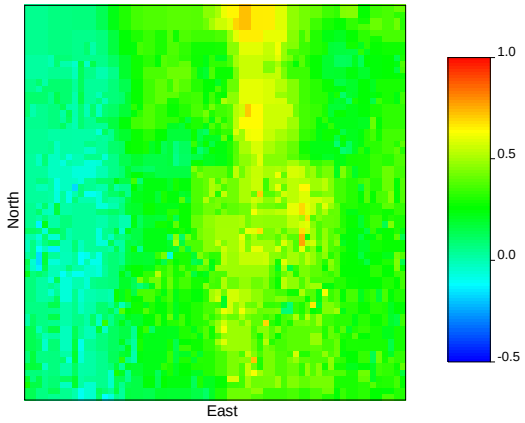

Figure 5.8: Reconstructed porosity distribution with 982 wavelet coefficients

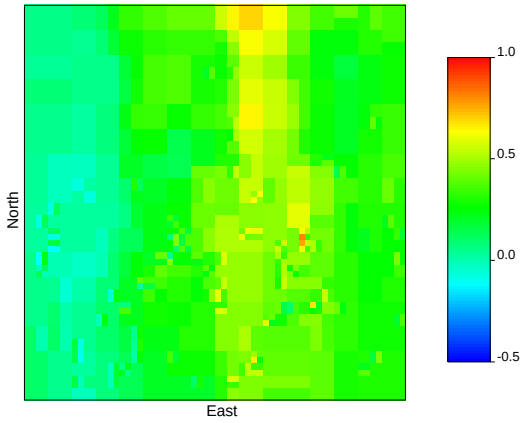

Figure 5.9: Reconstructed porosity distribution with 267 wavelet coefficients

Thus we do not have the image when the inverse procedure starts. Without the image, it is almost impossible to obtain a wavelet coefficient representation.

In this section, we will discuss several strategies to select a set of parameters for the reservoir parameter estimation problem. The test problem considered in this section represents two-dimensional, two-phase flow in a reservoir of uniform thickness $h = 100$ ft, with no-flow boundaries. The reservoir is rectangular with dimensions 1600×1600 ft^2. The reservoir is partitioned into 32×32 uniform gridblocks. No particular permeability distribution is assumed, except that the mean permeability is about 700 md and with a logarithm variance of 0.3, and no correlation between gridblocks. The permeability field is also assumed to be isotropic. The porosity distribution is assumed to be uniform at 0.22. Other relevant reservoir and fluid properties are as follows: system compressibility, $c_t = 3.0 \times 10^{-6}$ psi^{-1}; wellbore radius $r_{well} = 0.25$ ft; and the initial pressure is 5200 psi. Table 5.2 shows fluid viscosities and formation volume factors; Table 5.3 shows the relative permeability and capillary pressure (Fig. 5.10 shows the relative permeability curves). The standard deviation of measured data are assumed to be $\sigma_p = 1.0$ psi for bottomhole pressures, and $\sigma_{wc} = 0.3\%$ for watercuts.

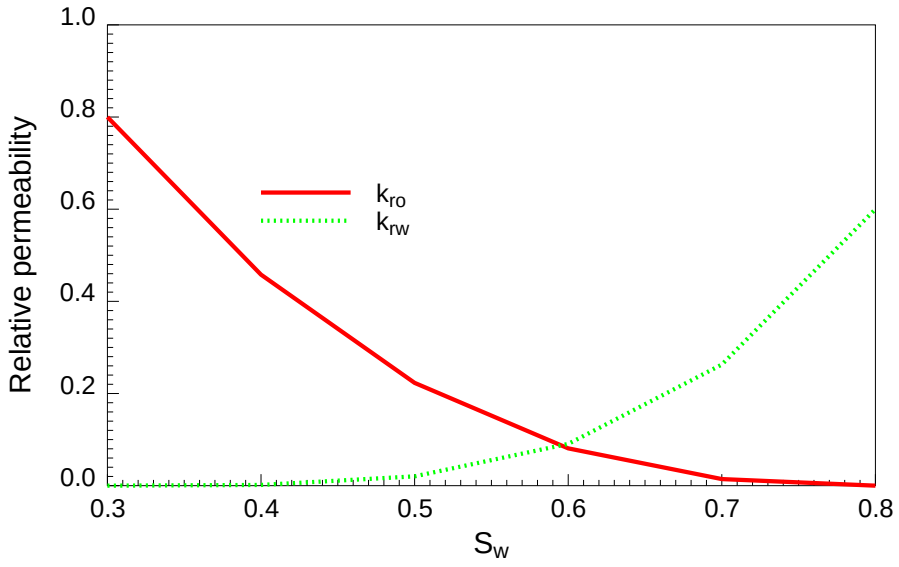

Figure 5.10: Relative permeability curves

Table 5.2: Fluid properties as functions of pressure

p_o (psi)	B_o	μ_o (cp)	B_w	μ_w (cp)
14.7	1.2204	1.500	1.0560	0.6
4014.7	1.1218	1.656	1.0391	0.6
8014.7	1.0232	1.812	1.0222	0.6

Table 5.3: Fluid properties as functions of water saturation

S_w	k_{ro}	k_{rw}	P_c (psi)
0.3	0.8	0.0	5.0
0.4	0.4579467	1.5558304E-03	4.0
0.5	0.2230838	2.0219631E-02	3.0
0.6	8.0954306E-02	9.0638116E-02	2.5
0.7	1.4310835E-02	0.2627751	2.0
0.8	0.0	0.6	1.0

Fig. 5.11 shows the true reservoir permeability distribution: a high-permeability strip with $k = 2500$ md is located in the middle of the reservoir, while the rest of the reservoir with $k = 500$ md. There are four active wells in the reservoir, with grid indices (3,3), (3,29), (29,3), (29,29) respectively. The well at the left-bottom corner is the injecting well with an injection rate at 10000 STB/day, the rest are all producing wells with production rates at 2000, 2000, 6000 STB/day respectively. The elapsed time is 400 days and all the producing wells have water breakthrough. Information for both the bottomhole pressure and watercut at each well is available. Figs. 5.13 through 5.16 show the production history at each well. We also assumed that 4-D seismic data is available. Throughout this chapter, when referring to 4-D seismic data, we mean the *interpretation* of 4-D seismic data, that is, the water saturation change inferred from two 3-D seismic surveys. Landa (1997) gave some methods to interpret 3-D seismic data to yield maps of water saturation changes. Because the seismic data are usually with high uncertainty, we need to associate the information from water saturation changes with larger variances. Fig. 5.12 shows the water saturation change between $t = 100$ days and $t = 200$ days.

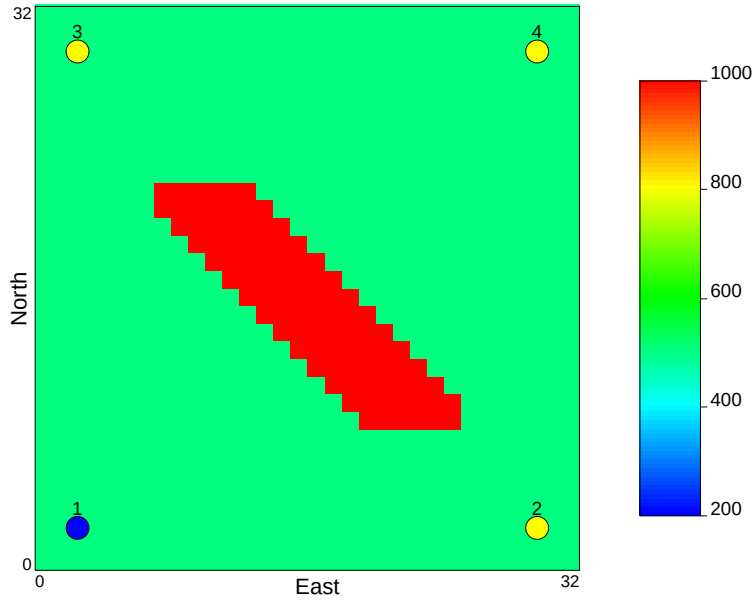

Figure 5.11: Reservoir permeability distribution (32×32 gridblocks)

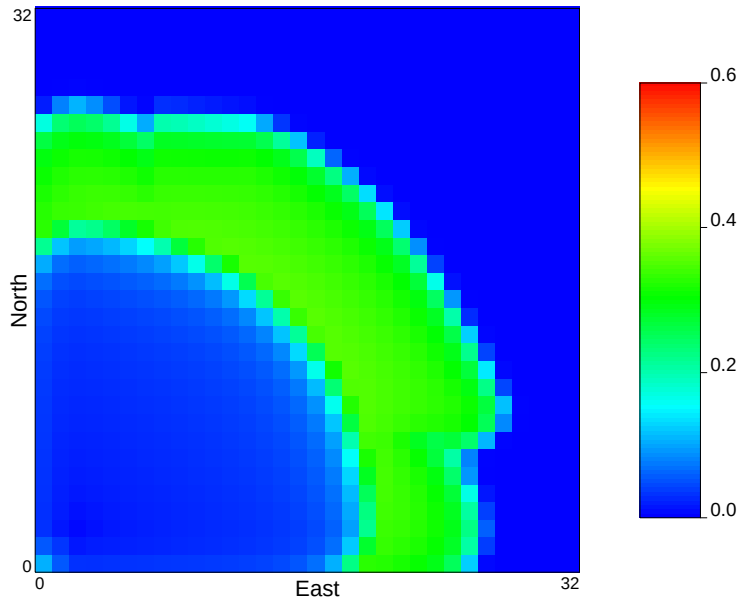

Figure 5.12: Water saturation change distribution during time t=100,200 days

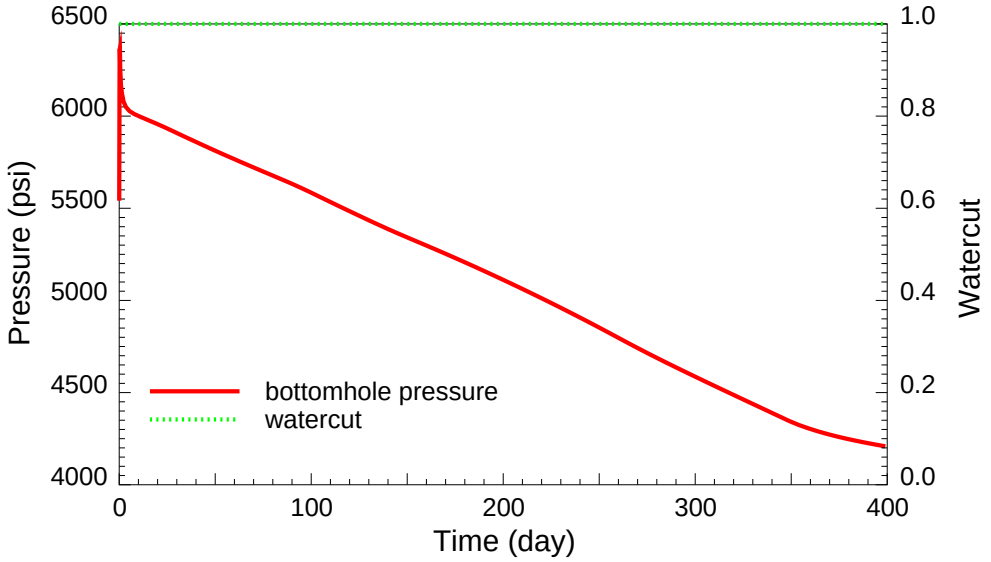

Figure 5.13: Bottomhole pressure and watercut at Well #1

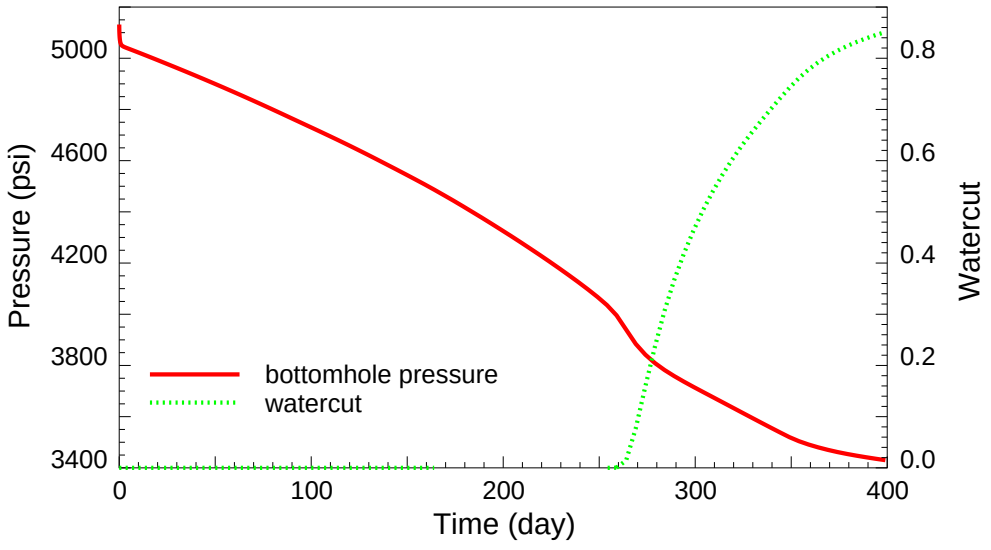

Figure 5.14: Bottomhole pressure and watercut at Well #2

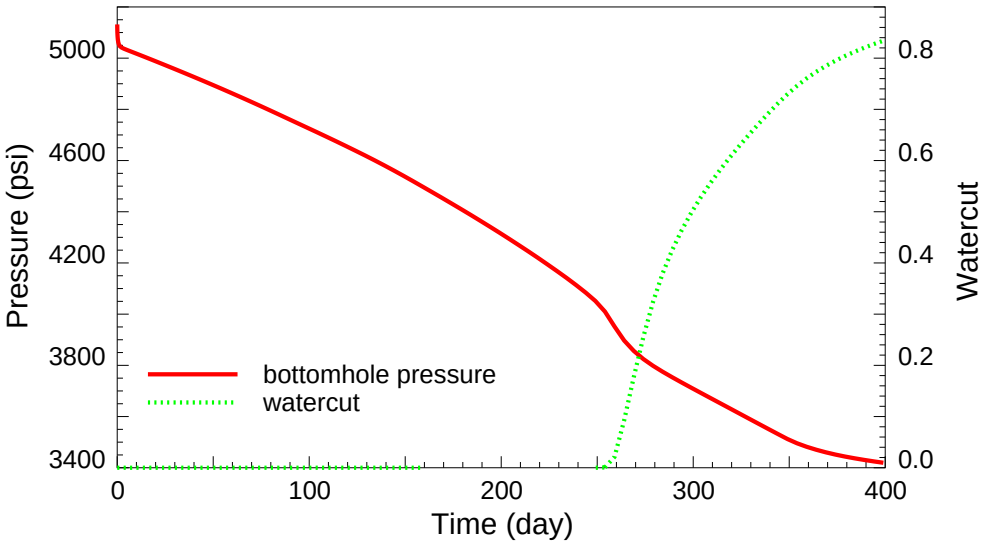

Figure 5.15: Bottomhole pressure and watercut at Well #3

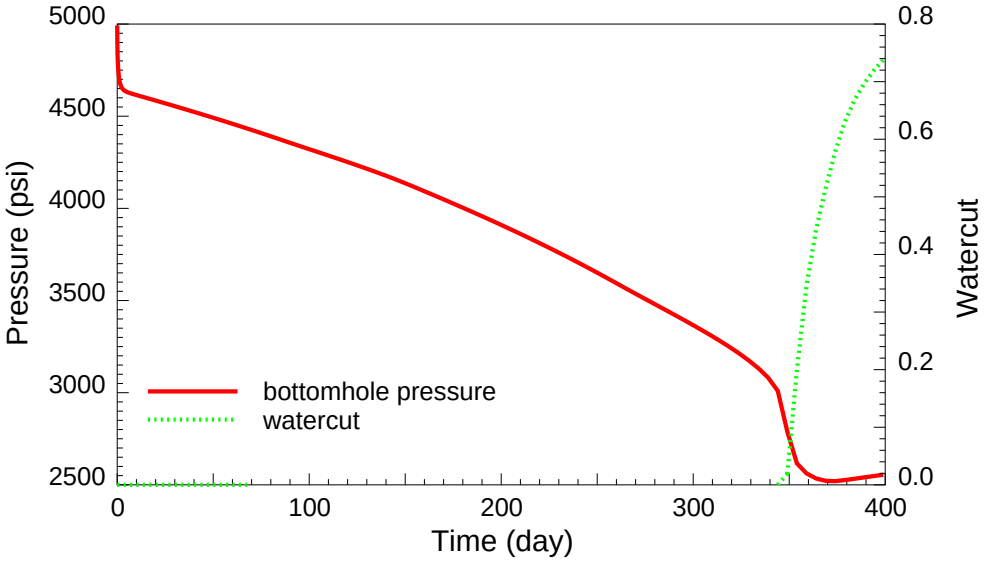

Figure 5.16: Bottomhole pressure and watercut at Well #4

In this section, all the dynamic data (bottomhole pressure, watercut, 4-D seismic data) are synthetic data, which were generated from the true reservoir property distribution without any errors. All the computations were carried out on a SGI Origin 200 workstation with four 270MHz MIPS R12000 (IP27) CPUs (we used two of the four CPUs for the test cases in this chapter). The program was parallelized using MPI and PETSCc (Balay, Gropp, McInnes & Smith 1999). Because we did not assume any prior distribution, we always started the nonlinear regression from a uniform permeability distribution $k^0 = 700$ md. Certainly this is not the best choice, it is always desirable to have a good initial guess. Nevertheless, the following examples show that our algorithms gave solutions matching the true permeability distribution very well, even without any prior distribution and with this poor initial guess.

5.2.1 Conventional Approach (Pixel Modeling)

Pixel modeling is perhaps the best-explored algorithm for this type of problem, as described by Landa (1997). The example problem contains 32×32 pixels or 1024 unknown permeability parameters.

In order to speed up the computation, we used a multistage approach for the pixel modeling. The first stage uses only the early-time bottomhole pressure data (including well test data). At this early stage, there is no water breakthrough in any producing wells, and the 4-D seismic data is not available yet. In this particular example, the test period is 50 days. This stage is very fast and takes only four iterations. Fig. 5.17 shows that the objective function is reduced to less than 40. Fig. 5.18 shows that each iteration takes only about 300 seconds. In total, the first stage takes less than 20 minutes to complete. However, the short-term pressure data does not have enough resolving power for all the parameters. As shown in Fig. 5.19, except for the wellbore neighborhoods, other parameters are not changed. In other words, most of the reservoir parameters are outside the investigation radius of the wells.

The second stage takes the long-term pressure data into consideration. This period is about 160 days. Fig. 5.20 shows that the investigation radius of wells reached the

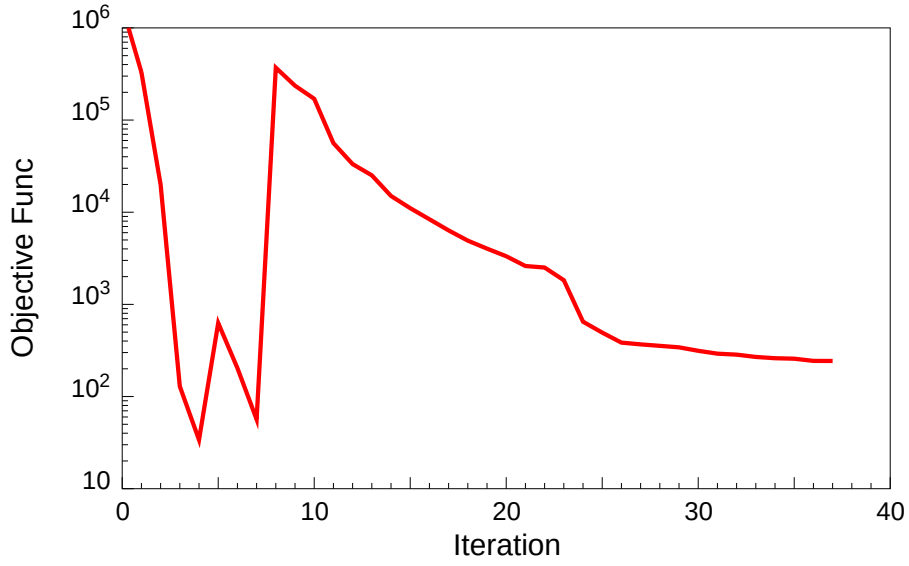

Figure 5.17: Convergence curve (pixel modeling, Case 5.2.1)

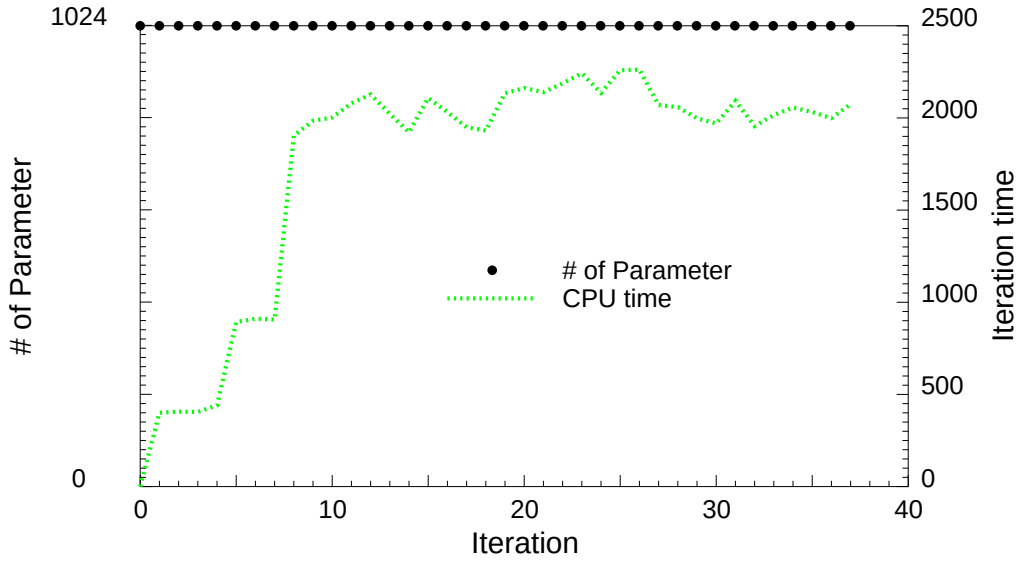

Figure 5.18: Number of parameters and CPU time at each iteration (pixel modeling, Case 5.2.1)

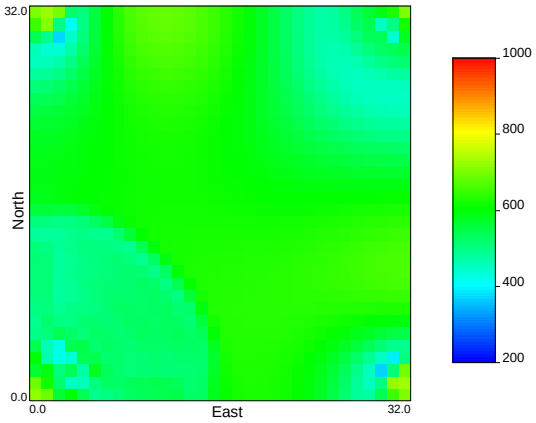

Figure 5.19: Permeability distribution after the 4^{th} iteration (pixel modeling, Case 5.2.1)

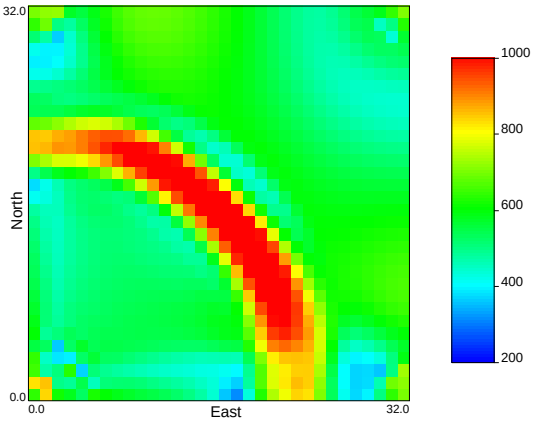

Figure 5.20: Permeability distribution after the 7^{th} iteration (pixel modeling, Case 5.2.1)

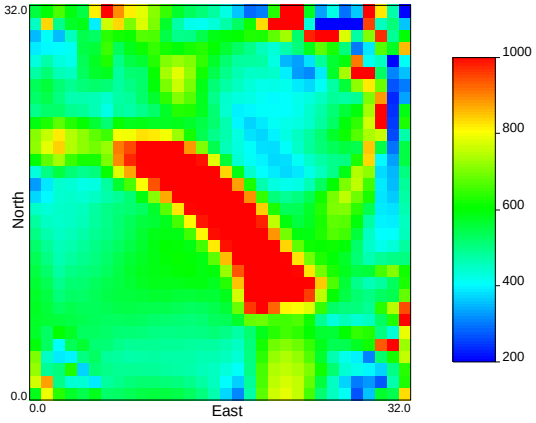

Figure 5.21: Permeability distribution after the 26^{th} iteration (pixel modeling, Case 5.2.1)

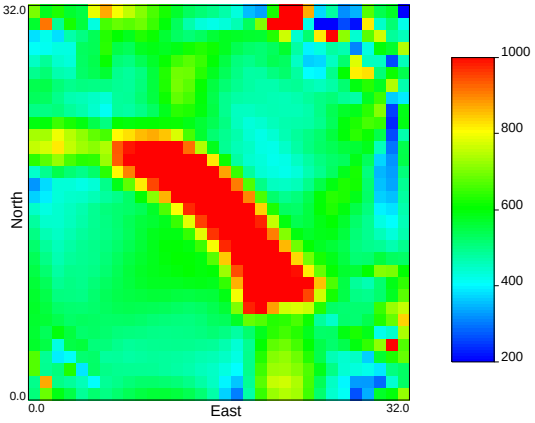

Figure 5.22: Permeability distribution after the 37^{th} iteration (pixel modeling, Case 5.2.1)

middle of the reservoir. This part of the reservoir is partially resolved. This stage takes three iterations. Therefore, the first and the second stages only take seven iterations together and the total computational time is about 72 minutes. In contrast to these two stages, the third stage takes much longer and the regression converges with great difficulty.

The third (final) stage makes use of all the dynamic data available: bottomhole pressures, watercuts, and the saturation change distribution. Though the first two stages have reduced the objective function to less than 100, due to the inclusion of watercut information, the overall objective function rises to more than 1×10^6. This example suggests that it is much easier to match the bottomhole pressure data than to match the watercut data. Also as shown in Section 3.2, the sensitivity of a parameter (in pixel modeling, the gridblock property) is much higher when the water front reaches the neighboring region of the particular parameter. In other words, the water front carries much more information and has a high resolving power. Vasco, Yoon & Datta-Gupta (1998) showed that most of the dominant features of the reservoir at the modeling scale become apparent even when only the first arrival times are matched.

Figs. 5.21 and 5.22 show the permeability distribution at the 26^{th} and 37^{th} iterations respectively. The estimated permeability distribution matches the true permeability distribution shown in Fig. 5.11 very well. Well #4 (the well at the right-top corner) breaks through last and there is only about 50-day watercut information available. Therefore, the mismatch around this area is much higher than the other wellbore neighborhoods.

Fig. 5.18 also shows the number of parameters used in each iteration. Because there is not any parameter reduction involved, the algorithm used all the parameters at each iteration.

5.2.2 Uniform-Resolution Wavelet Approach

In this study, we only consider orthogonal wavelet bases. Let $B = \{\vec{g}_m\}_{m \in N}$ be an orthogonal wavelet basis, the best linear estimation of a signal $\vec{f}$ is to estimate the

first M parameters. The resulting approximation is

$$\vec{f}_M = \sum_{m=1}^{M} \left\langle \vec{f}, \vec{g}_m \right\rangle \vec{g}_m = \sum_{m=1}^{M} \alpha_m \vec{g}_m, \tag{5.5}$$

where

$$\alpha_m = \left\langle \vec{f}, \vec{g}_m \right\rangle, \text{ for } m = 1, 2, ..., M. \tag{5.6}$$

That is, the dimension of the model space is reduced to M with $\{\alpha_m\}_{1 \le m \le M}$ as the model parameters, and the model space is spanned by $\{\vec{g}_m\}_{1 \le m \le M}$. The approximation error is

$$\vec{f} - \vec{f}_M = \sum_{m=M+1}^{N} \left\langle \vec{f}, \vec{g}_m \right\rangle \vec{g}_m, \tag{5.7}$$

and

$$\epsilon_M = \left\| \vec{f} - \vec{f}_M \right\|^2 = \sum_{m=M+1}^{N} \left| \vec{f}, \vec{g}_m \right|^2. \tag{5.8}$$

This estimation is efficient only if ϵ_M has a fast decay when M increases, which depends on the properties of $\vec{f}$ and of the basis. If the signal $\vec{f}$ is uniformly smooth, the linear estimation will be efficient. However, this is not usually the case in reservoir parameter estimation. Reservoir property distributions usually have irregularities and discontinuities, such as channels and faults.

Yoon, Datta-Gupta, Vasco & Behrens (1999) presented a multiscale approach to production data integration. Different numbers of parameters are used at different iterations, but the grid spacing for parameters remains uniform. Let us suppose the two-dimensional reservoir griding is composed of $2^M \times 2^N$ regular grids. At first, only $2^{M-L} \times 2^{N-L}$ parameters are used in the integration. Later on, more and more parameters are added into the integration. There are totally $L+1$ different settings, each with $2^{M-l} \times 2^{N-l}$ parameters, for $l = 0, 1, ..., L$. At the l^{th} setting, the parameters are as follows:

$$\alpha^{(l)}[i, j] = \frac{1}{4^l} \sum_{p=0}^{2^l-1} \sum_{q=0}^{2^l-1} a[i \cdot 2^l + p, j \cdot 2^l + q], \tag{5.9}$$

for $i = 0, 1, ..., 2^{M-l} - 1, j = 0, 1, ..., 2^{N-l} - 1$. Therefore, a parameter at the l^{th} scale is an arithmetic average of 4^l original parameters.

In general, the wavelet image representation of the original image a_0 is composed of $3L + 1$ subimages:

$$\left\{a_L, \left\{d_j^1, d_j^2, d_j^3\right\}_{1\le j\le L}\right\}. \tag{5.10}$$

If we only keep the parameters at the coarse level L, i.e., only the scaling functions $\{a_L\}$, the parameter estimation becomes a uniform estimation. Furthermore, the parameters have uniform resolution. The approach proposed by Yoon, Datta-Gupta, Vasco & Behrens (1999) is actually a special case of the linear estimation with uniform resolution, which we denoted as the uniform-resolution approach. If we use the Haar wavelet as the orthogonal wavelet basis, the parameterization is slightly different from Yoon, Datta-Gupta, Vasco & Behrens (1999)'s approach, defined as

$$a_l[i,j] = \frac{1}{2^l}\sum_{p=0}^{2^l-1}\sum_{q=0}^{2^l-1} a[i\cdot 2^l + p, j\cdot 2^l + q]. \tag{5.11}$$

As in Section 5.2.1, we used a multistage approach for the nonlinear regression. The first stage uses the early-time bottomhole pressure history data (including well test data). The second stage takes the long-term bottomhole pressure data into consideration. The third (final) stage uses all the dynamic data available: bottomhole pressures, watercuts, and the saturation change distribution. However, in this uniform-resolution wavelet approach, we used only 64 parameters ($L = 2$) for the first two stages. The final stage used the level by level approach, i.e., the nonlinear regression used the 64 parameters ($L = 2$) at first, then 256 parameters ($L = 1$), and at last, all the 1024 parameters ($L = 0$).

In the first two stages, there is no water breakthrough in any producing wells, and the 4-D seismic data is not available yet. The computation for the first two stages is very fast and takes only seven iterations. Unlike the first two stages in Case 5.2.1, although the bottomhole pressure data still does not have enough resolving power for each individual parameter at the gridblock level, it can resolve the arithmetic averages (which are included inherently among the wavelet coefficients). As shown in Fig. 5.25, in addition to the wellbore neighborhoods, the production data is also sensitive to the parameters in the middle of the reservoir, even though we were not able to distinguish each individual gridblock. In other words, this part of reservoir is only resolved at a coarse resolution level.

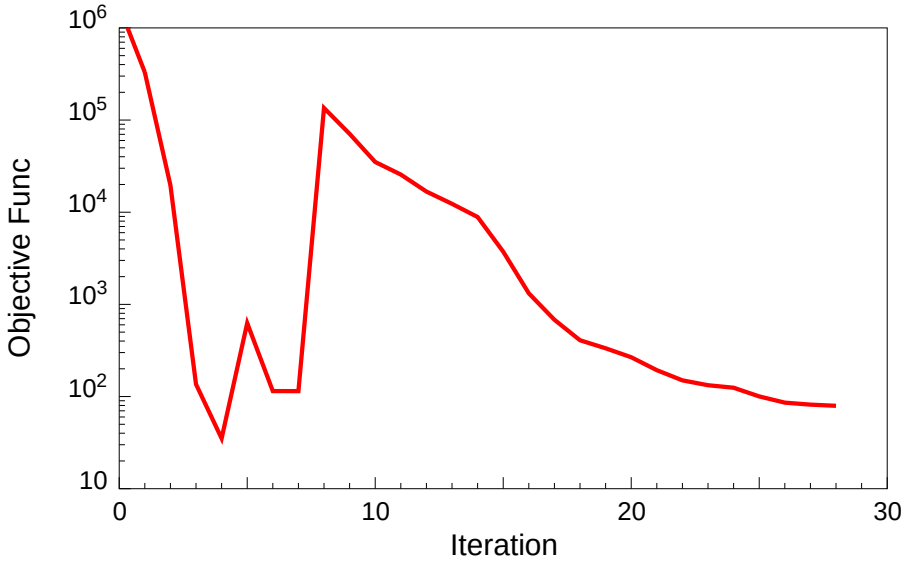

Figure 5.23: Convergence curve (level by level, Case 5.2.2)

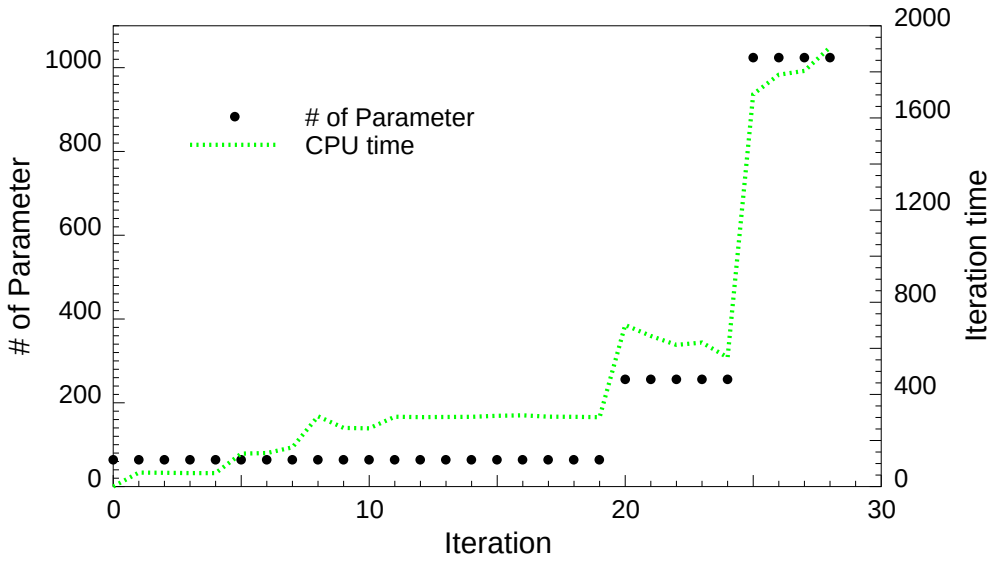

Figure 5.24: Number of parameters and CPU time at each iteration (level by level, Case 5.2.2)

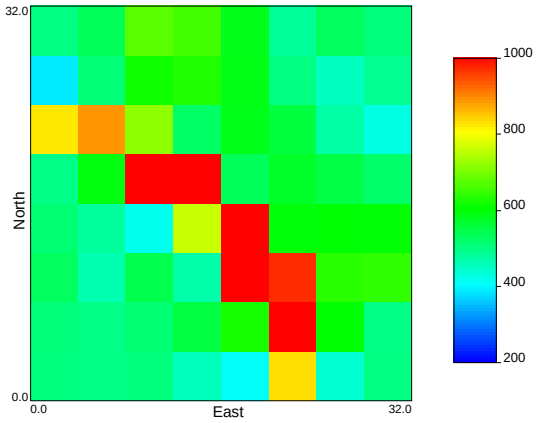

Figure 5.25: Permeability distribution after the 7^{th} iteration (level by level, 64 parameters, Case 5.2.2)

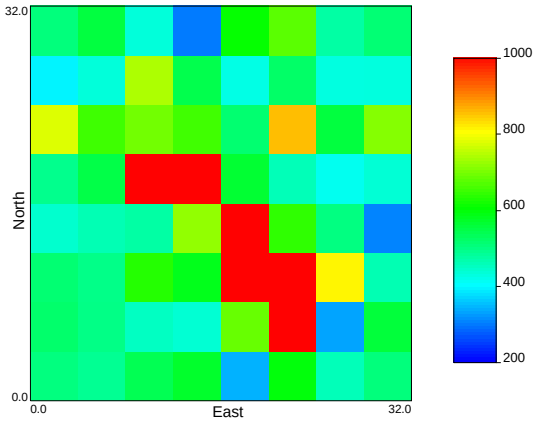

Figure 5.26: Permeability distribution after the 19^{th} iteration (level by level, 64 parameters, Case 5.2.2)

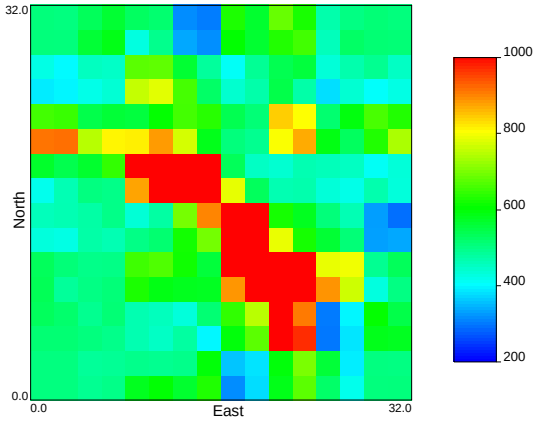

Figure 5.27: Permeability distribution after the 24^{th} iteration (level by level, 256 parameters, Case 5.2.2)

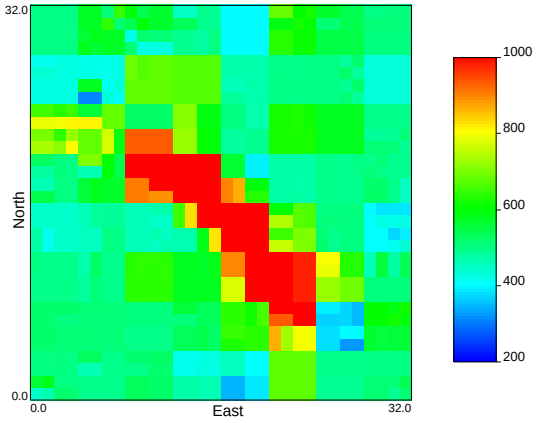

Figure 5.28: Permeability distribution after the 35^{th} iteration (level by level, 1024 parameters, Case 5.2.2)

There are three different resolution levels in the final stage. When the objective function does not decrease fast enough, or it is already below a certain threshold, the regression algorithm will switch to the higher-resolution level, as shown in Figs. 5.23 and 5.24. The regression results are outlined here:

- Fig. 5.26 shows the permeability distribution with only 64 parameters. As we can see, though with a very low resolution, the distribution captures the overall structure of the reservoir.

- Fig. 5.27 shows the permeability distribution with 256 parameters. At this point, the estimated permeability distribution matches the true permeability distribution shown in Fig. 5.11 very well.

- Fig. 5.28 shows the permeability distribution with all the 1024 parameters. There are only slight differences from the result shown in Fig. 5.27. Most of the permeability distribution is not changed much.

This uniform-resolution wavelet approach has several advantages over the pixel modeling. First of all, the regression is much faster, because we used no more than 256 parameters most of the time. Secondly, with lower resolutions, we can have a better understanding of the reservoir structure. However, there are also some drawbacks associated with this approach:

- This method does not consider the heterogeneity around wellbores. In the situations where the heterogeneity is very high, the low-resolution parameterization might not be able to match the data. Even with matched data, the history matching may result in unrealistic estimation.

- There is still some computational inefficiency associated with this method. For instance, as shown in Fig. 5.24, because we need to calculate sensitivity of all the parameters, the CPU time for each iteration at the highest resolution is very high (more than 2000 seconds for the test case), yet not all the parameters contribute to the parameter estimation.

- The uniform-resolution wavelet approach always uses uniform parameterization. This parameterization technique only applies when the reservoir distribution is uniformly smooth. This assumption usually does not hold in reservoir parameter estimation.

5.2.3 Nonuniform-Resolution Wavelet Analysis

In order to avoid some of the disadvantages of the uniform-resolution wavelet approach, we applied a nonuniform-resolution estimation using the wavelet analysis. Instead of always using the *first* M parameters to approximate a function $\vec{f}$, the nonuniform-resolution estimation uses the *best* M parameters. The approximation is

$$\vec{f}_M = \sum_{m=1}^{M} \left\langle \vec{f}, \vec{g}_{j[m]} \right\rangle \vec{g}_{j[m]} = \sum_{m=1}^{M} \alpha_{j[m]} \vec{g}_{j[m]}, \tag{5.12}$$

where $\{j[m]\}_{1 \le m \le M}$ is the best parameter sequence,

$$\alpha_{j[m]} = \left\langle \vec{f}, \vec{g}_{j[m]} \right\rangle, \text{ for } m = 1, ..., M \tag{5.13}$$

are wavelet coefficients on the M best bases. For the uniform-resolution estimation, the parameter sequence $j[m]$ is simply

$$j[m] = j, \text{ for } m = 1, 2, ..., M. \tag{5.14}$$

For a function with uniform regularities, these two approaches are not very different from each other. However, many signals and images we encounter in practice do not preserve such nice properties, e.g., smoothness and regularity. In reservoir parameter estimation, there are two major factors that make the uniform-resolution estimation unsuitable in many cases:

1. The observed data are usually concentrated around wellbores, e.g., bottomhole pressure, watercut, well logging, coring, etc. In this sense, the data is much more sensitive to well neighborhoods. Appropriate modeling for these areas is crucial to reservoir parameter estimation. For reservoirs with very strong heterogeneities, these parameters have to be (and can be) estimated with high resolution. The uniform-resolution estimation can not account for this local refinement.

2. There are many geological structures in reservoirs, e.g., channels, faults, etc. These geological structures have very strong discontinuities. Though the discontinuities may be at very small scales, the reservoir behaviors may depend on them heavily.

In this section we focus on the nonuniform-resolution estimation using wavelet analysis. For this nonuniform-resolution approach with wavelets, there are several issues to be addressed:

- *Choice of wavelet bases* $\vec{g}_m$: There are many different wavelet bases with different properties. In practice, our choices are limited to several wavelet families. First of all, we should use a wavelet basis with a compact support, because the number of gridblocks in the discrete system is not very large, usually no more than hundreds in any dimension. Secondly, the orthogonality is a desired property in this study, for the sake of variance analysis. Therefore, the best choice is an orthogonal wavelet basis with a compact support. Because the Daubechies wavelets have the most compact supports with given vanishing moments, we always used the Daubechies wavelets in this study.

- *Best parameter set* $j[m]$: It is a challenging task to select the best parameter set for the nonuniform-resolution estimation. In image processing, we have an image at hand. The reduced parameter set can be determined by the properties of the image. This is not the case for inverse problems, in which we do not have a very good understanding of the property distribution. Usually, the prior distribution can provide some guideline for the parameter selection. In this section, we will describe several strategies to select parameters, in two major categories.

- *Parameter estimation for* $\{\alpha_m\}$: The ultimate objective in this study is to estimate the parameters, in this case, the wavelet coefficients $\{\alpha_m\}$. The core of the nonlinear regression is the Gauss-Newton algorithm, with some necessary modifications.

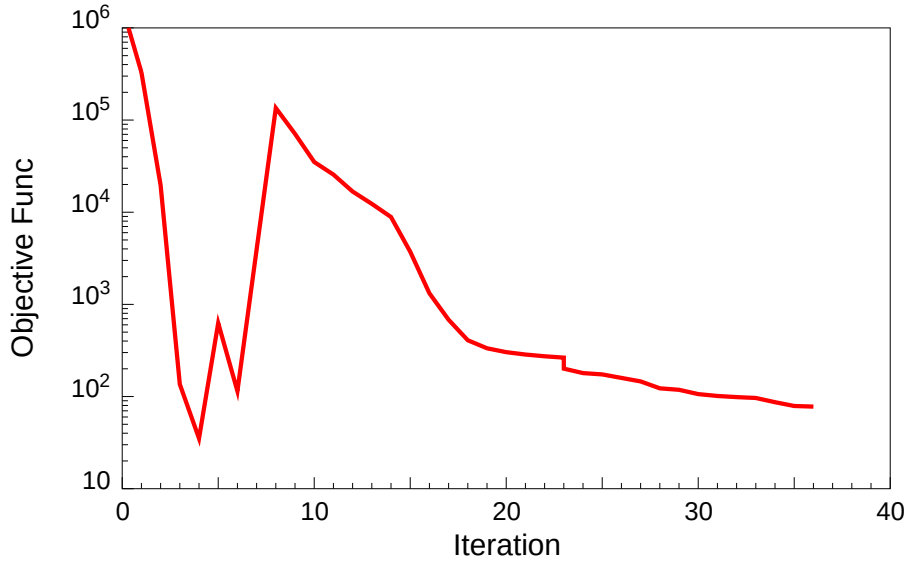

Figure 5.29: Convergence curve (4-D seismic and Haar wavelets, Case 5.2.3)

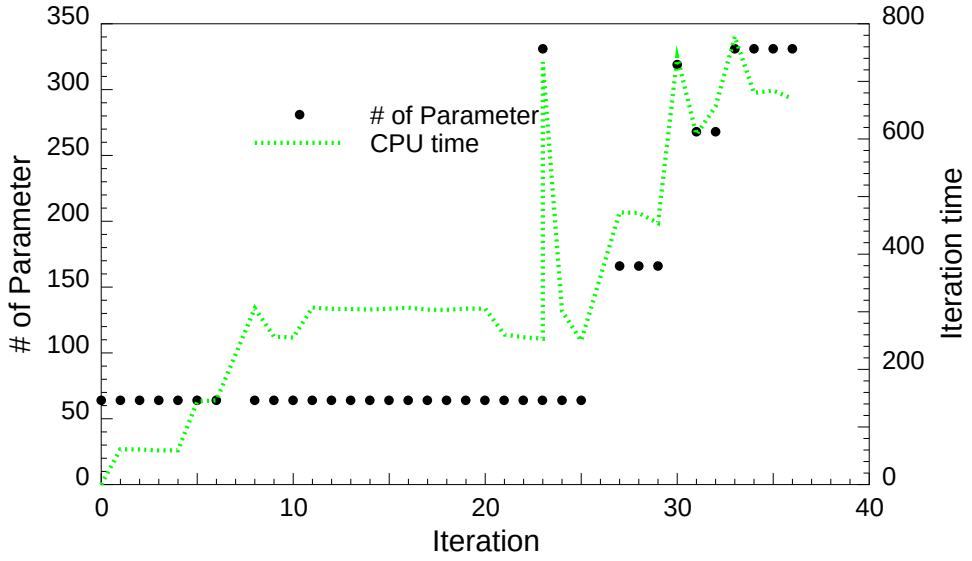

Figure 5.30: Number of parameters and CPU time at each iteration (4-D seismic and Haar wavelets, Case 5.2.3)

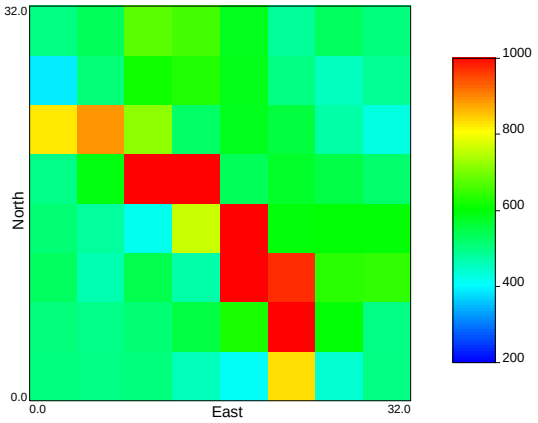

Figure 5.31: Permeability distribution after the 7^{th} iteration (4-D seismic and Haar wavelets, 64 parameters, Case 5.2.3)

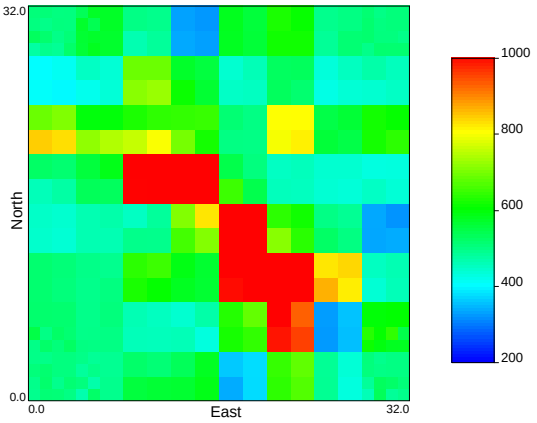

Figure 5.32: Permeability distribution after the 23^{th} iteration (4-D seismic and Haar wavelets, 331 parameters, Case 5.2.3)

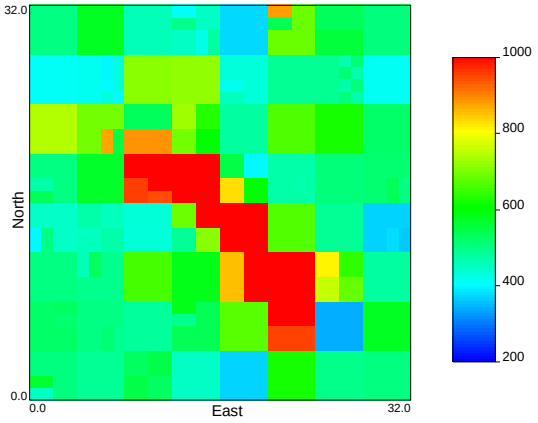

Figure 5.33: Permeability distribution after the 29^{th} iteration (4-D seismic and Haar wavelets, 166 parameters, Case 5.2.3)

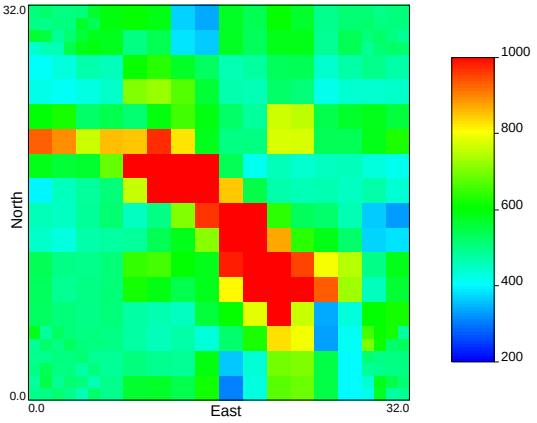

Figure 5.34: Permeability distribution after the 33^{th} iteration (4-D seismic and Haar wavelets, 331 parameters, Case 5.2.3)

In this section, we describe a method to account for the heterogeneities of wellbore neighborhoods. We used the Haar wavelet for this test case. As in the uniform-resolution estimation (Section 5.2.2), we used the first 64 parameters ($L = 2$) for the first two stages. Therefore, at first the regression process is exactly the same as in Section 5.2.2. Fig. 5.31 shows the permeability distribution after the first two stages.

After the first two stages, we have a clearer pattern of the overall reservoir structure. Any further regression will only change the local variation of the reservoir to further match the observed data. Because the flow pattern will not change much after the initial matching, we can perform the analysis of variance on the basis of the current permeability distribution. We can set up an initial parameter set $\{a_i^{(0)}\}_{1\le i\le M}$, with all the parameters at a low-resolution level (in this case, $L = 1$), plus some local refinement at wellbore neighborhoods. In most cases, this parameterization should have sufficient resolution to capture the major flow pattern. There are two different implementations in this case. The two implementations are slightly different in the way to reorder and to select the parameters. The first algorithm is outlined as follows:

1. Run the nonlinear regression for one iteration, with the parameter set $\{a^{(0)}\}$.

2. The uncertainty of each parameter in $\{a^{(0)}\}$ can be estimated using Eq. 4.12. We can discard the parameters with higher uncertainty (or lower sensitivities) from the parameter set $\{a^{(0)}\}$. Thus we have a new parameter set $\{a_i^{(1)}\}_{1\le i\le M'}$, with $M\prime$ much smaller than M.

3. Rerun the nonlinear regression for several iterations, with the parameter set $\{a^{(1)}\}$. Then calculate the uncertainty of each parameter in $\{a^{(1)}\}$, discard those with high certainty and put them back to the original parameter sequence. Choose the best parameters next in line in the original parameter set. Now we have a new parameter set $\{a^{(2)}\}$.

4. Set $\{a^{(1)}\} = \{a^{(2)}\}$ and go back to step 3.

5. Once we have used up all the parameters in the original set $\{a^{(0)}\}$, we need only include the parameters with higher resolutions.

The advantage of this method is that it is very easy to implement. However, we have to assume the variance sequence to be fairly static during the regression. The second method can select the parameters adaptively, even beyond the scope of $\{a^{(0)}\}$. The algorithm is as follows:

1. Run the nonlinear regression for one iteration, with the parameter set $\{a^{(0)}\}$.

2. The uncertainty of each parameter in $\{a^{(0)}\}$ can be estimated using Eq. 4.12. We can discard the parameters with higher uncertainty from the parameter set $\{a^{(0)}\}$. Thus we have a new parameter set $\{a_i^{(1)}\}_{1\leq i\leq M'}$, with $M\prime$ much smaller than M.

3. As we know, the initial parameter set is not complete, because it does not include all parameters with high certainty. The high-resolution parameters are not included in the parameter set. We can then do local refinement for those parameters with lowest uncertainty. That is, if a parameter with high certainty is not at the highest resolution, we include the higher-resolution wavelet coefficients at the same location. Now we have a parameter set $(\{a_i^2\}_{1\leq i\leq M''}$

4. Set $\{a^{(0)}\} = \{a^{(2)}\}$, and go back to Step 1.

In this particular test case, the two algorithms are not very different from each other in performance. The following figures show the results from the first algorithm.

Fig. 5.32 shows the permeability distribution after the 23^{th} iteration. This is the first iteration after running the regression with the low-level parameters. In fact, in this iteration, we included 331 parameters in the active parameter set. Fig. 5.32 shows that the resolutions are different in different regions, because we refined the wellbore neighborhoods locally.

Fig. 5.33 shows the permeability distribution with 166 parameters. At this point, the objective function has been reduced to 118, and the estimated permeability distribution matches the true permeability distribution shown in Fig. 5.11 very well.

Fig. 5.34 shows the permeability distribution with all the 331 parameters in $\{a^{(0)}\}$. Because the objective function is already very small, we did not perform any further refinement of the parameterization.

In summary, there are several advantages in using nonuniform-resolution estimation. First of all, the computational load is significantly reduced, compared with the pixel modeling and the level-by-level uniform-resolution approach. Secondly, the nonuniform resolution can account adaptively for the resolution of the observed data by refining the parameterization only at regions with most impact on reservoir responses. Therefore, we do not need to refine uniformly across the reservoir. We do not even need to keep the same resolution in different directions. Thirdly, this approach can also effectively avoid local minima by running the regression at lower resolution first.

5.2.4 Integrated Static Information (3-D Seismic Data)

In previous approaches, we used the 4-D seismic data as additional dynamic data. The 4-D seismic data provides the spatial resolving power lacking in both the pressure and watercut data, as shown in Section 4.3.1. However, the 4-D seismic data is not always available in practice. It is also very difficult to interpret the 4-D seismic data. In practice 3-D seismic data is more widely used and is much easier to obtain.

In this work, we developed an algorithm to integrate 3-D seismic data directly into the inverse procedure, using the wavelet analysis. Given a seismic data set (assuming the seismic data is at the same resolution as the numerical simulation), we can perform the wavelet transform on this data set. In most cases, the large-amplitude coefficients or local maxima indicate the locations of discontinuities or edges. In reservoirs, the discontinuities are usually the major geological features such as faults and channels, which are important to the reservoir characterization and performance. We can use the 3-D seismic image as a basis to pick up the first p most significant wavelet coefficients as the model parameters in our inverse procedure. The procedure is as follows. First, we have to center the variations of the seismic data $\vec{\gamma}$ to zero, that is, to subtract the mean μ_γ from $\vec{\gamma}$:

$$\vec{\gamma}^0 = \vec{\gamma} - \mu_\gamma. \tag{5.15}$$

Then we perform the wavelet transform on $\vec{\gamma}^0$ as

$$\begin{bmatrix} \vec{\omega}_p \\ \vec{\omega}_0 \end{bmatrix} = \begin{bmatrix} W_p \\ W_0 \end{bmatrix} \cdot \vec{\gamma}^0, \tag{5.16}$$

where $\vec{\omega}_p$ are the first p significant wavelet coefficients for $\vec{\gamma}^0$. Eq. 5.16 implies that

$$\begin{aligned} \vec{\gamma}^0 &= \left[W_p^T \cdot W_0^T\right] \cdot \begin{bmatrix} \vec{\omega}_p \\ \vec{\omega}_0 \end{bmatrix} \\ &= W_p^T \vec{\omega}_p + W_0^T \vec{\omega}_0 \approx W_p^T \vec{\omega}_p. \end{aligned} \tag{5.17}$$

If we assume the 3-D seismic data (or its interpretation) is correlated with the reservoir property distribution, the transformation matrix W_p usually characterizes the reservoir as well. Therefore, we can choose $\vec{c}_p = W_p \cdot \vec{\alpha}$ as the new model parameters. The real reservoir model parameters become $\vec{\alpha} = W_p^T \cdot \vec{c}_p$. In other words, instead of using the seismic data directly, we use the pattern captured by the transformation matrix W_p as a mask to pick up the active parameter set in the wavelet space.

Unlike the spectral decomposition technique(Oliver 1994b, Reynolds, He, Chu & Oliver 1995), we do not need to assume any explicit prior probability distribution for the model parameters. Furthermore, because we do not use the actual seismic distribution to infer any individual parameter, but only the overall structure (the mask as defined by W_p), this technique will work well even if the 3-D seismic data is only loosely correlated with the reservoir property distribution.

This approach has several advantages:

- The total uncertainty of estimation is reduced. As shown in Section 4.2.2, the estimation with a smaller number of parameters always has a lower uncertainty:

$$\tilde{\Sigma}_M(p) < \Sigma_M, \text{ for } p < N. \tag{5.18}$$

- The seismic data reflects the geological structure of the reservoir, therefore, the wavelet mask can preserve the same geological structure as well.

- It is not required to have any assumption about the prior probability distribution, as long as the seismic data is coupled with the reservoir property distribution.

Table 5.4: Conjugate mirror filters h[n] for the two wavelet bases

n	$h[n]$ Haar	$h[n]$ Daubechies-4
0	0.70710678	0.48296291
1	0.70710678	0.83651630
2		0.22414387
3		-0.12940952

Fig. 5.35 shows the synthetic 3-D seismic map of the reservoir shown in Fig. 5.11. In this example, the synthetic seismic map was generated by smoothing the permeability field and adding random noises. This seismic map is closely correlated to the permeability distribution. We used two different wavelet bases to analyze the seismic map: the Haar wavelet and the Daubechies-4 wavelet. The conjugate mirror filters $h[n]$ for the two wavelet bases are shown in Table 5.4.

Integration with the Haar Wavelet

Fig. 5.36 shows the wavelet representation of the 3-D seismic map using the Haar wavelet. It is very obvious that most of the wavelet coefficients have very small amplitudes. This fact implies that we can have a good reconstruction with a very small number of wavelet coefficients. Fig. 5.37 shows the nonzero pattern with only 241 wavelet coefficients. Fig. 5.38 shows the reconstructed seismic map with the 241 wavelet coefficients. As we can see, we did not lose much information in the reconstruction. The reconstructed map still keeps the overall geological structure. Because the seismic map is closely correlated with the permeability distribution, we can use the same mask to select the active parameter set in the nonlinear regression.

Figs. 5.41 and 5.42 show the permeability distribution estimation using only the low-level parameters (the first 64), the same as in Section 5.2.3.

Fig. 5.43 shows the permeability distribution estimation using the most significant 102 parameters from the seismic map. In this case, the resolution around the high-permeability strip is very high, even with such a small number of parameters.

Fig. 5.44 shows the permeability distribution estimation using the most significant

241 parameters from the seismic map. Now we have a fairly accurate estimation of the permeability field. Though there are some mismatches with the true permeability distribution, the regression captures the overall structure. Furthermore, the usage of the seismic map helps identify the discontinuities in the reservoir, with a higher resolution.

Fig. 5.39 shows that the objective function decreases very rapidly, even with such a small number of parameters. At the 35^{th} iteration, the objective function is reduced to around 100. Fig. 5.40 shows the number of active parameters and CPU time at each iteration. Because the algorithm keeps the number of active parameters as small as possible, the computational load is very low for all the iterations.

In this example, the Haar wavelet is a very efficient way to reduce the number of parameters, to identify the geological discontinuities, and to avoid the local minima. Nevertheless, the Haar wavelet has some undesirable properties. Because the Haar wavelet is piecewise linear and has a vanishing moment of 1, the Haar wavelet has a very compact support and is not smooth at all. Therefore it is not suitable to use the Haar wavelet in more realistic situations. As shown in Fig. 5.44, the regression yields some unrealistic blocky distributions, and also some straight edges.

Integration with the Daubechies-4 Wavelet

Because the Haar wavelet has some undesirable properties, we investigated the feasibility of another wavelet basis: the Daubechies-4 wavelet. The Daubechies-4 wavelet has a vanishing moment of 2 and is smoother than the Haar wavelet. The inversion procedure is exactly the same as for the Haar wavelet.

Fig. 5.46 shows the Daubechies-4 wavelet representation of the 3-D seismic map. Just as we expected, most of the wavelet coefficients have very small amplitude. Fig. 5.47 shows the nonzero pattern with only 236 wavelet coefficients. Fig. 5.48 shows the reconstructed seismic map with the 236 wavelet coefficients. Obviously, this reconstruction is much better than the reconstruction using the Haar wavelet (Fig. 5.38). No blocky distributions are introduced in this reconstruction.

Figs. 5.51 and 5.52 show the permeability distribution estimation using only the low-level parameters (the first 64). We should notice that the regression results are

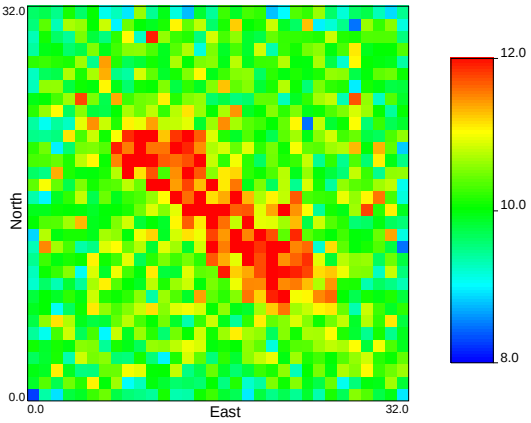

Figure 5.35: Synthetic 3-D seismic data distribution

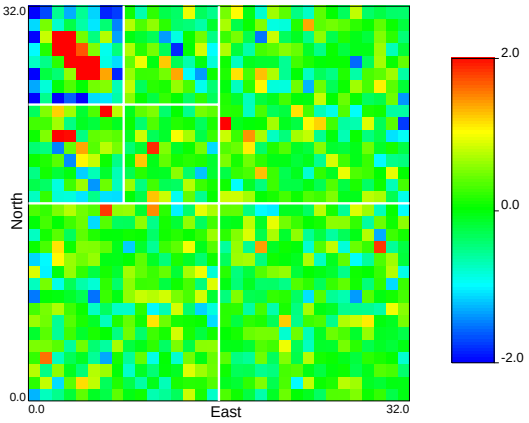

Figure 5.36: Haar wavelet transform of the 3-D seismic data

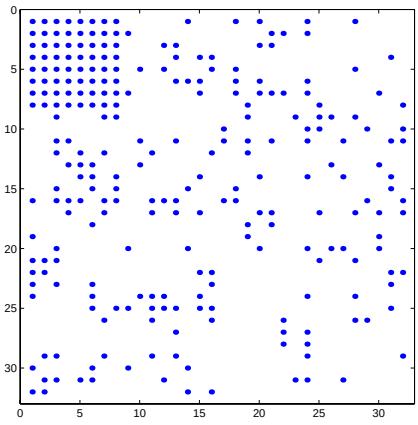

Figure 5.37: Nonzero pattern of the wavelets of 3-D seismic data

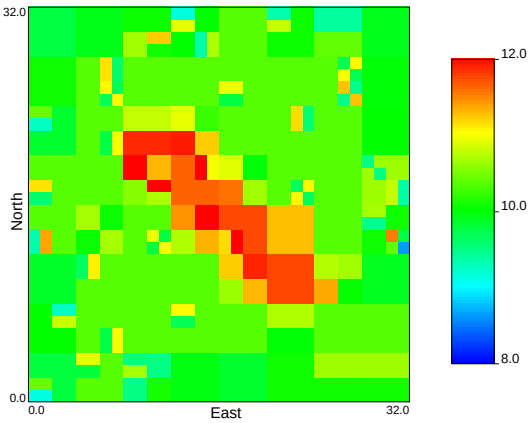

Figure 5.38: 3-D seismic reconstruction

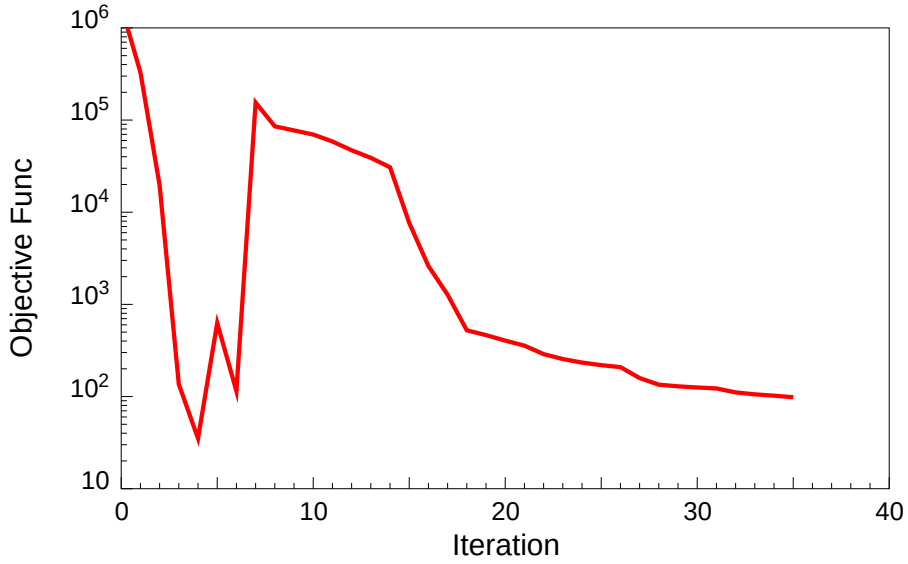

Figure 5.39: Convergence curve (3-D seismic and Haar wavelets, Case 5.2.4)

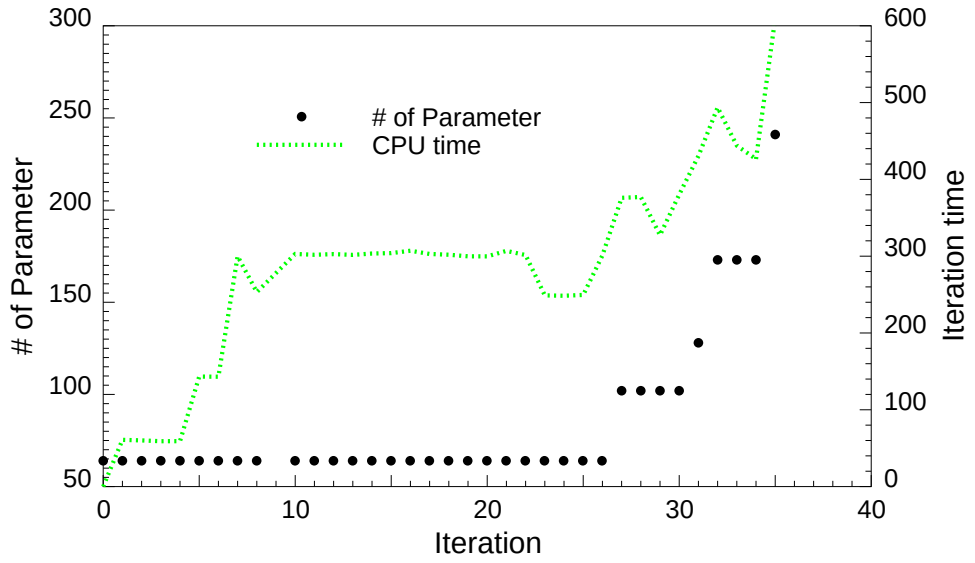

Figure 5.40: Number of parameters and CPU time at each iteration (3-D seismic and Haar wavelets, Case 5.2.4)

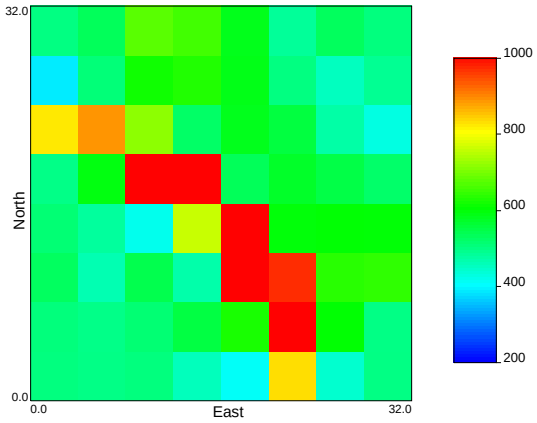

Figure 5.41: Permeability distribution after the 7^{th} iteration (3-D seismic and Haar wavelets, 64 parameters, Case 5.2.4)

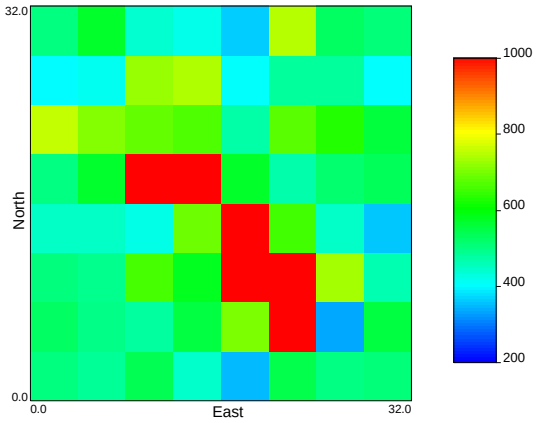

Figure 5.42: Permeability distribution after the 26^{th} iteration (3-D seismic and Haar wavelets, 64 parameters, Case 5.2.4)

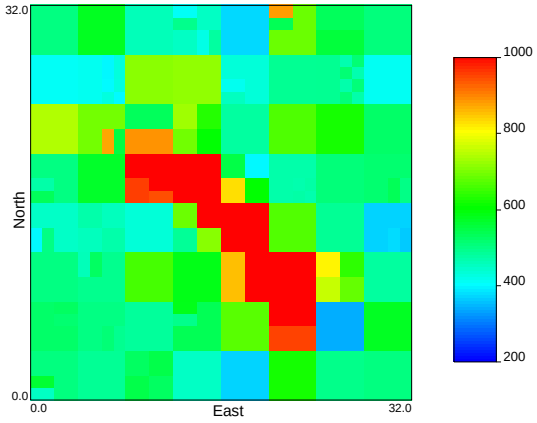

Figure 5.43: Permeability distribution after the 30^{th} iteration (3-D seismic and Haar wavelets, 102 parameters, Case 5.2.4)

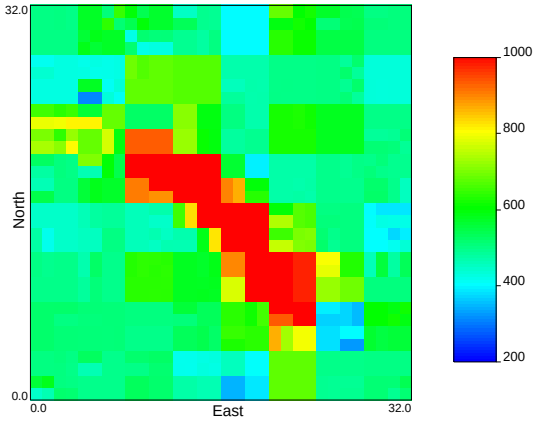

Figure 5.44: Permeability distribution after the 35^{th} iteration (3-D seismic and Haar wavelets, 241 parameters, Case 5.2.4)

not the same as in Section 5.2.3. The reason is that the low-resolution Daubechies-4 wavelet coefficients are not the arithmetic averages of gridblocks. Instead, the wavelet coefficients are certain linear combinations of gridblocks. Fig. 5.52 also shows some variations along the reservoir boundaries. The boundary effects are due to the fact that we used the periodic wavelets to decompose the image over an interval (Appendix A.2.2). This construction has the disadvantage of creating high-amplitude wavelet coefficients along the boundaries, because the border wavelets have separate components with no vanishing moments. This problem can be solved either using folded wavelets or boundary wavelets. In this study, we did not attempt to apply either of these techniques.

Fig. 5.53 shows the permeability distribution estimation using only 97 parameters. The regression has already captured the shape of the high-permeability stripe. At the same time, there are still some local variations in the estimation. Fig. 5.53 shows the permeability distribution using 236 parameters picked by the seismic map shown in Fig. 5.47. Now we have a fairly accurate estimation of the permeability field.

Fig. 5.49 shows a similar convergence behavior as for the Haar wavelet. The objective function can be reduced to below 100, even with a very small number of parameters. The Daubechies wavelet also has a slower convergence rate, compared to the Haar wavelet. Fig. 5.50 shows the number of active parameters and CPU time at each iteration.

In this example, the Daubechies wavelet is also a very efficient way to reduce the number of parameters, to identify the geological discontinuities, to avoid the local minima. Furthermore, the Daubechies wavelet avoids the unrealistic blocky distribution and straight edges. However, because we used the periodic wavelets, the implementation introduced some undesirable boundary effects. Another disadvantage of the Daubechies wavelet is that the true permeability distribution has some very sharp discontinuities. It is not very easy for a smooth wavelet like the Daubechies-4 wavelet to approximate a strong discontinuity.

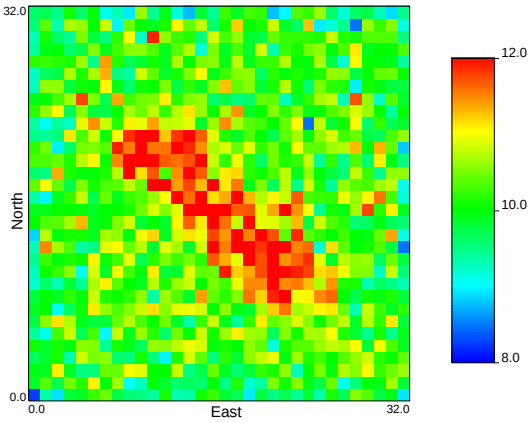

Figure 5.45: 3-D seismic data distribution

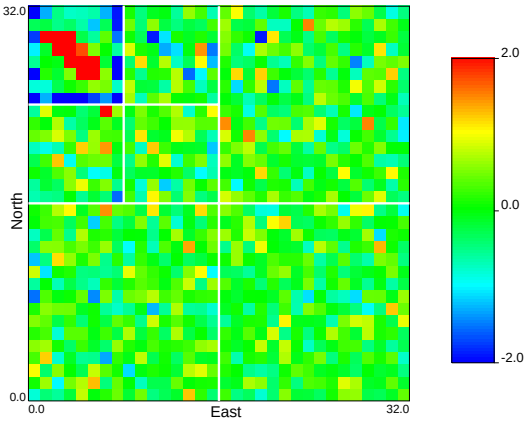

Figure 5.46: Daubechies-4 wavelet transform of the 3-D seismic data

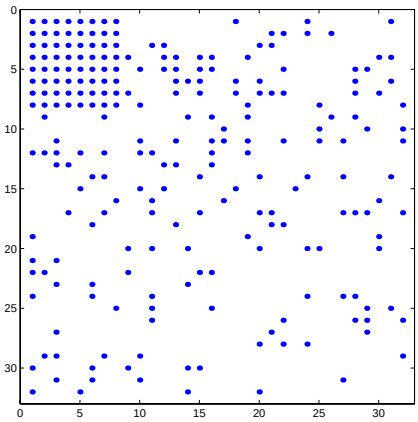

Figure 5.47: Nonzero pattern of the wavelets of 3-D seismic data

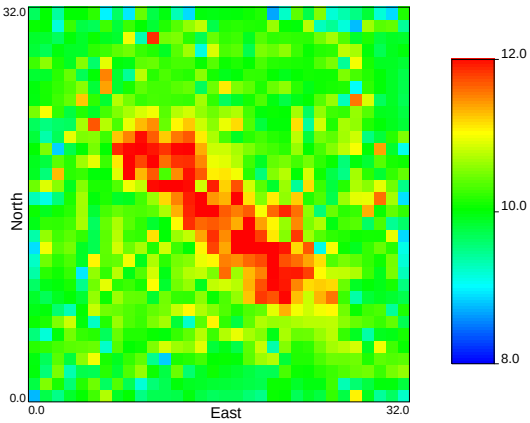

Figure 5.48: 3-D seismic reconstruction

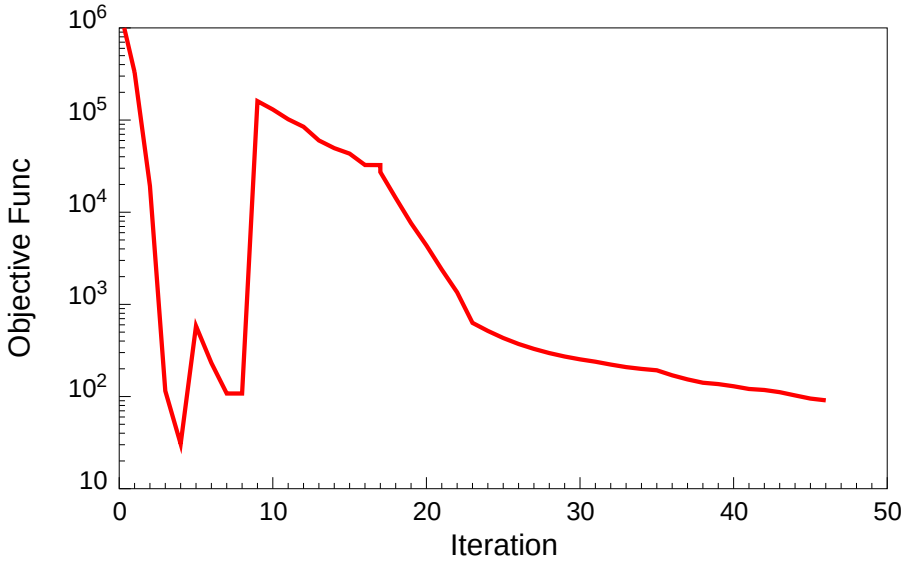

Figure 5.49: Convergence curve (3-D seismic and Daubechies-4, Case 5.2.5)

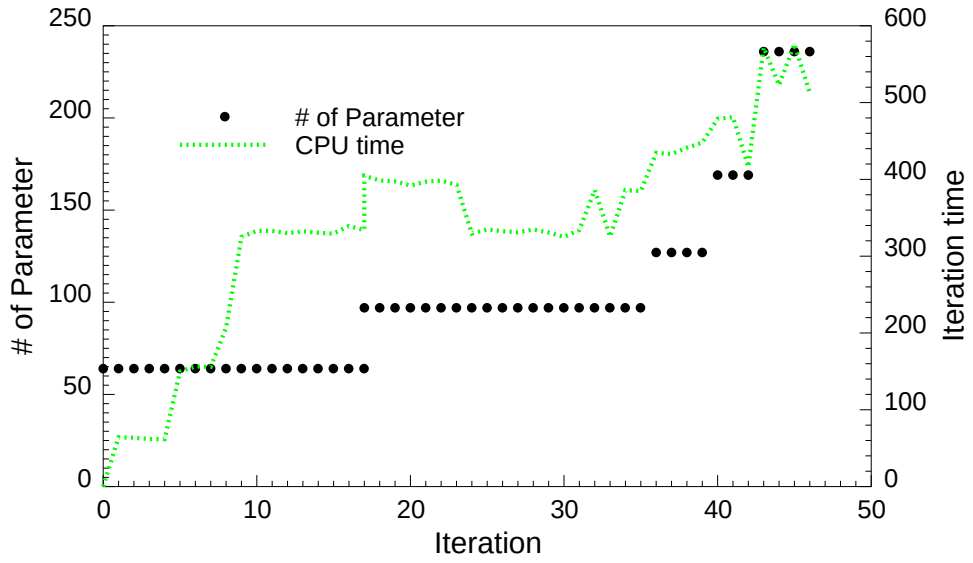

Figure 5.50: Number of parameters and CPU time at each iteration (3-D seismic and Daubechies-4 wavelets, Case 5.2.5)

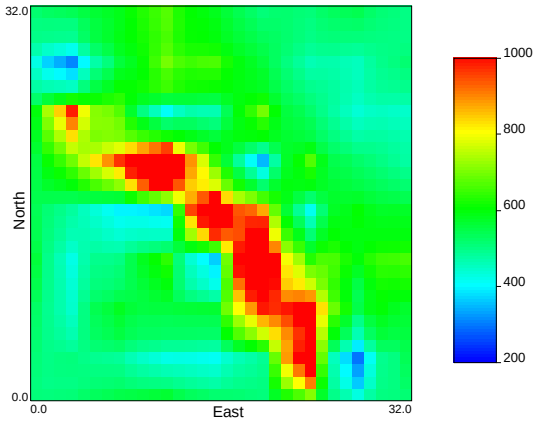

Figure 5.51: Permeability distribution after the 7^{th} iteration (3-D seismic and Daubechies-4 wavelets, 64 parameters, Case 5.2.5)

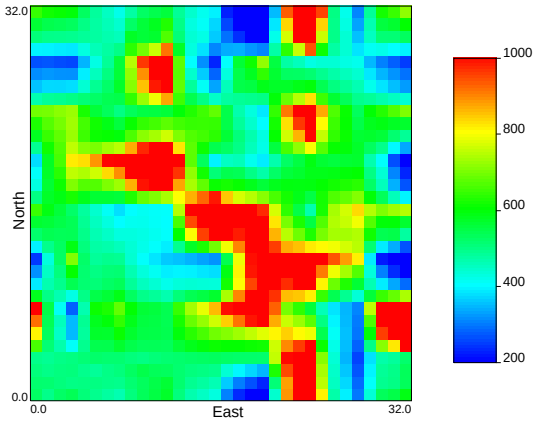

Figure 5.52: Permeability distribution after the 17^{th} iteration (3-D seismic and Daubechies-4 wavelets, 64 parameters, Case 5.2.5)

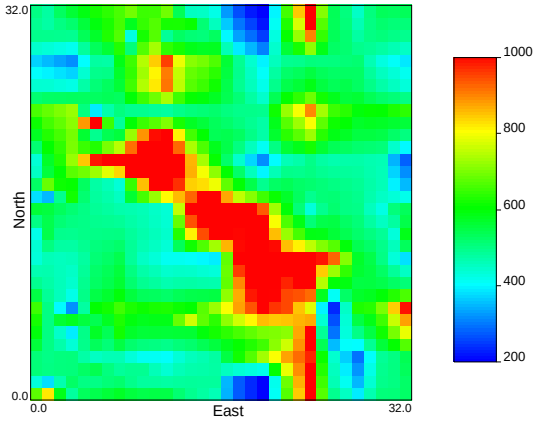

Figure 5.53: Permeability distribution after the 35^{th} iteration (3-D seismic and Daubechies wavelets, 97 parameters, Case 5.2.5)

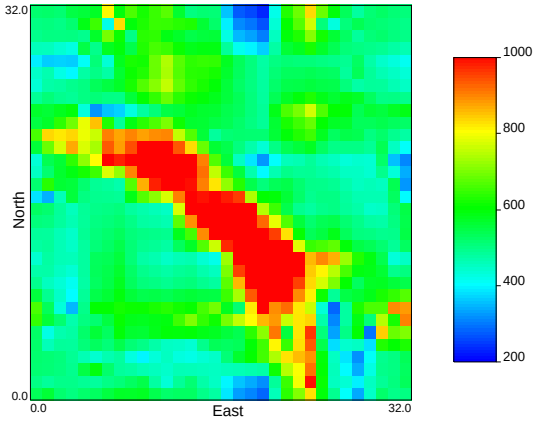

Figure 5.54: Permeability distribution after the 46^{th} iteration (3-D seismic and Daubechies wavelets, 236 parameters, Case 5.2.5)

Table 5.5: Comparison for the test cases in Section 5.2

Case No.	# of iteration	total CPU time (min.)	# of unknowns
5.2.1	37	1104.5	1024
5.2.2	28	243.3	$64 \sim 1024$
5.2.3	35	200.0	$64 \sim 331$
5.2.4	35	165.5	$64 \sim 241$
5.2.5	46	269.0	$64 \sim 236$

5.2.5 Summary of Different Approaches

Table 5.5 shows the performance of different approaches. The performance of each approach can be summarized as follows:

- *Estimated permeability distribution*: All the approaches yielded very good estimation of permeability distribution, though with different characteristics. The pixel modeling gave a realization with uniform fine-resolution. However, it also gave some unnecessary details. The level-by-level approach and the Haar wavelet gave quite similar results. They all had the tendency to yield blocky distributions. The Daubechies wavelet gave a more realistic distribution, but with some undesirable boundary effects.

- *Number of iterations*: There were no significant differences in the number of iterations. The number of iterations ranges from 28 to 46. The Daubechies approach took longer to converge, because it is not very suitable to approximate a strong discontinuity.

- *Computational time*: It is very obvious that the pixel modeling is the most inefficient approach, because it uses all the 1024 parameters in each iteration. There were no substantial differences among the four wavelet approaches. This is mainly due to the size of the reservoir discretization. For larger reservoirs, the nonuniform-resolution wavelet approach should hold advantages over the level-by-level approaches. At very high resolution levels, it is very difficult to use the level-by-level approach.

- *Feasibility for large reservoirs*: There is no doubt that the pixel modeling is not applicable for very large reservoirs, as the computational load is just too great to be practical. The level-by-level approach solved this problem partially by accounting for the data resolution with different resolution levels. But this method is still very inefficient when the number of parameters is large. The nonuniform-resolution wavelet approach can use either the variance analysis or correlated data to reduce the number of parameters down to a reasonable level. Even with a very small number of parameters, the nonuniform-resolution wavelet approaches could still capture the overall geological structures.

- *Data integration*: All the approaches can integrate dynamic data: bottomhole pressures, watercuts, and saturation change distribution. The wavelet approach also provides a powerful tool to correlate with other static data. As shown in this study, we can effectively embed the 3-D seismic map into the inversion procedure. This approach is not limited only to 3-D seismic data, other geological information can be integrated as easily.

Chapter 6

Applications and Test Cases

In Chapter 5, we have outlined the basic methodologies for reservoir parameter estimation using the wavelet analysis. A simple test case has been used to test different methods. In this chapter, several additional test problems are presented to further investigate the multiresolution approach. In this chapter, except for the permeability and porosity distribution, all the other relevant reservoir and fluid properties are the same as defined in Section 5.2. In all the test cases, the objective functions have been reduced to very small values. Hence only the observed data are plotted in this chapter, because there is almost no visible mismatch between the observed and calculated data.

6.1 Effects of Noisy Data

In Section 5.2, we used the "perfect" data for all the test cases. The perfect data were generated from the true property distribution, with no model or measurement errors. In most test cases, we were able to reduce the objective functions to very small values. However, this is not the case in practice. The observed data always include measurement errors, noise, etc. Furthermore, the forward mathematical model (the reservoir simulator) has inherent errors, such as discretization errors, dispersion errors, etc. When the system is under-determined, which is usually the case in reservoir parameter estimation, parameters are sensitive to these errors. In other words, small

errors in the observed data may cause large errors in estimated parameters. In this study, the objective function is defined as

$$E = \left(\vec{d}^{cal} - \vec{d}^{obs}\right)^T C_D^{-1} \left(\vec{d}^{cal} - \vec{d}^{obs}\right). \tag{6.1}$$

The mean value of the objective function is $nData$, the number of data. Therefore, in real cases, we can not reduce the objective function to a value significantly smaller than $nData$. Even if we do have a very small objective function, we might have overfitted the observed data.

In order to investigate the impact of errors in data, we set up a test case in which all the observed data have errors associated with them. The reservoir configuration is exactly the same as the one in Section 5.2. The only modification is that we added some errors into the data as follows:

- *Bottomhole pressure*: Random noises with a standard deviation of 1.0 psi.
- *Watercut*: Random noises with a standard deviation equivalent to 1% of the amplitude.
- *Water saturation change*: The saturation change was remapped with a resolution of 0.1:

  $$\Delta S_w^{new} = \lfloor \Delta S_w / 0.1 + 0.5 \rfloor \times 0.1, \tag{6.2}$$

  which is shown in Fig. 6.1.
- *3-D seismic map*: No additional noises. Fig. 6.2 shows the 3-D seismic map.

We repeated the five different cases from Section 5.2, with the new data with artificial errors. The results are summarized as follows:

1. *Pixel modeling*: Fig. 6.3 shows the permeability distribution only using the early-time data. There are some local variations around wellbores, especially around the injecting well. Fig. 6.4 shows the final regression result, which is not as good as the estimation without data errors (Fig. 5.22). There are many unresolved parameters in the pixel modeling. In this example, the system is under-determined and the unresolved parameters are very sensitive to errors in the observed data.

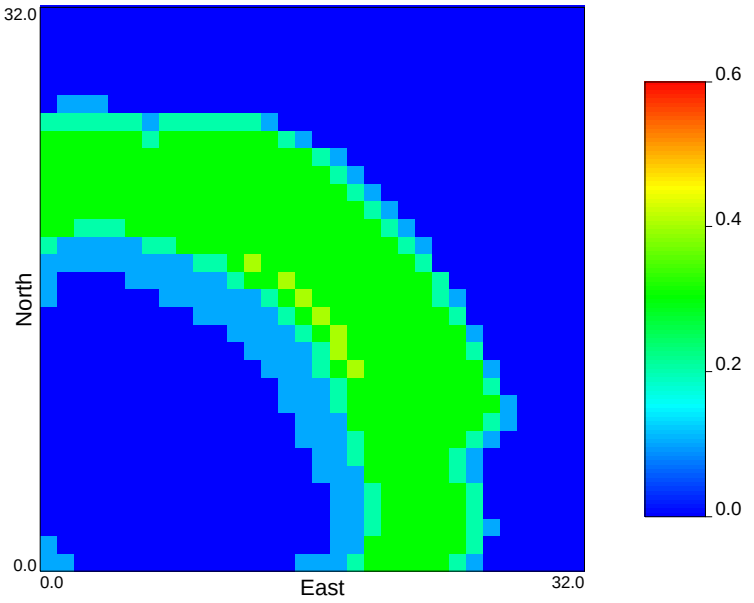

Figure 6.1: Low-resolution water saturation change map

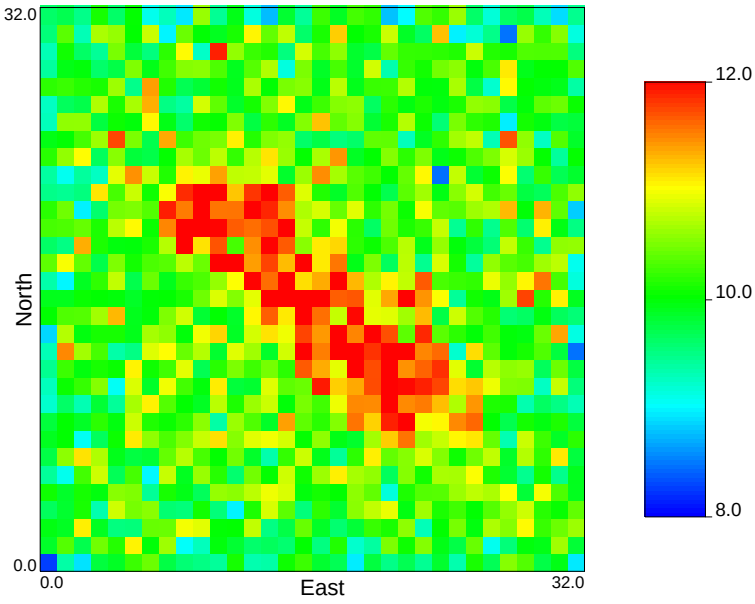

Figure 6.2: Synthetic 3-D seismic map

2. *Level by level*: Fig. 6.5 shows the permeability distribution using only 64 low-resolution parameters. Even though there are errors in the observed data, the data still has enough resolving power to resolve most of the low-resolution parameters. However, at the high-resolution level, as shown in Fig. 6.6, the algorithm was not able to resolve most of the fine-scale parameters.

3. *Nonuniform-resolution Haar wavelet*: Fig. 6.7 shows the permeability distribution using the Haar wavelet with only 331 parameters. Because the system is much smaller than the original one, and most of the parameters are well defined with low uncertainty, the algorithm was able to resolve most of the parameters with high certainty.

4. *Nonuniform-resolution Haar wavelet with 3-D seismic data*: Fig. 6.8 shows the permeability distribution using the 3-D seismic mask. The 3-D seismic map provides some spatial information for the reservoir, therefore, we were able to resolve some fine-scale parameters in the middle of reservoir.

5. *Nonuniform-resolution Daubechies-4 wavelet with 3-D seismic data*: This test case is not much different from the previous one with the Haar wavelet. Figs. 6.9 and 6.10 show the permeability distributions using 64 low-resolution parameters and 258 parameters respectively.

Regarding the sensitivity to data errors, the nonuniform-resolution wavelet approaches hold an advantage over the pixel modeling and the uniform-resolution level-by-level approach. The wavelet approaches are able to produce reliable estimations even with noisy observed data. These wavelet approaches can account for the data resolution using different resolutions in different regions. The wavelet approaches also stabilize the inversion in three main ways: reducing the number of parameters, integrating as much data as possible, and keeping the uncertainty as low as possible.

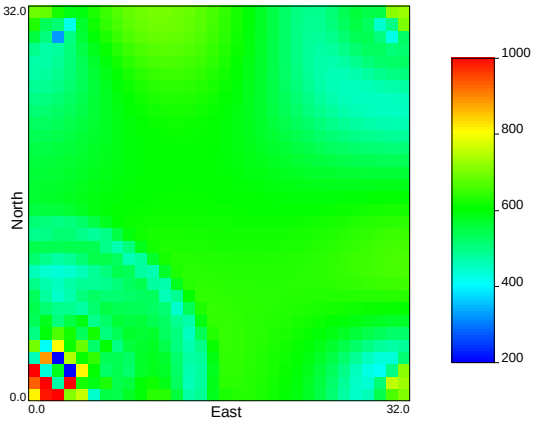

Figure 6.3: Permeability distribution after the 5^{th} iteration (pixel modeling, Case 6.1.1)

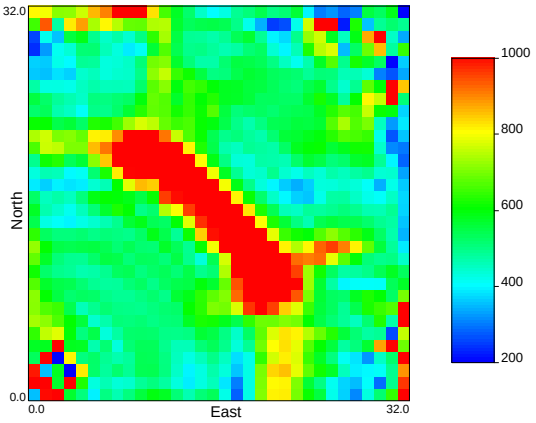

Figure 6.4: Permeability distribution after the 24^{th} iteration (pixel modeling, Case 6.1.1)

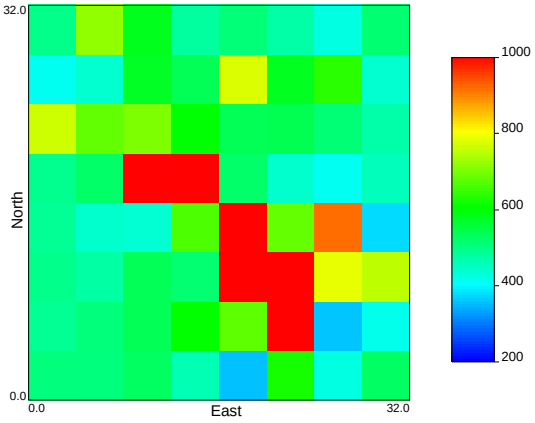

Figure 6.5: Permeability distribution after the 12^{th} iteration (level by level, Case 6.1.2)

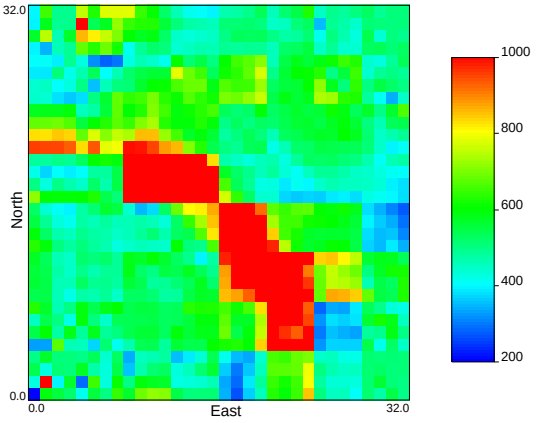

Figure 6.6: Permeability distribution after the 16^{th} iteration (level by level, Case 6.1.2)

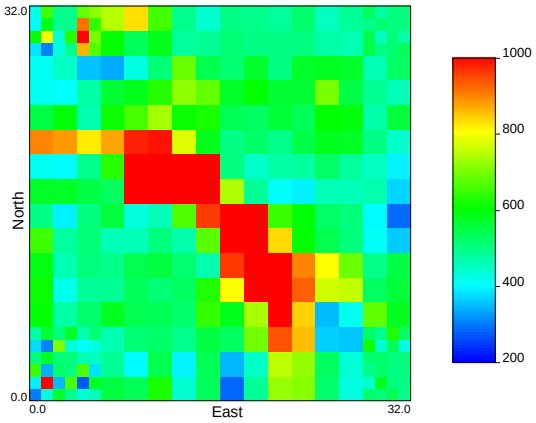

Figure 6.7: Permeability distribution after the 22^{th} iteration (Haar Wavelet, Case 6.1.3)

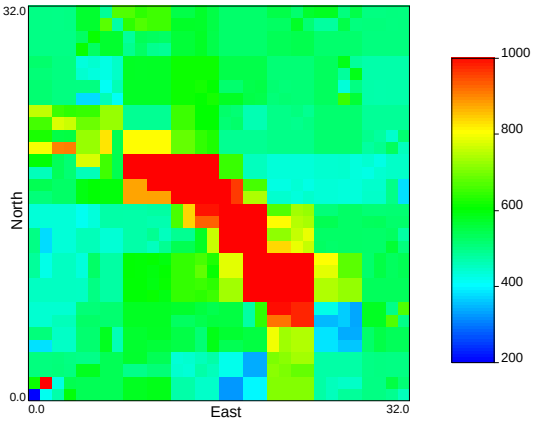

Figure 6.8: Permeability distribution after the 23^{th} iteration (Haar wavelet (3-D seismic), Case 6.1.4)

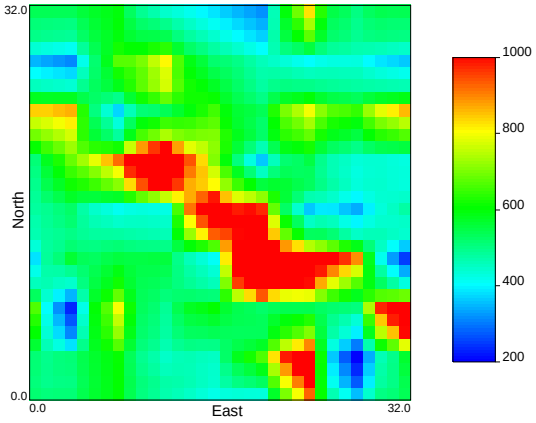

Figure 6.9: Permeability distribution after the 20^{th} iteration (Daubechies-4 wavelet, Case 6.1.5)

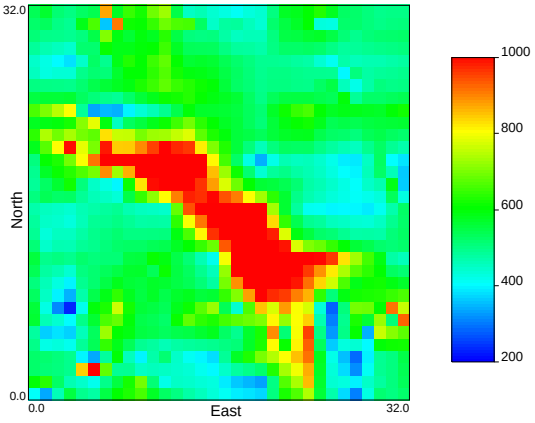

Figure 6.10: Permeability distribution after the 30^{th} iteration (Daubechies-4 wavelet, Case 6.1.5)

6.2 A Reservoir with a Fault

In this example, we tested a different reservoir configuration. Fig. 6.11 shows the true reservoir permeability distribution: a low-permeability strip with $k \approx 25$ md is located in the middle of the reservoir, while the rest of the reservoir has a mean permeability of $\bar{k} = 500$ md. There are four active wells in the reservoir, with grid indices (3,3), (3,29), (29,3), (29,29) respectively. The well at the left-bottom corner is the injecting well with an injection rate at 10000 STB/day, the rest are all producing wells with production rates at 2000, 2000, 6000 STB/day respectively. The elapsed time is 500 days and all the producing wells have water breakthrough. Information for both the bottomhole pressure and watercut at each well is available. Figs. 6.13 through 6.16 show the production history at each well. We also assumed that water saturation change distribution is available. Fig. 6.12 shows the water saturation change between $t = 100$ days and $t = 200$ days.

This is a very challenging problem, due to the very special reservoir configuration. Because the permeability of the strip is so low, it can almost be treated as a no-flow fault. The downstream side of the fault is not well defined, because the amount of fluid flowing across this strip is very small. Furthermore, Wells #2 and #3 break through very early, while Well #4 breaks through after $t = 400$ days.

We used two different methods to estimate the permeability distribution: the pixel modeling and the wavelet approach.

1. *Pixel Modeling*: Fig. 6.17 shows the permeability distribution with only 150-day production history. Even with only the bottomhole pressure data, the algorithm can "see" the existence of the low-permeability region. However, the region downstream of the fault has very low sensitivities and does not have much impact on the observed data. Fig. 6.18 shows the final permeability distribution with all the data available: bottomhole pressures, watercuts, and the saturation change distribution. Though the upstream side of the fault is reasonably well defined, the downstream side remains unresolved. The variance distribution map (shown in Fig. 6.21) indicates different impacts of the dynamic data. The observed data from wells can only resolve the parameters around wellbores.

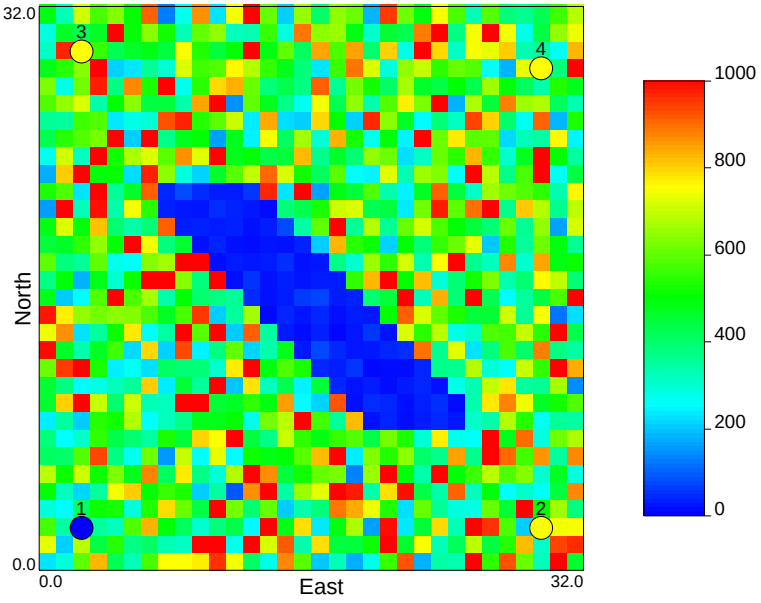

Figure 6.11: Reservoir permeability distribution (32×32 gridblocks)

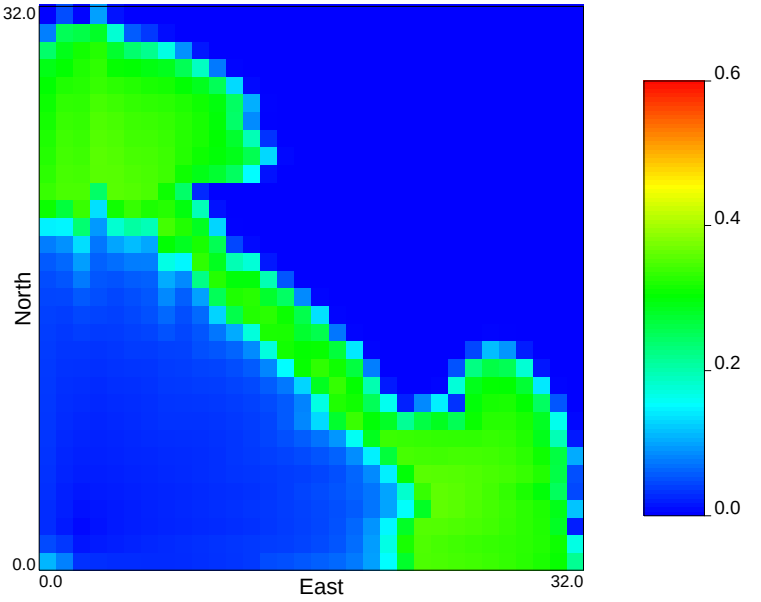

Figure 6.12: Water saturation change between $t = 100$ and 200 days

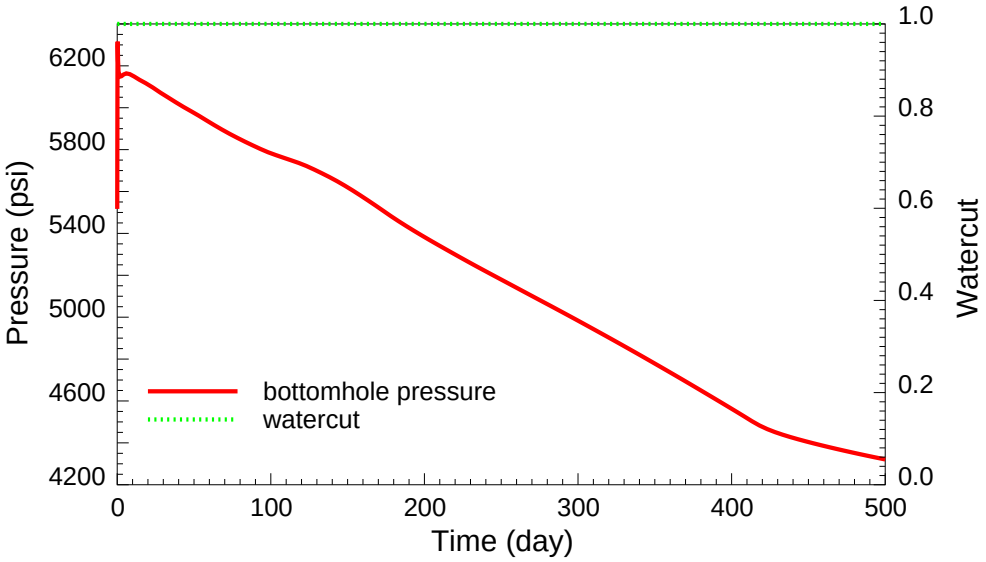

Figure 6.13: Bottomhole pressure and watercut at Well #1

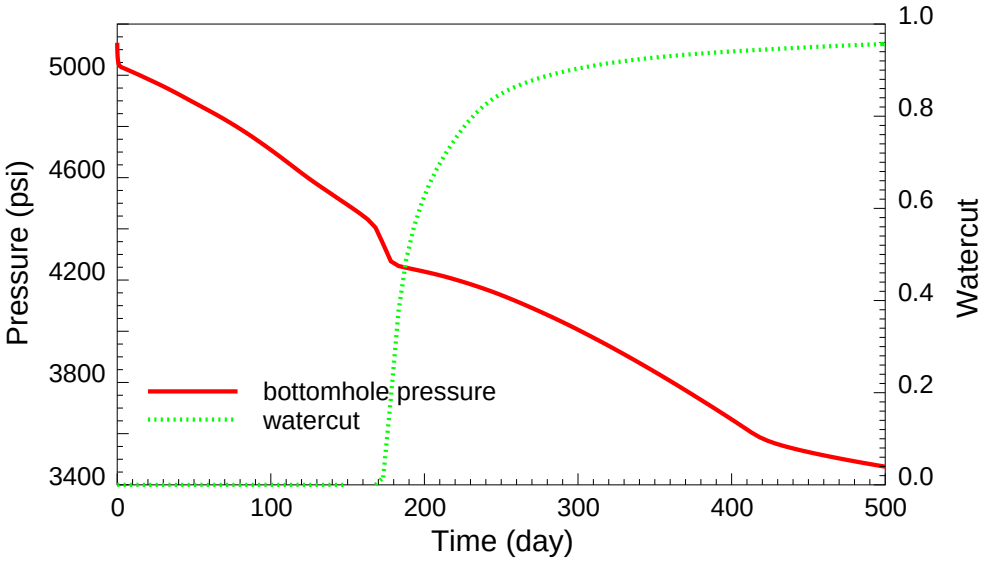

Figure 6.14: Bottomhole pressure and watercut at Well #2

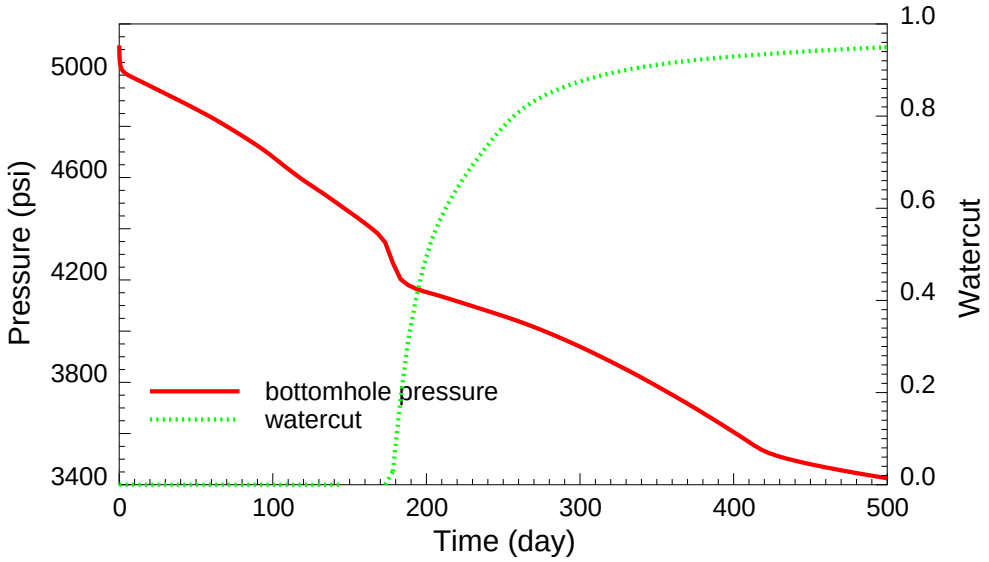

Figure 6.15: Bottomhole pressure and watercut at Well #3

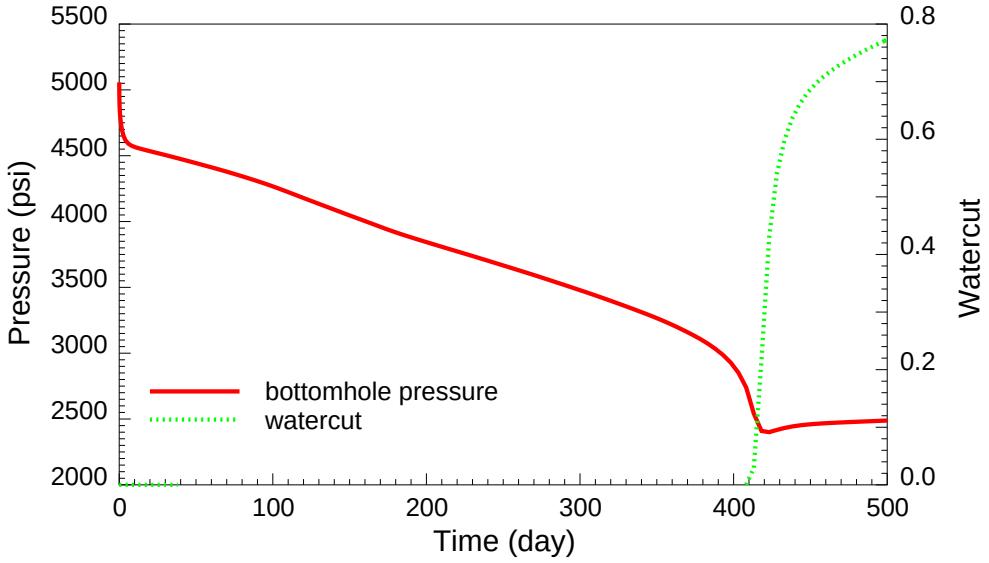

Figure 6.16: Bottomhole pressure and watercut at Well #4

These data can not constrain the parameters in middle of the reservoir. The saturation change information helps resolve the upstream side of the fault. None of the dynamic data has sufficient resolving power for the downstream side of the fault.

2. *Wavelet Approach*: Fig. 6.19 shows the estimate of the permeability distribution with 91 low-resolution parameters. Because there are some heterogeneities around the wellbores, we need to do some local refinement in the wellbore neighborhoods. With only 91 parameters, the upstream side of the fault is very well defined, with the downstream unresolved. Fig. 6.20 shows the final permeability distribution with 325 parameters. Obviously, there is not much improvement for the downstream side of the fault. The variance distribution map (Fig. 6.22) shows that the observed data can not resolve any high-resolution parameters with high certainty, except for the low-resolution parameters and wellbore neighborhoods,

Fig. 6.24 shows the variances of parameters for the two approaches. The pixel modeling can only resolve a small number of parameters with high certainty. The wavelet analysis performs much better than the pixel modeling. In both cases, the observed data do not allow us to resolve all the parameters. Fig. 6.23 shows the total uncertainties for the two approaches. For a small number of parameters, the total uncertainty of the wavelet approach is substantially lower than that of the pixel modeling. For instance, using 100 parameters, the total uncertainty of the wavelet approach is only 13.06, while the pixel modeling gives a total uncertainty of 27.42.

6.3 A Reservoir without 4-D Seismic Information

So far, we have only used test cases with simple geometries. In this section, a more realistic reservoir configuration is tested. Fig. 6.25 shows the true reservoir permeability distribution. There are six active wells in the reservoir, as shown in Fig. 6.26. Well #1 is injecting with a constant injection rate at 10200 STB/day. Wells #2, #3,

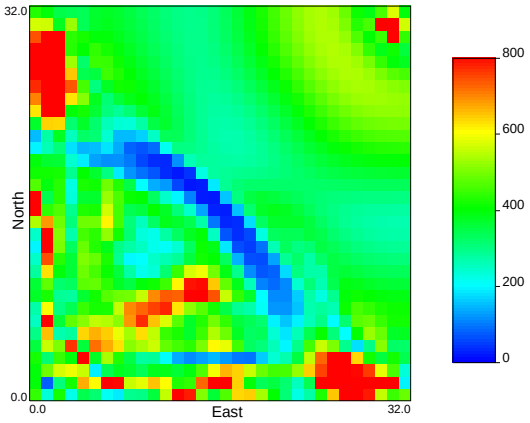

Figure 6.17: Permeability distribution with 150-day history(pixel modeling, Case 6.2.1)

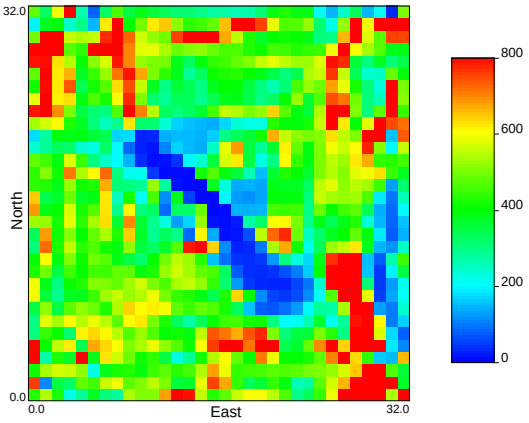

Figure 6.18: Permeability distribution with all observed data (pixel modeling, Case 6.2.1)

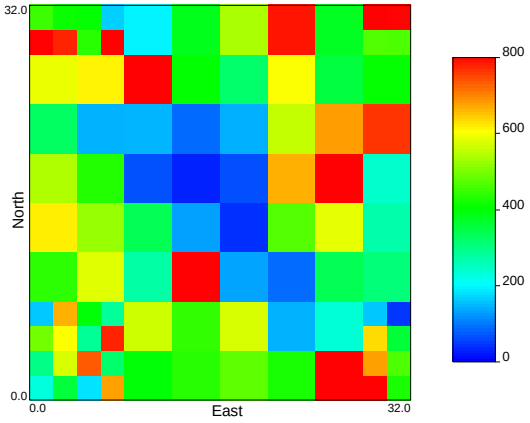

Figure 6.19: Permeability distribution with 91 parameters (Haar wavelet, Case 6.2.2)

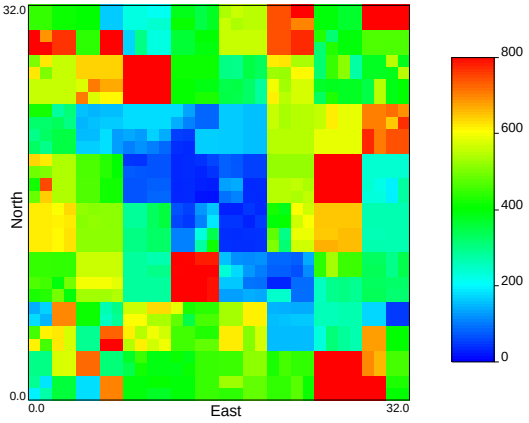

Figure 6.20: Permeability distribution with 325 parameters (Haar wavelet, Case 6.2.2)

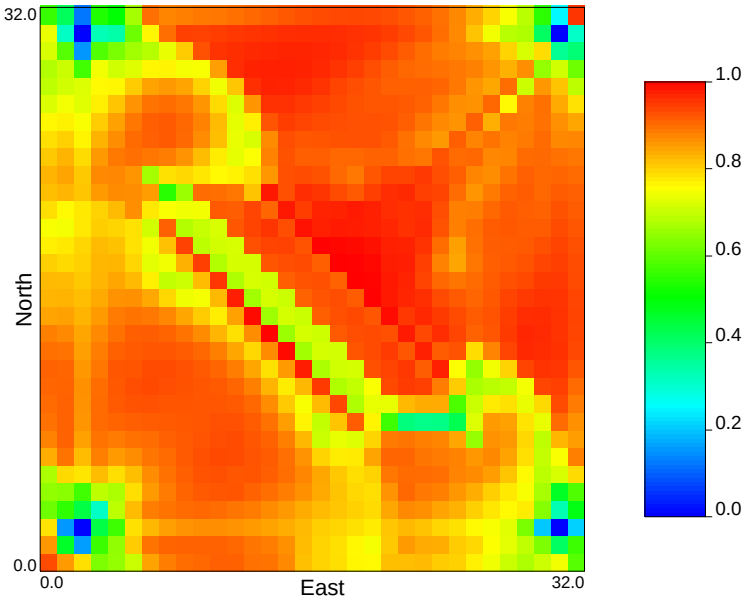

Figure 6.21: Variance distribution map for pixel modeling

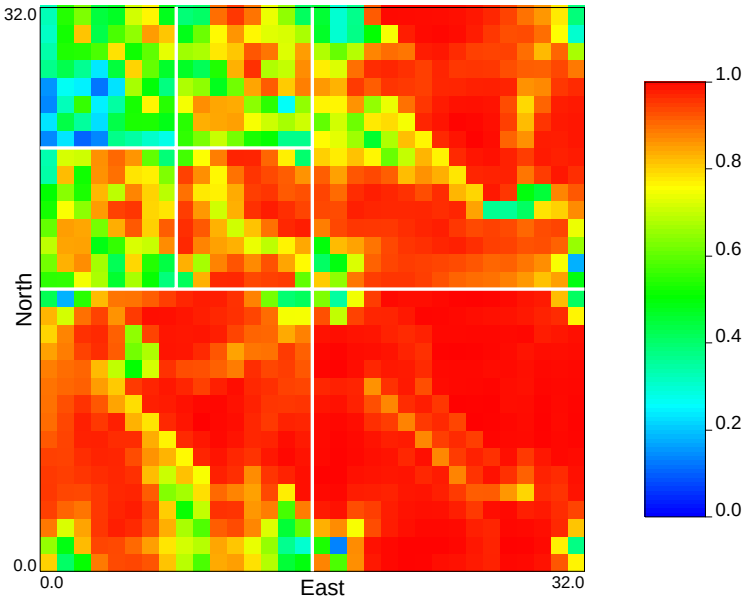

Figure 6.22: Variance distribution map for Haar wavelets

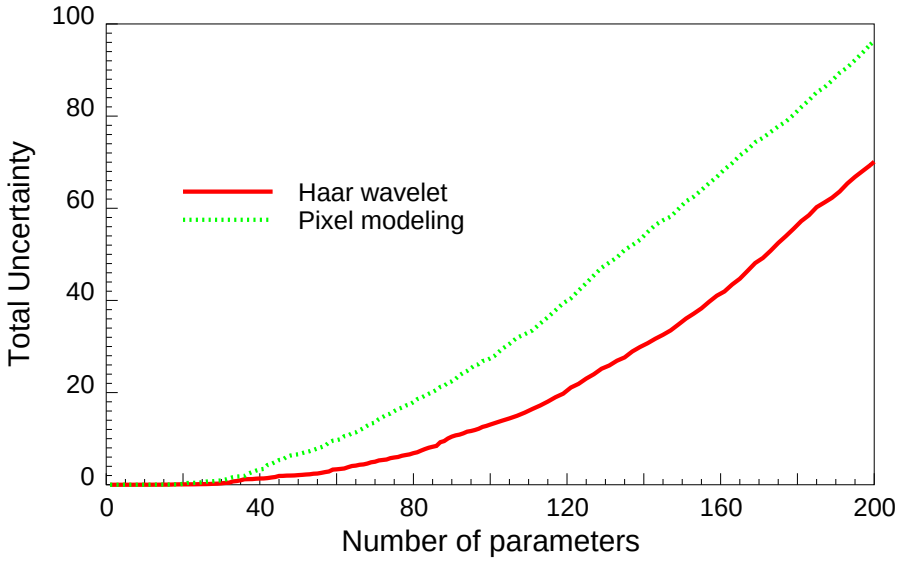

Figure 6.23: Total uncertainty curves

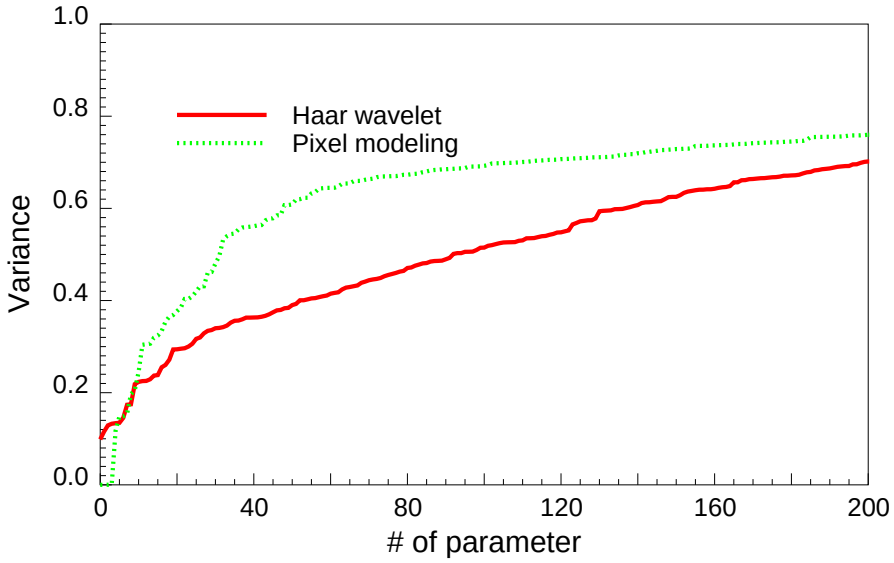

Figure 6.24: Variance of parameter estimation

#4 are producing with total production rates at 2000, 2000, 6000 STB/day respectively. Wells #5, #6 are observation wells, each with a low production rate at 100 STB/day.

The elapsed time is 400 days and all the producing wells have water breakthrough. Information for both the bottomhole pressure and watercut at each well is available. We also assumed that the water saturation change distribution (from 4-D seismic) is not available. Instead, we had a 3-D seismic interpretation, as shown in Fig. 6.27.

In this test case, two different wavelet bases were used. For each wavelet basis, two different runs were performed, with and without 3-D seismic data, respectively. The results are summarized as follows:

- *Haar wavelet*: Figs. 6.28 and 6.29 show the results for nonlinear regression with and without the 3-D seismic information. There is no substantial difference between the two runs, because this reservoir is well defined with six active wells in it. Even without 3-D seismic data, the parameters can be resolved very well. However, the 3-D seismic data does help reduce the number of parameters. The regression integrating the 3-D seismic data used only 223 parameters to reconstruct the reservoir permeability distribution, while the other method used 382 parameters. Both methods yield unrealistic blocky distributions, due to the fact that the Haar wavelet has a very compact support and a vanishing moment of 1.

- *Daubechies-4 wavelet*; Figs. 6.30 and 6.31 show the results for nonlinear regression with the Daubechies-4 wavelet, with and without 3-D seismic data. The 3-D seismic data does not have much impact in this case either. The estimated permeability distribution is more realistic in appearance than those obtained using the Haar wavelet.

6.4 Larger Problem

In this example, the algorithms were applied to a problem involving a larger reservoir, with more complex configurations. Figs. 6.32 and 6.33 show the true reservoir

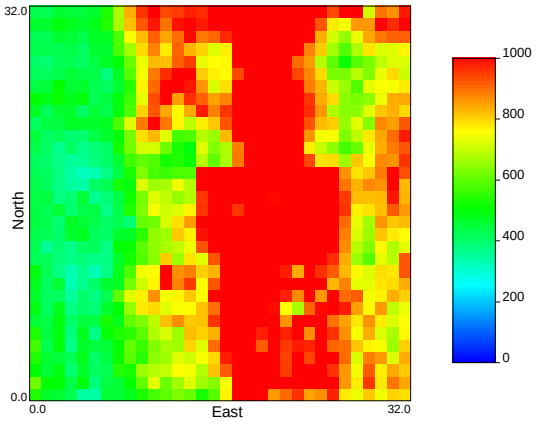

Figure 6.25: Reservoir permeability distribution

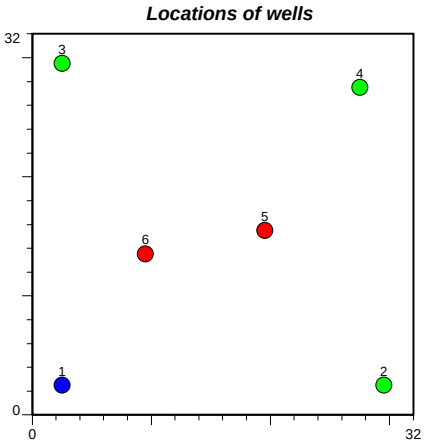

Figure 6.26: Well locations

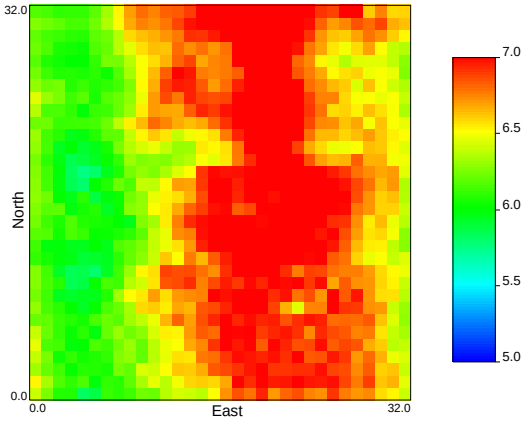

Figure 6.27: Synthetic 3-D seismic map

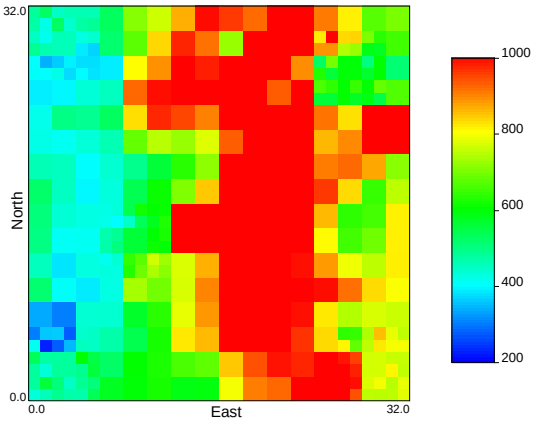

Figure 6.28: Permeability distribution with Haar wavelet, (no 3-D seismic, 382 parameters, Case 6.3.1)

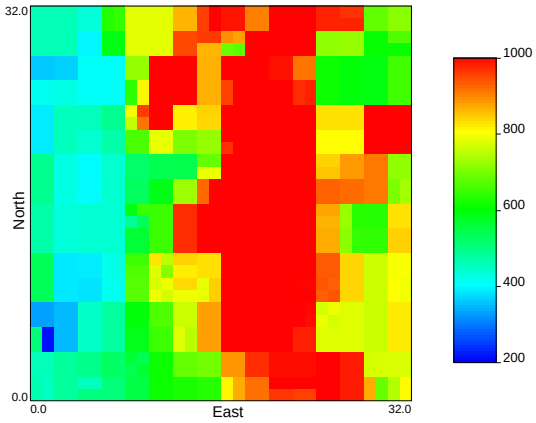

Figure 6.29: Permeability distribution with Haar wavelet, (3-D seismic, 223 parameters, Case 6.3.2)

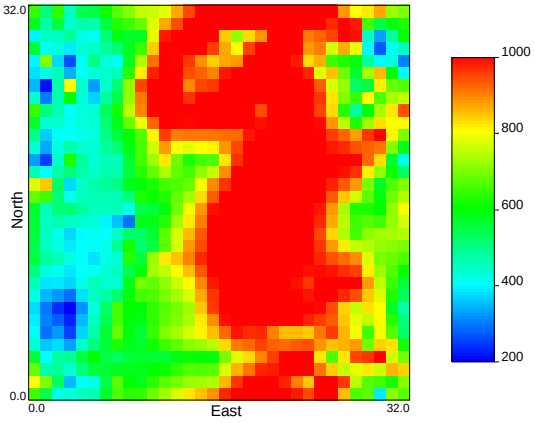

Figure 6.30: Permeability distribution with Daubecheis wavelet, (no 3-D seismic, 208 parameters, Case 6.3.3)

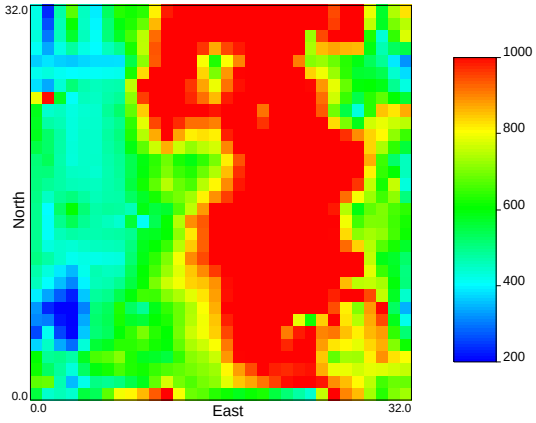

Figure 6.31: Permeability distribution with Daubecheis wavelet, (3-D seismic, 220 parameters, Case 6.3.4)

permeability and porosity distributions respectively. Fig. 6.34 shows the strong correlation between the permeability and porosity fields. Instead of using permeability and porosity as independent parameters, we assumed there is a fixed relationship between them:

$$\phi = \frac{\ln(k) - b}{a}. \tag{6.3}$$

In this example, we set $a = 10.0$ and $b = 2.5$.

There are 11 active wells in the reservoir, as shown in Fig. 6.35. Wells #1 through #4 are injection wells, with bottomhole pressures and injection rates shown in Figs. 6.38 through 6.41. Wells #5 through #11 are producing with constant total production rates at 1000, 1000, 1000, 1000, 1500, 1000, 500 STB/day respectively.

The elapsed time is 800 days and only four producing wells have water breakthrough (Wells #5, #8, #9 and #10). The bottomhole pressures and watercuts for these wells are shown in Figs. 6.42 through 6.45.

In this test case, we assumed the availability of both 3-D seismic and 4-D seismic data. Fig. 6.36 shows the 3-D seismic map, and Fig. 6.37 shows the water saturation change between $t = 400$ days and $t = 600$ days. However, in this particular case, neither of the seismic data has much impact on the parameter estimation. First of all, the 3-D seismic map is only loosely correlated with the reservoir property distribution (permeability and porosity). Therefore, the 3-D seismic data can only provide low-resolution information. Because there are active wells across the whole reservoir, most of the low-resolution information can be provided by the observed data at the wells, even without 3-D seismic data. Secondly, the water sweep area in this example is very small, compared with previous examples. Several producing wells did not have water breakthrough. The 4-D seismic data can only provide some spatial information in very limited regions.

This sample reservoir is under-determined and it was very difficult for us to obtain a reliable estimation. There are a total of 4096 parameters and the correlation between permeability and porosity is not very precise. Nevertheless, the wavelet approach effectively reduced the number of parameters down to 328, as shown in Fig. 6.47. Though we could not obtain any detailed estimation about the reservoir, the algorithm was able to capture the major geological trends. We also had a very good

match in both bottomhole pressures and watercuts at all the wells, with a fairly small residual objective function, as shown in Fig. 6.46. Fig. 6.48 and Fig. 6.49 show the estimations for permeability and porosity fields respectively. Because we only used a reduced number of parameters in the inversion procedure, the estimation can not fully account for the reservoir heterogeneity, resulting in a much smoother realization than the true distribution.

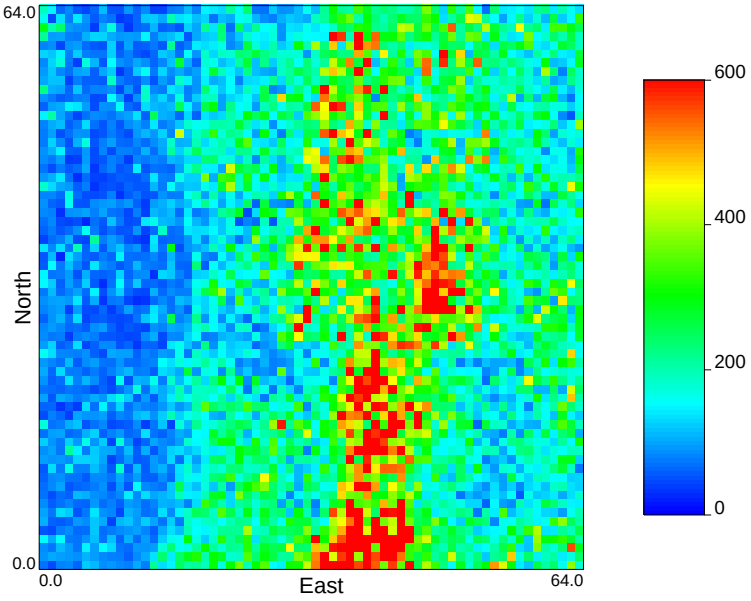

Figure 6.32: Reservoir permeability distribution (64×64 gridblocks)

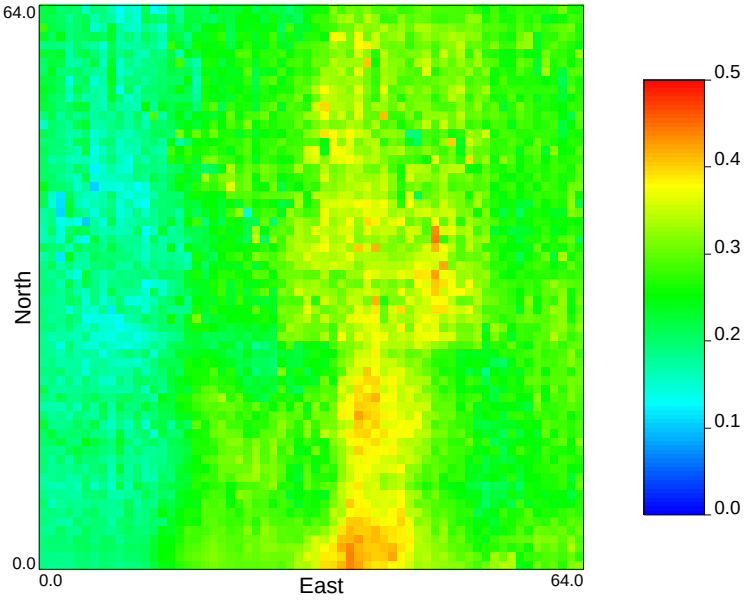

Figure 6.33: Reservoir porosity distribution (64×64 gridblocks)

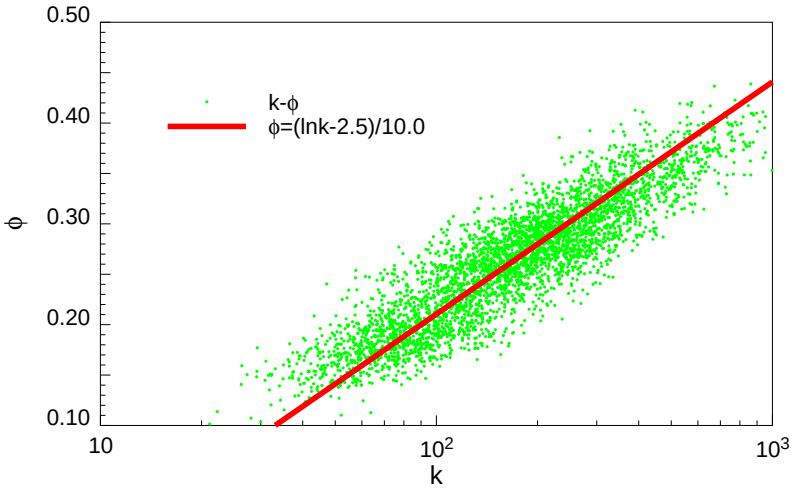

Figure 6.34: Permeability and porosity correlation plot

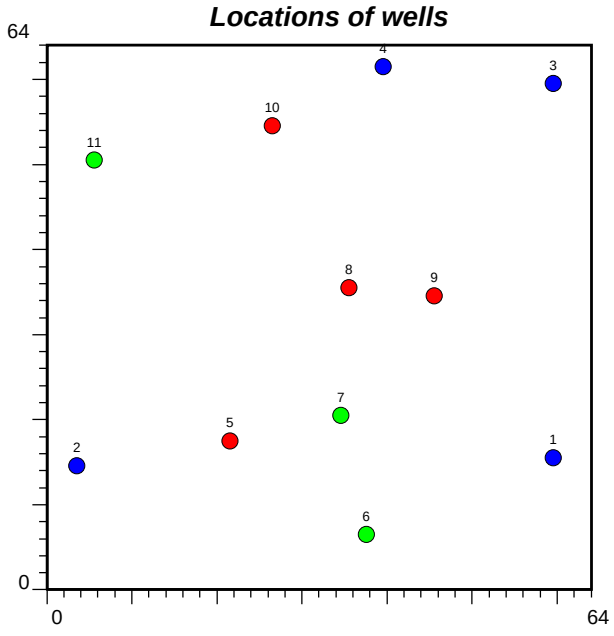

Figure 6.35: Well locations (64×64 gridblocks)

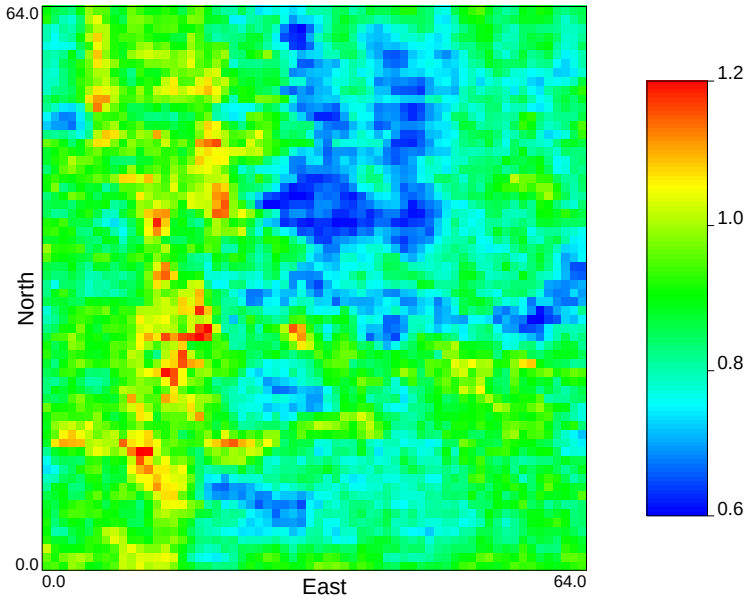

Figure 6.36: 3-D seismic map (high frequency)

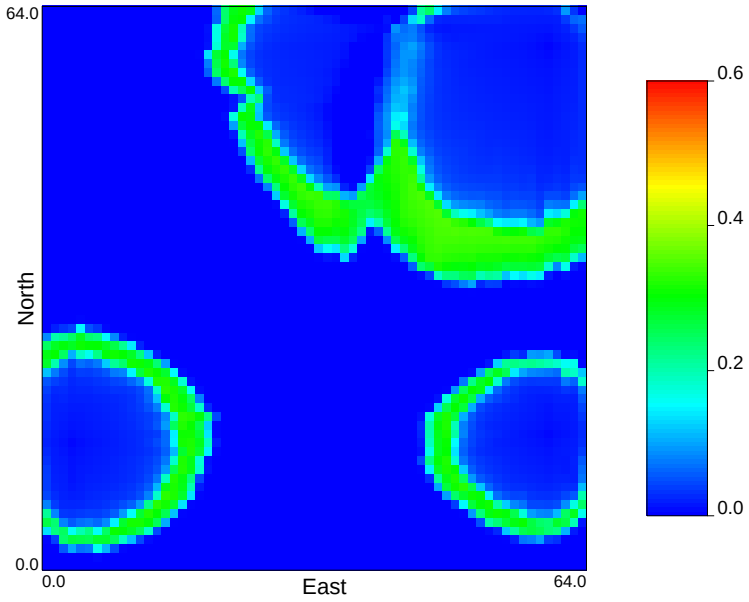

Figure 6.37: Water saturation change between $t = 400$ days and $t = 600$ days

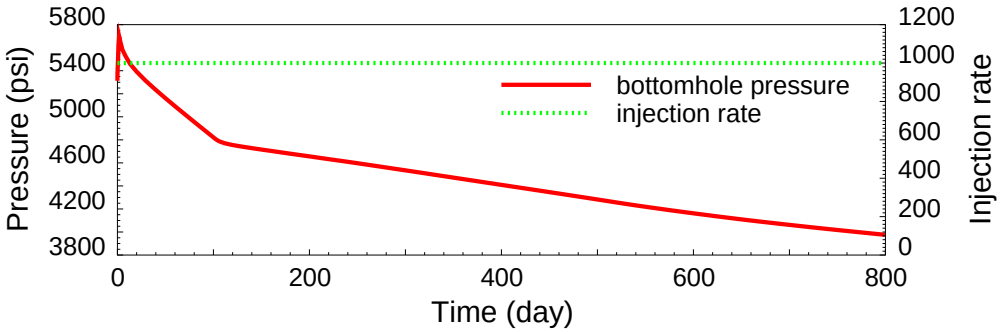

Figure 6.38: Bottomhole pressure and injection rate for Well #1

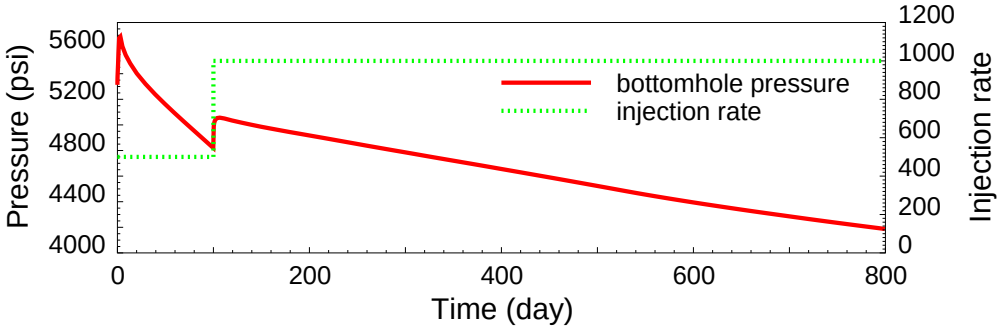

Figure 6.39: Bottomhole pressure and injection rate for Well #2

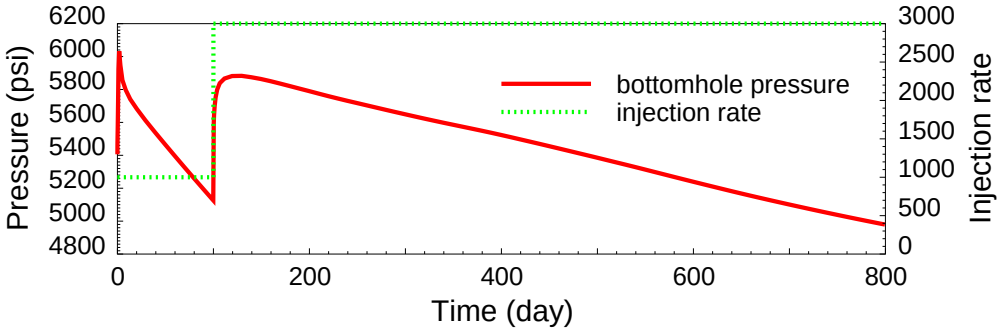

Figure 6.40: Bottomhole pressure and injection rate for Well #3

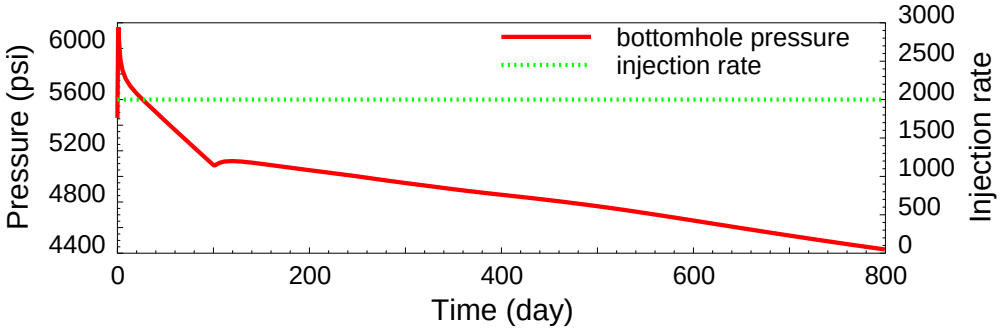

Figure 6.41: Bottomhole pressure and injection rate for Well #4

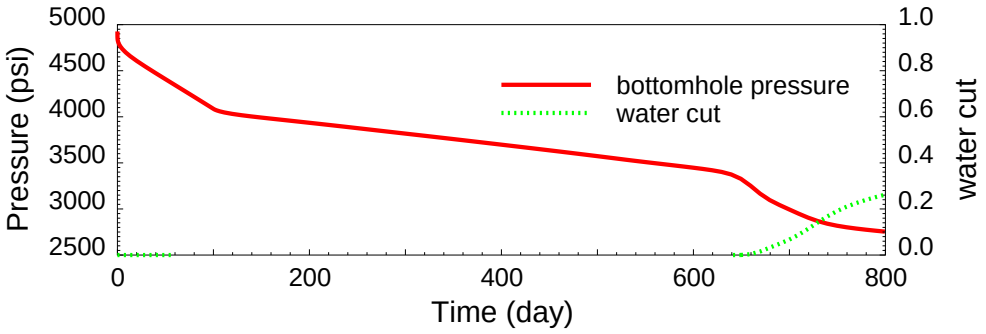

Figure 6.42: Bottomhole pressure and watercut for Well #5

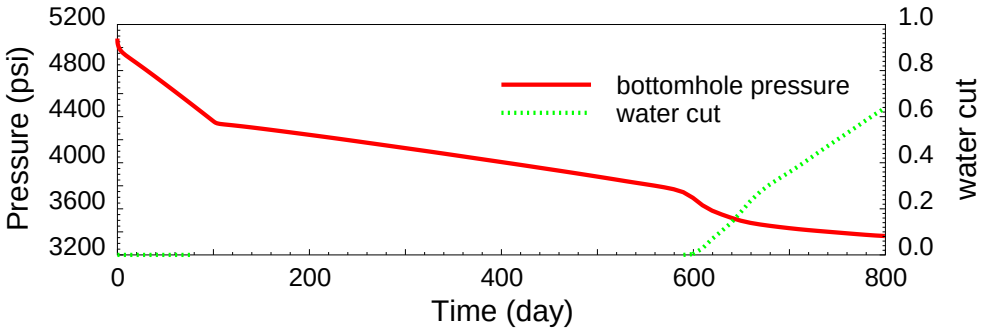

Figure 6.43: Bottomhole pressure and watercut for Well #8

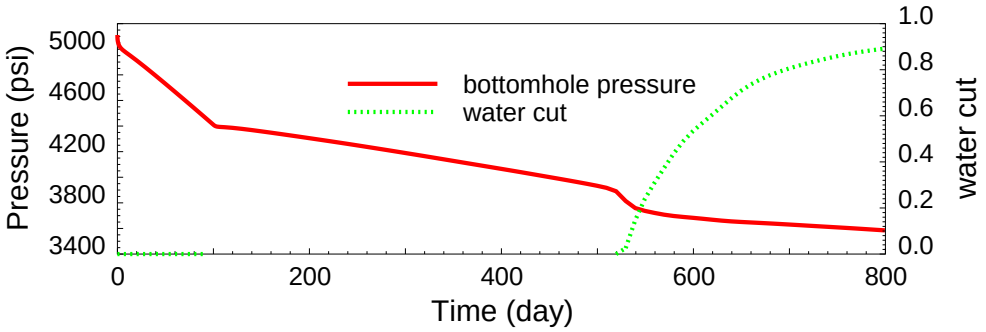

Figure 6.44: Bottomhole pressure and watercut for Well #9

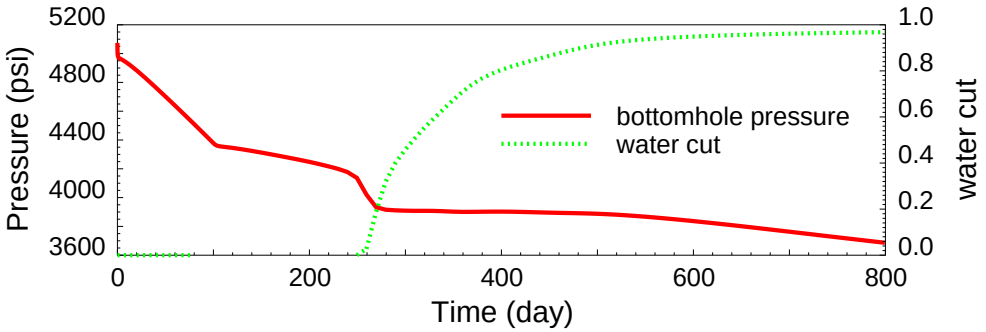

Figure 6.45: Bottomhole pressure and watercut for Well #10

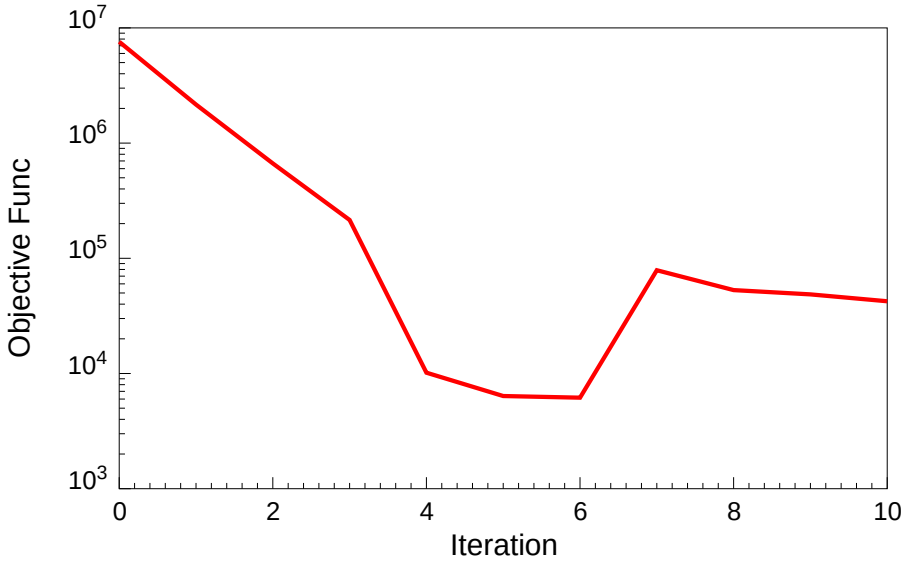

Figure 6.46: Convergence curve (wavelet modeling, Case 6.4)

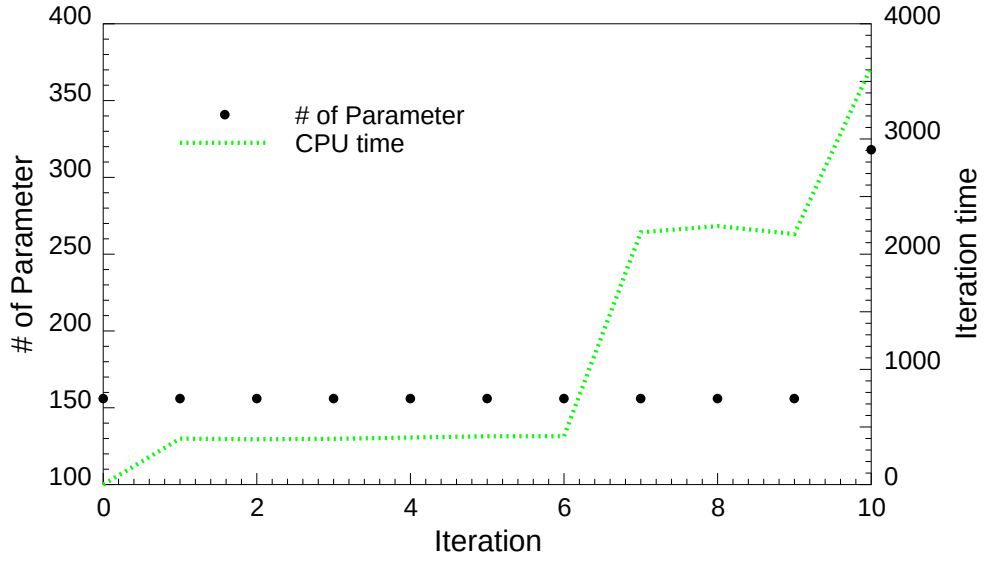

Figure 6.47: Number of parameters and CPU time at each iteration (wavelet modeling, Case 6.4)

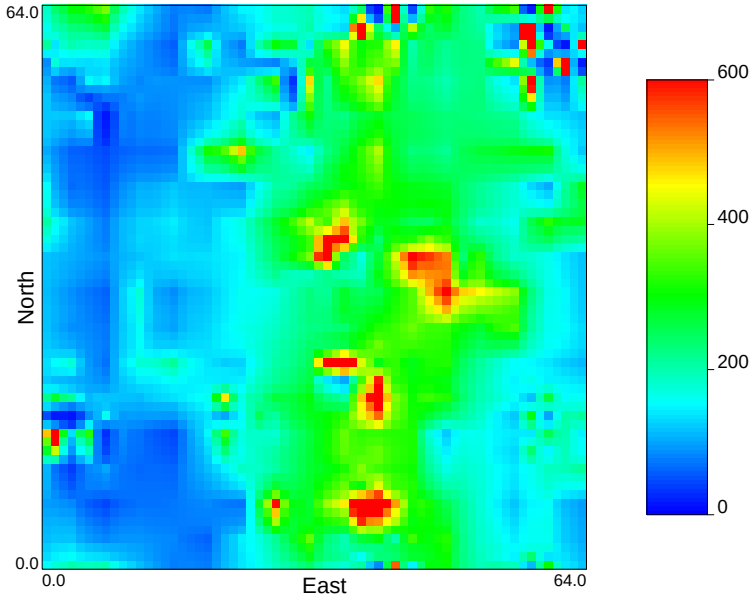

Figure 6.48: Permeability distribution with 318 parameters

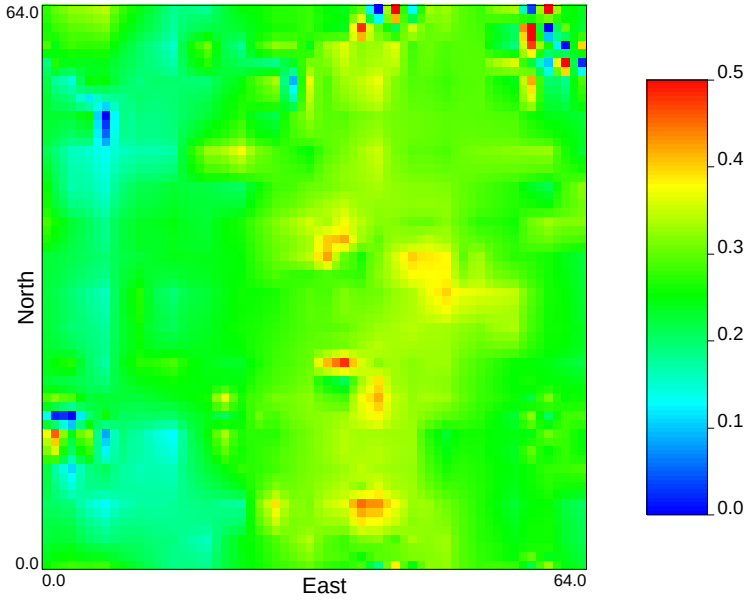

Figure 6.49: Porosity distribution with 318 parameters

Chapter 7

Conclusions and Open Issues

7.1 Conclusions

In this study, we found out that most of the properties of the wavelet analysis are very desirable in reservoir characterization and data integration. Not only can the wavelet analysis overcome most of the shortcomings of the conventional parameter estimation methods, but also can make the algorithm much more efficient and stable. Furthermore, we can balance the trade-off between resolution and uncertainty very easily. Although there are still several issues to investigate, the application of wavelet analysis is very promising.

- *Fast and efficient algorithms for the wavelet transform.* As mentioned in Section 2.4, the computational cost for the wavelet transform of N data is $O(N)$. Therefore there is almost no additional computational cost for the wavelet transform, in comparison to the numerical simulator.

- *Wavelets are powerful in reducing the number of parameters.* All the wavelet coefficients are distributed at different resolution levels, therefore the sensitivity of the reservoir responses are different on different levels. In the case of the sensitivity coefficients of the downhole pressure, except for the region nearby wellbores, the sensitivity is very small at high levels (as demonstrated in Chapter 4). In other words, it is impossible for us to have a high-resolution description

in most of the reservoir. Therefore it makes no practical sense to consider those insignificant wavelet coefficients. Using different configuration and thresholds, we can reduce the number of wavelets to 15-20% of the original set. This mechanism improves the performance in several ways: first of all, it reduces the number of unknowns, so the regression will converge much faster; secondly, in each iteration it takes far less time to compute sensitivity coefficients on the reduced set of wavelets.

- *The wavelet analysis is powerful and flexible in data integration.* Due to the complex nature of the multiple scales of heterogeneity in reservoirs, different types of information may be sensitive to different scales of heterogeneities, or localized at different resolution levels. For example, production data give low-level information, while core and log data give high-resolution information, but only in a very small region. 4-D seismic data is more likely to be localized on an intermediate level, but more powerful in spatial resolution than core and log data. Each type of information has its own advantage and limitation. The wavelet analysis gives us a powerful tool to integrate all this information at the appropriate scale. We also proposed a new algorithm to integrate 3-D seismic data into the nonlinear regression in the form of a "mask" on the wavelet coefficient map.

- *The wavelet analysis provides a method to incorporate geological information and preserve the geological structure of reservoirs.* The wavelet analysis has a wide range of application in image analysis, especially in pattern recognition. The pattern recognition is very useful in reservoir characterization, in which we try to capture the major geological structures, e.g., faults, channels, and bounding surfaces. All of the geological structures can be treated as localized discontinuities, or edges. The wavelet analysis will show large coefficients at certain levels, which represent object edges, thus providing a means of edge detection.

- *Optimal set of parameters.* An optimal set of parameters are crucial in parameter estimation problems. In this study, we developed several methods to select

the optimal parameter set. The wavelet decomposition provides a very convenient way of estimating the natural combination of parameters that not only describe the reservoir but also honor the resolution power of the constraining data.

- *The algorithm is very easy to parallelize.* All the numerical computations in this work were carried out on a SGI Origin 200 workstation with four 270MHz MIPS R12000 (IP27) CPUs. The program we developed has very good scalability and was parallelized to use all four processors in the machine. A parallelized program would be the most appropriate choice for practical problems, which would have thousands of cells, requiring large memory storage and computing power.

- *The new approach is very general.* First of all, though we used a finite-difference reservoir simulator as the forward mathematical model, the application is not restricted to this category. It is very easy to apply this approach to other types of simulators, such as streamline simulators and finite-element simulators. Secondly, because this approach is powerful to reduce the number of parameters to be estimated, the wavelet analysis will improve efficiency and performance even for non-gradient methods, such as neural-network methods, simulated annealing methods, genetic algorithms, etc.

7.2 Future Work

Although the parameter estimation algorithm and the wavelet analysis are very powerful tools and enable us to explore the problem of reservoir characterization and data integration in depth, some aspects of this subject still need further investigation. Some open issues are:

- *Different wavelet bases.* In this work, we only demonstrated the use of the Daubechies wavelets, which are the simplest orthogonal wavelet bases. The

Daubechies wavelets have some properties unsuitable for our purpose. For instance, the Daubechies wavelets are very asymmetric and have compact supports. In this sense, some other wavelet bases should be taken into consideration. In order to avoid the boundary effects introduced by periodic wavelets, boundary wavelets should be implemented.

- *Adaptive wavelets.* Due to the complex nature of the heterogeneity in reservoirs, the statistical distributions are not strictly stationary in most cases. Therefore it is desirable to choose wavelet bases adaptively for different probability distributions. For example, in regions with strong discontinuities, a wavelet with shorter support is a better choice to avoid large wavelet coefficients. This approach will depend strongly on geological information to identify geological structures.

- *Pattern recognition techniques.* The pattern recognition technique is very useful in reservoir characterization. Though this study has applied some simple techniques, this aspect of the problem requires further investigation.

Nomenclature

A	Cross-section area, ft^2
B	Formation factor
C_D	Data covariance (total)
C_d	Data measurement covariance
C_M	Model parameter covariance
C_T	Theoretical model covariance
c_t	Total isothermal compressibility, psi^{-1}
D	Data space
$\vec{d}$	Observed parameters (data)
E	Objective function
$\vec{F}$	Gradient of objective function
f	Probability density function
G	Sensitivity coefficient matrix
g	Gravity, ft/s^2
$g[n]$	Discrete wavelet filter
H	Hessian matrix
h	Height, ft
$h[n]$	Discrete wavelet filter
J	Jacobian matrix
k	Permeability, md
k_r	Relative permeability
L	Length, ft
M	Model space

$\mathcal{N}$	Gaussian distribution
p	Pressure, psi
p_c	Capillary pressure, psi
q	Flow rate, STB/day
Q	Total flow rate, STB/day
R	Resolution operator
r_0	Equivalent wellbore radius, ft
r_w	Wellbore radius, ft
S	Sensitivity coefficient matrix
S_o	Oil saturation
S_w	Water saturation
T_p	Transmissibility of Phase p
t	Time, day
$\vec{u}$	Reservoir response
V	Volume, ft^3
$\vec{v}$	Known reservoir properties
W	Wavelet transformation matrix
WI	Well index
wc	Watercut
z	Height, ft

GREEKS

$\vec{\alpha}$	Model parameters
$\delta(x, x')$	Dirac delta function
ϕ	Porosity
$\phi(t)$	Scaling function
Ψ_p	Phase potential
$\psi(t)$	Wavelet function
μ	Fluid viscosity, cp

ϵ	Small error
ω	Frequency
ρ	Density, lbm/ft^3

SUBSCRIPTS

0	Initial
D	Data
i, j, l	Integer indices
M	Model
ML	Maximum likelihood
n	Iteration number
o	Oil phase
p	phase
r	Reference
T	Theoretical
w	Water phase
$well$	Wellbore

SUPERSCRIPTS

cal	Calculated
est	Estimated
GN	Gauss-Newton method
(k)	Timestep number
obs	Observed
pri	Prior
T	Transpose
$*$	Conjugate
$\bar{}$	Laplace transformed variable

$\hat{}$ Fourier transformed variable

Appendix A

Wavelet Analysis

A.1 Wavelet Bases

The wavelet transform $Wf(u,\, s)$ is a two-dimensional representation of a one-dimensional signal f. This indicates the existence of some redundancy that can be reduced and even removed by subsampling the parameters of the transform. Subsampling the transform defines a complete signal representation if any signal can be reconstructed from linear combinations of discrete families of wavelet atoms $\{\psi_{u_n,s_j}\}_{(j,n)\in\mathcal{Z}^2}$. Completely eliminating the redundancy is equivalent to building a basis of the signal space. One can construct wavelets ψ such that the dilated and translated family

$$\{\psi_{j,n}(t) = \frac{1}{\sqrt{2^j}}\psi(\frac{t-2^j n}{2^j})\}_{(j,n)\in\mathcal{Z}^2} \tag{A.1}$$

is an orthonormal basis of $\mathbf{L}^2(\mathcal{R})$.

A.1.1 Multiresolution Approximations

Orthogonal wavelets dilated by 2^j carry signal variations at the resolution 2^{-j}. The construction of these bases can thus be related to *multiresolution approximations*. A multiresolution approximation of a function f at a resolution 2^{-j} is specified by a discrete grid of samples that provides local averages of f over neighborhoods of size proportional to 2^j. Formally, the approximation of a function at a resolution 2^{-j}

is defined as an orthogonal projection on a space $\mathbf{V}_j \subset \mathbf{L}^2(\mathcal{R})$, which regroups all possible approximations at the resolution 2^{-j}. The orthogonal projection of f is the function $f_j \in \mathbf{V}_j$ that minimizes $\|f - f_j\|$. The following definition introduced by Mallat (1987) and Meyer (1992) specifies the mathematical properties of multiresolution spaces. A sequence $\{\mathbf{V}_j\}_{j\in\mathcal{Z}}$ of closed subspaces of $\mathbf{L}^2(\mathcal{R})$ is *a multiresolution approximation* if the following six properties are satisfied:

$$\forall (j,k) \in \mathcal{Z}^2 \ , \ f(t) \in \mathbf{V}_j \ \Leftrightarrow \ f(t - 2^j k) \in \mathbf{V}_j, \tag{A.2}$$

$$\forall j \in \mathcal{Z} \ , \ \mathbf{V}_{j+1} \subset \mathbf{V}_j, \tag{A.3}$$

$$\forall j \in \mathcal{Z} \ , \ f(t) \in \mathbf{V}_j \ \Leftrightarrow \ f(\frac{t}{2}) \in \mathbf{V}_{j+1}, \tag{A.4}$$

$$\lim_{j\to+\infty} \mathbf{V}_j = \bigcap_{j=-\infty}^{+\infty} \mathbf{V}_j = \{0\}, \tag{A.5}$$

$$\lim_{j\to-\infty} \mathbf{V}_j = Closure(\bigcup_{j=-\infty}^{+\infty} \mathbf{V}_j) = \mathbf{L}^2(\mathcal{R}). \tag{A.6}$$

There exists θ such that $\{\theta(t-n)\}_{n\in\mathcal{Z}}$ is a Riesz basis of $\mathbf{V}_0$.

The approximation of f at the resolution 2^{-j} is defined as the orthogonal projection $P_{\mathbf{V}_j} f$ on $\mathbf{V}_j$. To compute this projection, we must first find an orthonormal basis of $\mathbf{V}_j$. The following theorem orthogonalizes the Riesz basis $\{\theta(t-n)\}_{n\in\mathcal{Z}}$ and constructs an orthogonal basis of each space $\mathbf{V}_j$ by dilating and translating a single function ϕ called a *scaling function*.

Let $\{\mathbf{V}_j\}_{j\in\mathcal{Z}}$ be a multiresolution approximation and ϕ be the scaling function whose Fourier transform is

$$\hat{\phi}(\omega) = \frac{\hat{\theta}(\omega)}{(\sum_{k=-\infty}^{+\infty} |\hat{\theta}(\omega + 2k\pi)|^2)^{\frac{1}{2}}}. \tag{A.7}$$

Let us denote

$$\phi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \phi\left(\frac{t-n}{2^j}\right). \tag{A.8}$$

The family $\{\phi_{j,n}\}_{n\in\mathcal{Z}}$ is an orthonormal basis of $\mathbf{V}_j$ for all $j \in \mathcal{Z}$.

The orthogonal projection of f over $\mathbf{V}_j$ is obtained with an expansion in the scaling orthogonal basis

$$P_{\mathbf{V}_j} f = \sum_{n=-\infty}^{+\infty} \langle f,\, \phi_{j,n} \rangle \phi_{j,n} = \sum_{n=-\infty}^{+\infty} a_j[n] \phi_{j,n}, \tag{A.9}$$

where the inner products

$$a_j[n] = \langle f,\, \phi_{j,n} \rangle \tag{A.10}$$

provide a discrete approximation at the scale 2^j. A multiresolution approximation is entirely characterized by the scaling function ϕ that generates an orthogonal basis of each space $\mathbf{V}_j$. Any scaling function is specified by a discrete filter called a *conjugate mirror filter*.

A.1.2 Conjugate Mirror Filters

The multiresolution causality property Eq. A.3 imposes that $\mathbf{V}_j \subset \mathbf{V}_{j-1}$. In particular $\frac{1}{\sqrt{2}} \phi\left(\frac{t}{2}\right) \in \mathbf{V}_1 \subset \mathbf{V}_0$. Because $\{\phi(t-n)\}_{n \in \mathcal{Z}}$ is an orthonormal basis of $\mathbf{V}_0$, we can decompose

$$\frac{1}{\sqrt{2}} \phi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} h[n] \phi(t-n), \tag{A.11}$$

with

$$h[n] = \langle \frac{1}{\sqrt{2}} \phi\left(\frac{t}{2}\right), \phi(t-n) \rangle. \tag{A.12}$$

This sequence $h[n]$ will be interpreted as a discrete filter. The Fourier transform of both sides of Eq. A.11 yields

$$\hat{\phi}(2\omega) = \frac{1}{\sqrt{2}} \hat{h}(\omega) \hat{\phi}(\omega) \tag{A.13}$$

for

$$\hat{h}(\omega) = \sum_{n=-\infty}^{+\infty} h[n] e^{-in\omega}. \tag{A.14}$$

As Mallat (1987) and Meyer (1992) proved, if $\phi \in \mathbf{L}^2(\mathcal{R})$ is an integrable scaling function, the Fourier series of $h[n] = \langle \frac{1}{\sqrt{2}} \phi\left(\frac{t}{2}\right), \phi(t-n) \rangle$ satisfies

$$\forall \omega \in \mathcal{R}\ ,\ |\hat{h}(\omega)|^2 + |\hat{h}(\omega+\pi)|^2 = 2, \tag{A.15}$$

and

$$\hat{h}(0) = \sqrt{2}. \tag{A.16}$$

Conversely, if $\hat{h}(\omega)$ is 2π periodic and continuously differentiable in a neighborhood of $\omega = 0$, if it satisfies Eq. A.15 and (A.16) and if

$$\inf_{\omega \in [-\frac{\pi}{2}, \frac{\pi}{2}]} |\hat{h}(\omega)| > 0 \tag{A.17}$$

then

$$\hat{\phi}(\omega) = \prod_{p=1}^{+\infty} \frac{\hat{h}\left(2^{-p}\omega\right)}{\sqrt{2}} \tag{A.18}$$

is the Fourier transform of a scaling function $\phi \in \mathbf{L}^2(\mathcal{R})$. Discrete filters whose transfer functions satisfy Eq. A.15 are called *conjugate mirror filters*.

A.1.3 Orthogonal Wavelets

The approximation of f at the scales 2^j and 2^{j-1} are respectively equal to their orthogonal projections on $\mathbf{V}_j$ and $\mathbf{V}_{j-1}$. Let $\mathbf{W}_j$ be the orthogonal complement of $\mathbf{V}_j$ and $\mathbf{V}_{j-1}$:

$$\mathbf{V}_{j-1} = \mathbf{V}_j \bigoplus \mathbf{W}_j. \tag{A.19}$$

The orthogonal projection of f on $\mathbf{V}_{j-1}$ can be decomposed as the sum of orthogonal projections on $\mathbf{V}_j$ and $\mathbf{W}_j$:

$$P_{\mathbf{V}_{j-1}} f = P_{\mathbf{V}_j} f + P_{\mathbf{W}_j} f. \tag{A.20}$$

The complement $P_{\mathbf{W}_j} f$ provides the details of f that appear at the scale 2^{j-1} but which disappear at the coarser scale 2^j. Mallat (1987) and Meyer (1992) proved that one can construct an orthonormal basis of $\mathbf{W}_j$ by scaling and translating a wavelet ψ. Let ϕ be a scaling function and h the corresponding conjugate mirror filter. Let ψ be the function whose Fourier transform is

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}} \hat{g}\left(\frac{\omega}{2}\right) \hat{\phi}\left(\frac{\omega}{2}\right), \tag{A.21}$$

with

$$\hat{g}(\omega) = e^{-i\omega} \hat{h}^*(\omega + \pi). \tag{A.22}$$

Let us denote

$$\psi_{j,n}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t-2^j n}{2^j}\right). \tag{A.23}$$

For any scale 2^j, $\{\psi_{j,n}\}_{n\in\mathcal{Z}}$ is an orthonormal basis of $\mathbf{W}_j$. For all scales, $\{\psi_{j,n}\}_{n\in\mathcal{Z}}$ is an orthonormal basis of $\mathbf{L}^2(\mathcal{R})$. It can also be shown that $\hat{g}$ is the Fourier series of

$$g[n] = \langle \frac{1}{\sqrt{2}}\psi\left(\frac{t}{2}\right), \psi(t-n)\rangle, \tag{A.24}$$

which are the decomposition coefficients of

$$\frac{1}{\sqrt{2}}\psi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} g[n]\phi(t-n). \tag{A.25}$$

Calculating the inverse Fourier transform of Eq. A.22 yields

$$g[n] = (-1)^{1-n}h[1-n]. \tag{A.26}$$

This mirror filter plays an important role in the fast wavelet transform algorithm.

A.1.4 Daubechies Wavelets

In this work, we used the Daubechies wavelet family for the wavelet analysis. Daubechies wavelets have several properties that are suitable for our purpose. First of all, Daubechies wavelets have a compact support of minimum size for any given number n_v of vanishing moments. Furthermore, Daubechies wavelets are also orthogonal transforms and it is much easier to analyze the variance and resolution of the parameter estimation procedures.

Usually, we denote the Daubechies wavelet with a vanishing moment of p as D_p, and the discrete filter $h[n]$ has $2p$ non-zero coefficients. In this work, we used only Daubechies wavelets with vanishing moments of 1, 2, and 3.

A.1.5 Separable Wavelet Bases

A separable wavelet orthonormal basis of $\mathbf{L}^2(\mathcal{R}^2)$ is constructed with separable products of a scaling function ϕ and a wavelet ψ. We define three wavelets:

$$\psi^1(x,y) = \phi(x)\psi(y)\ ,\ \psi^2(x,y) = \psi(x)\phi(y)\ ,\ \psi^2(x,y) = \psi(x)\psi(y), \tag{A.27}$$

and for $1 \leq k \leq 3$

$$\psi^k_{j,n,m}(x,y) = \frac{1}{2^j}\psi^k\left(\frac{x-2^j n}{2^j}, \frac{y-2^j m}{2^j}\right). \tag{A.28}$$

The wavelet family

$$\{\psi^1_{j,n,m}(x,y), \psi^2_{j,n,m}(x,y), \psi^3_{j,n,m}(x,y)\}_{(n,m)\in\mathcal{Z}^2} \tag{A.29}$$

is an orthonormal basis of $\mathbf{W}^2_j$ and

$$\{\psi^1_{j,n,m}(x,y), \psi^2_{j,n,m}(x,y), \psi^3_{j,n,m}(x,y)\}_{(j,n,m)\in\mathcal{Z}^2} \tag{A.30}$$

is an orthonormal basis of $\mathbf{L}^2(\mathcal{R}^2)$.

A.2 Fast Orthogonal Wavelet Transform

A.2.1 One-Dimensional Discrete Transform

Decomposition coefficients in a wavelet orthogonal basis are computed with a fast algorithm that cascades discrete convolutions with h and g, and subsamples the output. Given a discrete signal $a_0[n]$, let us define

$$f(t) = \sum_{n=-\infty}^{+\infty} a_0[n]\phi(t-n) \in \mathbf{V}_0. \tag{A.31}$$

Because $\{\phi(t-n)\}_{n\in\mathcal{Z}}$ is orthonormal,

$$a_0[n] = \langle f(t), \phi(t-n)\rangle = f * \bar{\phi}(n). \tag{A.32}$$

The discrete wavelet coefficients of a_0 are defined to be the wavelet coefficients of f:

$$d_j[n] = \langle f, \psi_{j,n}\rangle. \tag{A.33}$$

$P_{\mathbf{V}_j}$, the projection of f on the space $\mathbf{V}_j$, is characterized by

$$a_j[n] = \langle f, \phi_{j,n}\rangle. \tag{A.34}$$

Mallat(1989, 1998) showed that the wavelet decomposition and reconstruction is calculated with discrete convolutions. We denote

$$\bar{x}[n] = x[-n] \tag{A.35}$$

and

$$\check{x}[n] = \begin{cases} x[p] & if\ n = 2p \\ 0 & if\ n = 2p+1 \end{cases} . \tag{A.36}$$

In the decomposition,

$$a_{j+1}[p] = \sum_{n=-\infty}^{+\infty} h[n-2p]a_j[n] = a_j * \bar{h}[2p], \tag{A.37}$$

$$d_{j+1}[p] = \sum_{n=-\infty}^{+\infty} g[n-2p]a_j[n] = a_j * \bar{g}[2p]. \tag{A.38}$$

In the reconstruction,

$$\begin{aligned} a_j[p] &= \sum_{n=-\infty}^{+\infty} h[p-2n]a_{j+1}[n] + \sum_{n=-\infty}^{+\infty} g[p-2n]d_{j+1}[n] \\ &= \check{a}_{j+1} * h[p] + \check{d}_{j+1} * g[p]. \end{aligned} \tag{A.39}$$

A.2.2 Wavelet Bases on an Interval

To decompose signals f defined over an interval $[0, N]$, it is necessary to construct the wavelet basis of $\mathbf{L}^2[0, N]$. Such bases are synthesized by modifying a wavelet basis $\{\psi_{j,n}\}_{(j,n)\in\mathcal{Z}^2}$ of $\mathbf{L}^2(\mathcal{R})$. If $\psi_{j,n}$ has a support included in $[0, N]$, it is not modified. The "border" wavelets $\psi_{j,n}$ whose supports overlap $t = 0$ or $t = N$ are transformed into functions with support in $[0, N]$, providing the necessary complement to generate a basis of $\mathbf{L}^2[0, N]$. There are several different ways to construct such wavelet bases.

- *Periodic wavelets*: A wavelet basis $\{\psi_{j,n}\}_{(j,n)\in\mathcal{Z}^2}$ is transformed into a wavelet basis of $\mathbf{L}^2[0, N]$ by periodizing each $\psi_{j,n}$. The periodization of f over $[0, N]$ is

$$f^{per}(t) = \sum_{k=-\infty}^{+\infty} f(t + kN). \tag{A.40}$$

The periodic wavelets are

$$\psi_{j,n}^{per}(t) = \frac{1}{\sqrt{2^j}} \sum_{k=-\infty}^{+\infty} \psi\left(\frac{t - 2^j n + kN}{2^j}\right). \tag{A.41}$$

Thus periodizing a wavelet orthogonal basis of $\mathbf{L}^2(\mathcal{R})$ defines a wavelet orthogonal basis of $\mathbf{L}^2[0.N]$. However, this basis construction has the disadvantage of creating high amplitude wavelet coefficients in the neighborhood of $t = 0$ and $t = N$ because the border wavelets have separate components with no vanishing moments.

- *Folded wavelets*: To avoid creating discontinuities with simple periodization, the signal is folded with respect to $t = 0$: $f_0(t) = f(t) + f(-t)$. The support of f_0 is $[-N, N]$ and it is transformed into a $2N$ periodic signal

$$f^{fold}(t) = \sum_{k=-\infty}^{+\infty} f_0(t - 2kN) = \sum_{k=-\infty}^{+\infty} f(t - 2kN) + \sum_{k=-\infty}^{+\infty} f(2kN - t). \tag{A.42}$$

Decomposing f^{fold} in a wavelet basis $\{\psi_{j,n}\}_{(j,n)\in\mathcal{Z}^2}$ is equivalent to decomposing f on a folded wavelet basis $\{\psi_{j,n}^{fold}\}_{(j,n)\in\mathcal{Z}^2}$.

- *Boundary Wavelets*: The restriction of periodic and folded "border" wavelets in the neighborhood of $t = 0$ and $t = N$ have respectively 0 and 1 vanishing moment. To avoid creating large amplitude wavelet coefficients at the borders, one can synthesize border wavelets that have as many vanishing moments as the original wavelet ψ.

Each of the cascade algorithms decomposes a_0 into a discrete wavelet transform $a_J, \{d_j\}_{1\leq j\leq J}$ with $O(N)$ operations. The reconstruction algorithm is also with $O(N)$ operations.

A.3 Banach and Hilbert Spaces

A *Banach space* $\mathbf{H}$ is a vector space that admits a norm. A norm satisfies the following properties:

$$\forall x \in \mathbf{H},\ \|x\| \geq 0 \ \textit{and}\ \|x\| = 0 \ \Leftrightarrow \ x = 0, \tag{A.43}$$

$$\forall \lambda \in \mathcal{C} \quad \|\lambda x\| = |\lambda| \cdot \|x\|, \tag{A.44}$$

$$\forall x, y \in \mathbf{H}, \quad \|x + y\| \leq \|x\| + \|y\|. \tag{A.45}$$

With such a norm, the convergence of $\{x_n\}_{n\in\mathcal{N}}$ to x in $\mathbf{H}$ means that

$$\lim_{n\to+\infty} x_n = x \quad \Leftrightarrow \quad \lim_{n\to+\infty} \|x_n - x\| = 0. \tag{A.46}$$

To guarantee that we remain in $\mathbf{H}$ when taking such limits, we impose a completeness property, using the notion of *Cauchy sequences*. A sequence $\{x_n\}_{n\in\mathcal{N}}$ is a Cauchy sequence if for any $\epsilon > 0$, if n and p are large enough, then $\|x_n - x_p\| < \epsilon$. The space $\mathbf{H}$ is said to be *complete* if every Cauchy sequence in $\mathbf{H}$ converges to an element of $\mathbf{H}$.

A *Hilbert space* $\mathbf{H}$ is a Banach space with an *inner product*. The inner product of two vectors $\langle x, y\rangle$ is linear with respect to its first argument:

$$\forall \lambda_1, \lambda_2 \in \mathcal{C}, \quad \langle \lambda_1 x_1 + \lambda_2 x_2, y\rangle = \lambda_1 \langle x_1, y\rangle + \lambda_2 \langle x_2, y\rangle. \tag{A.47}$$

The inner product has an Hermitian symmetry:

$$\langle x, y\rangle = \langle y, x\rangle^*. \tag{A.48}$$

Moreover

$$\langle x, x\rangle \geq 0 \ \textit{ and } \ \langle x, x\rangle = 0 \quad \Leftrightarrow \quad x = 0. \tag{A.49}$$

It can be easily verified that $\|x\| = \langle x, x\rangle^{1/2}$ is a norm.

Appendix B

Parallelization

B.1 Parallel Computing

Parallel computations have become a mainstay of scientific and engineering computing, including physics, chemistry, material science, biology, earth sciences, and others. A direct correspondence can be drawn between levels of computational performance and the problems that can be studied and solved through simulation. A report from Federal Office of Science and Technology Policy (1993) summarizes the grand challenge applications: a collection of important scientific and engineering problems, including 72-hour weather forecast, oil reservoir modeling, fluid turbulence, chemical dynamics, structural biology, etc. These applications have very challenging performance and storage requirements, which can not usually be fulfilled by a single-processor computer system.

In this study, we proposed a new approach to improve the efficiency and performance of reservoir parameter estimation. Nevertheless, the computational load is still very intensive. Parallel computing proved to be a very effective approach to further improve the performance. The parameter estimation algorithms used in this study have a very high level of parallelism.

In our approach, the computational load per iteration can be divided into two categories: the simulation run and the calculation of sensitivity coefficients, as follows:

$$T_{iter} = T_{sim} + n_m \cdot T_{sen}, \tag{B.1}$$

where n_m is the number of model parameters. Eq. B.1 is a simplified formula with the assumption that the computational load for sensitivity calculations is proportional to the number of model parameters. Eq. B.1 can be generalized for multiprocessor cases:

$$T_{iter}(p) = \frac{T_{sim}}{\gamma_{sim}(p)} + \frac{n_m}{\gamma_{sen}(p)} \cdot T_{sen}, \tag{B.2}$$

where p is the number of processors, $\gamma_{sim}(p)$ and $\gamma_{sen}(p)$ are speedup factors for the calculations of simulation and sensitivities respectively. The speedup factor on p processors is defined as (Culler, Singh & Gupta 1999)

$$\gamma(p) = \frac{\text{Performance}(p)}{\text{Performance}(1)}. \tag{B.3}$$

For a single fixed problem, the performance of the system is simply the reciprocal of the time to complete the problem, so we have the following important special case:

$$\gamma_{fixed}(p) = \frac{\text{Time}(1)}{\text{Time}(p)}. \tag{B.4}$$

Therefore, the speedup of the overall computation in our approach is

$$\gamma(p) = \frac{T_{iter}}{T_{iter}(p)} = \frac{T_{sim} + n_m \cdot T_{sen}}{T_{sim}/\gamma_{sim}(p) + n_m/\gamma_{sen}(p) \cdot T_{sen}}, \tag{B.5}$$

Another key indicator for the performance is the parallel efficiency, defined as

$$\eta(p) = \gamma(p)/p = \frac{T_{sim} + n_m \cdot T_{sen}}{T_{sim}/\eta_{sim}(p) + n_m/\eta_{sen}(p) \cdot T_{sen}}. \tag{B.6}$$

B.2 Implementation for Simulation

The parallel efficiency of the simulation is usually much less than one when the number of processors is large. The implementation of an efficient parallelization requires very efficient parallel solvers, good partitions, and hardware support for parallel architectures in some cases. Most of these subjects are out of the scope of this study. In order to simplify the implementation and focus on the major objectives, we used the parallel computing software package *PETSc* (Balay, Gropp, McInnes & Smith 1999) in this study. For uniprocessor cases, we used the preconditioned Krylov

subspace solver GMRES (Saad & Schultz 1986), with the $ILU(n)$ preconditioner. For multiprocessor cases, *PETSc* supports the block Jacobi and overlapping additive Schwarz methods in parallel. However, only the uniprocessor version of the block Gauss-Seidel method is currently in place. By default, the PETSc implementations of these methods employ ILU(0) factorization on each individual block.

We investigated the impact of the problem size and the number of processors on parallel efficiencies using two different reservoir examples. Case 1 is a reservoir with 32×32 gridblocks as shown in Section 5.2, and Case 2 is a reservoir with 64×64 gridblocks as shown in Section 6.4. In both the cases, the simulation will run for a simulated time of 500 days. The computations were carried out on a SGI Origin 200 workstation, with four 270 MHZ MIPS processors. Fig. B.1 shows the computational time for the two cases, with different numbers of processors. Fig. B.2 shows the parallel efficiency for different configurations. With two processors, Case 2 yields a very good parallel efficiency of 0.813, while Case 1 yields a parallel efficiency of 0.631. With four processors, both cases yield very poor parallel efficiencies below 0.5. We can make several observations from the numerical experiments:

- First of all, with the same number of processors, larger problems have better parallel efficiency, because the boundary/volume ratio is smaller for larger problems. The relative communication costs are roughly proportional to the boundary/volume ratio.

- Secondly, for a fixed-size problem, as the number of processors increases, the parallel efficiency decreases. This is mainly due to communication overhead and partition costs.

- There is one more factor in the parallel performance: the communication bandwidth. Although *PETSc* uses the message passing interface (MPI) to implement the parallel solvers, the SGI Origin 200 employs the shared address space architecture with a shared bus. This architecture implies that the communication bandwidth per processor is only $1/p$ of the overall bandwidth, which further deteriorates the performance.

B.3 Implementation for Sensitivity Calculation

If we use the same parallelization scheme for the sensitivity calculation, the overall parallel efficiency would be the same as the parallel efficiency of the simulation, as given by Eq. B.5 and Eq. B.6. Section B.2 shows that it is very difficult to improve the parallel efficiency of the simulation, with general parallel solvers. In order to overcome this disadvantage, we implemented a different algorithm to calculate sensitivity coefficients in parallel. Because we use the modified GPST method to calculate sensitivities, we need to solve the linear system with the same left-hand side many times. This system has a very high level of parallelism. The parallel implementation per simulation timestep is outlined as follows:

1. *Parallel solvers for the Newton-Raphson algorithm*: The nonlinear mass balance equation is solved iteratively using the Newton-Raphson algorithm, while the Jacobian matrix is updated accordingly, as shown in Section 3.1.2. The iterative linear system shown in Eqs. 3.17 through 3.20 is restated here:

$$J_{\nu}^{(k+1)} \cdot \delta\vec{u}_{\nu}^{(k+1)} = -\vec{f}_{\nu}^{(k+1)}, \tag{B.7}$$

   with

$$u_{\nu+1}^{(k+1)} = u_{\nu}^{(k+1)} + \delta u_{\nu}^{(k+1)}. \tag{B.8}$$

   At this step, Eq. B.7 is solved using *PETSc* parallel solvers, as described in Section B.2.

2. *Sequential Solvers for the Modified GPST algorithm*: As shown in Section 3.2.2, the modified GPST algorithm is expressed as:

$$J^{(k+1)}S^{(k+1)} = -D^{(k+1)}S^{(k)} - Y^{(k+1)}, \tag{B.9}$$

   where $J^{(k+1)}$ is the last Jacobian matrix computed in the Newton-Raphson method in the numerical simulator. Suppose we need to calculate sensitivity coefficients for n_m parameters, then Eq. B.9 need to be solved n_m times, with the same Jacobian matrix $J^{(k+1)}$ and n_m different right-hand-side vectors. The Jacobian matrix can be distributed to all the processors using the Message

Passing Interface (Message Passing Interface Forum 1997). Therefore, each processor will keep a local copy of the Jacobian matrix. The n_m parameters can also be distributed to all the processors, with each processor keeping n_m/p parameters. On each processor, Eq. B.9 will be solved using a sequential solver (in this case, we used the GMRES solver with the $ILU(1)$ preconditioner). The computational time would be $n_m/p \cdot T_{sen}$, assuming the processors have good load balance and light communication overhead. Therefore, the speedup factor for sensitivity calculation is

$$\gamma_{sen}(p) = \frac{\text{Time}(1)}{\text{Time}(p)} \approx \frac{n_m \cdot T_{sen}}{n_m/p \cdot T_{sen}} = p, \tag{B.10}$$

and the parallel efficiency $\eta_{sen}(p)$ is very close to one.

The overall speedup can be rewritten as

$$\gamma(p) \approx \frac{T_{sim} + n_m \cdot T_{sen}}{T_{sim}/\gamma_{sim}(p) + n_m/p \cdot T_{sen}}, \tag{B.11}$$

and the parallel efficiency becomes

$$\eta(p) = \gamma(p)/p \approx \frac{T_{sim} + n_m \cdot T_{sen}}{T_{sim} \cdot /\eta_{sim}(p) + n_m \cdot T_{sen}}. \tag{B.12}$$

For fixed T_{sim} and T_{sen}, Eq. B.12 implies that the overall parallel efficiency depends on two major factors: the parallel efficiency for the simulation η_{sim} and the number of parameters n_m. If the simulation parallel efficiency is very good, then the overall parallel efficiency would be very close to one, regardless of n_m. With a given parallel solver and a fixed problem, the parallel efficiency for the solver is fixed and will not have further impact on the performance. However, in our study, we usually have $n_m >> 1$, therefore the overall parallel efficiency is very good, even using a parallel solver with a low parallel efficiency. In real life, there is always some overhead associated with communications and partitions. Therefore, the real parallel efficiency is usually less than that given by Eq. B.12.

We used the two cases shown in Section B.2 to study the parallel efficiency of our approach. For each case, we ran the simulation with sensitivity calculations of 100 model parameters. Fig. B.3 shows the computational times for both the cases, while

Fig. B.4 shows the parallel efficiencies. For both the test cases, the parallel efficiency is fairly good with different configurations.

In summary, the parallelization approach used in this study has very good parallel efficiency. However, there are some aspects that could be further improved. First of all, a better parallel solver could further improve the parallel efficiency of the simulation. Secondly, an optimized implementation could reduce the communication costs. Furthermore, a different hardware architecture is required to implement high-performance parallel algorithms.

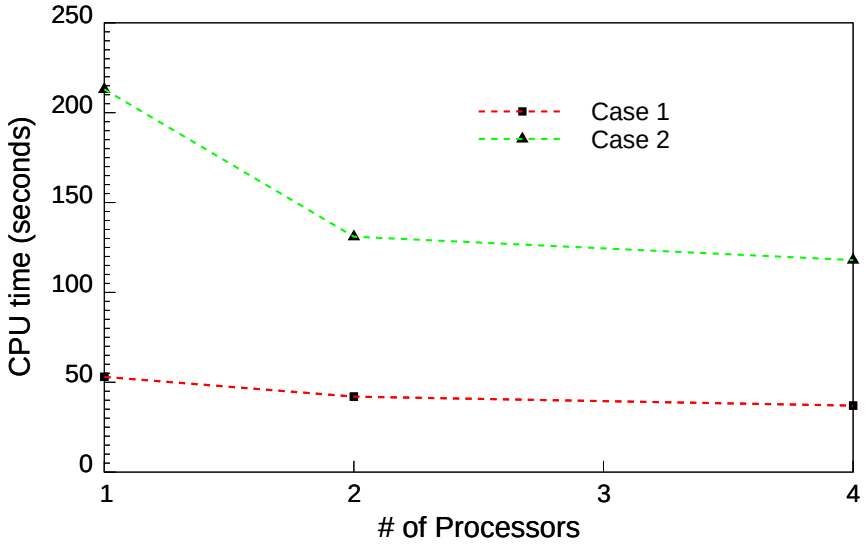

Figure B.1: CPU times of one single simulation run, for reservoirs with 1024 and 4096 blocks respectively

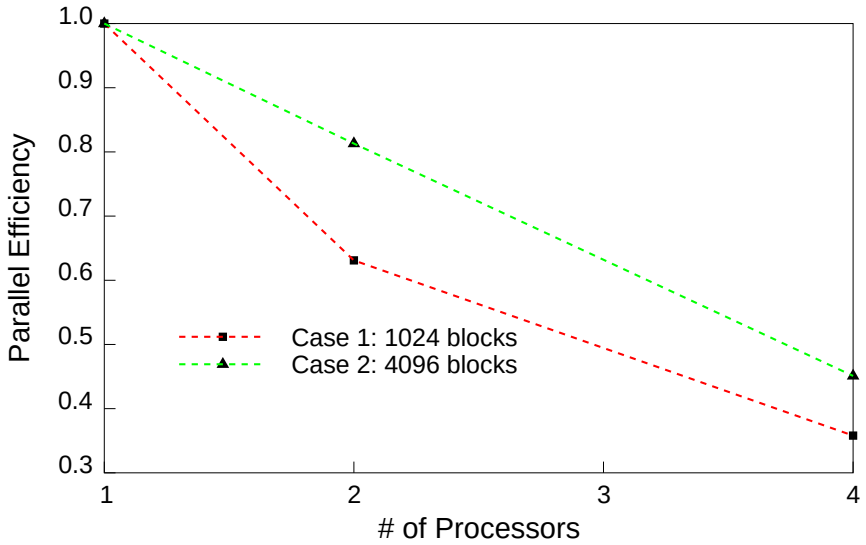

Figure B.2: Parallel efficiency curves of one single simulation run, for reservoirs with 1024 and 4096 blocks respectively

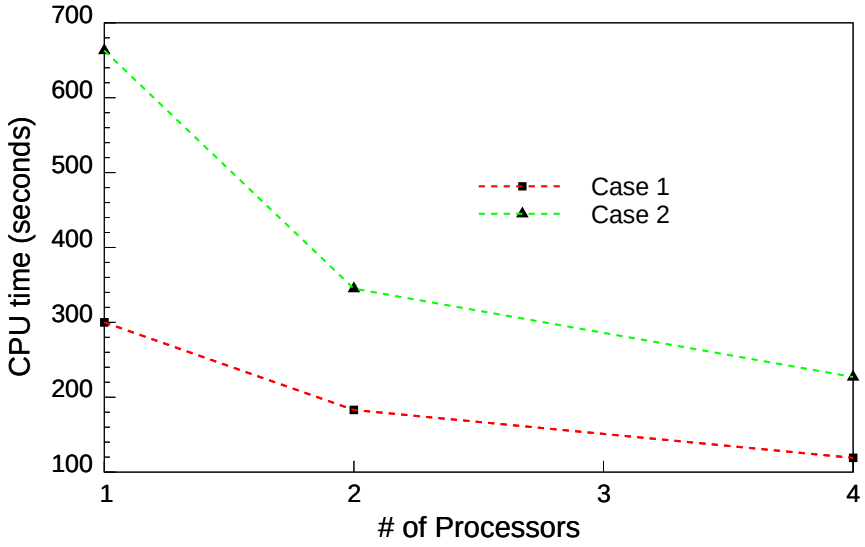

Figure B.3: CPU times of one simulation run with sensitivity calculation of 100 model parameters, for reservoirs with 1024 and 4096 blocks respectively

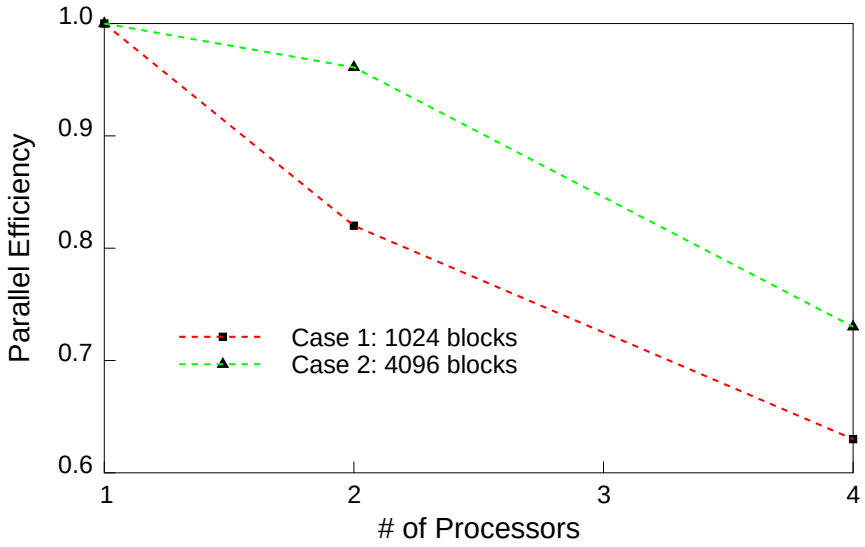

Figure B.4: Parallel efficiency curves, for simulation with sensitivity calculation of 100 model parameters

Bibliography

Anterion, F., Eymard, R., and Karcher, B., "Use of Parameter Gradients for Reservoir History Matching", paper SPE 18433 presented at the 1989 SPE Symposium on Reservoir Simulation, Houston, February 6-8, 1989.

Athichanagorn, S., "Development of an Interpretation Methodology for Long-Term Pressure Data from Permanent Downhole Gauges", PhD dissertation, Stanford University, California, 1999.

Aziz, K. and Settari, A., *Petroleum Reservoir Simulation*, Elsevier Applied Science Publishers, New York, 1979.

Balay, S., Gropp, W.D., McInnes, L.C., and Smith, B.F., *PETSc Users Manual*, ANL-95/11 - Revision 2.0.24, Argonne National Laboratory, 1999.

Bissel, R., Sharma, Y., and Killough, J.E., "History Matching Using the Method of Gradients: Two Case Studies", paper SPE 28590 presented at the 1994 SPE Annual Technical Conference and Exhibition, New Orleans, LA, September 25-28, 1994.

Bissel, R., "History Matching A Reservoir Model by the Positioning of Geological Objects", paper presented at the 5th European Conference on the Mathematics of Oil Recovery, Mining University, Leoben, Austria, September 3-6, 1996.

Capilla, J.E., Gómez-Hernández, J.J., and Sahuquillo, A., "Stochastic Simulation of Transmissivity Fields Conditional to Both Transmissivity and Piezometric Data, 2. Demonstration in a Synthetic Case," *J. of Hydrology*, 203:175-188, 1998.

Carter, R.D., Pierce, A.C., Kemp, L., and Williams, D.L., "Performance Matching With Constraints", *Soc. Pet. Eng. Journal* (April 1974) 187-196.

Chen, W.H., Gavalas, G.R., Seinfeld, J.H., and Wasserman, M.L., "A New Algorithm for Automatic History Matching", paper SPE 4545 presented at the 1973 SPE-AIME 48th Annual Fall Meeting, Las Vegas, NV, September 30 - October 3, 1973.

Chu, L., Reynolds, A.C., and Oliver, D.S., "Computation of Sensitivity Coefficients for Conditioning the Permeability Field to Well-Test Pressure Data", *In Situ* (1995) Vol. 19, 179-223.

Chu, L. and Komara, M., "An Efficient Technique for Inversion of Reservoir Properties Using Iteration Method", SPE 36512 presented at the 1996 SPE Annual Technical Conference and Exhibition, Denver, October 6-9, 1996.

Chu, L., Schatzinger, R.A., and Tham, M.K.: "Application of Wavelet Analysis to Upscaling of Rock Properties", Paper SPE 36517 presented at the 1996 SPE Annual Technical Conference and Exhibition, Denver, October 6-9, 1996.

Culler, D.E., Singh, J.P, and Gupta, A., *Parallel Computer Architecture, A Hardware/Software Approach*, Morgan Kaufmann Publishers, Inc., San Francisco, 1999.

Datta-Gupta, A., Vasco, D.W. and Long, J.C.S., "Sensitivity and Spatial Resolution of Transient Pressure and Tracer Data For Heterogeneity Characterization", paper SPE 30589, presented at the 1995 SPE Annual Technical Conference and Exhibition, Dallas, October 22-25, 1995.

Daubechies, I., "Orthonormal Bases of Compactly Supported Wavelets," *Comm. on Pure and Appl. Math.*, 41:909-996, November 1988.

de Marsily, G., Lavedan, G., Boucher, M., and Fasanino, G., "Interpretation of Interference Tests in a Well Field Using Geostatistical Techniques to Fit the Permeability Distribution in a Reservoir Model," *Geostatistics for Natural Resources Characterization, Part 2*, p831-849, 1984.

Deschamps, T., Grussaute, T., Mayers, D., and Bissel, R., "The Results of Testing Six Different Gradient Optimisers on Two History Matching Problems," Proceedings of the 6th European Conference on the Mathematics of Oil Recovery, B-24.

Dietrich, C.R., "Optimum Selection of Basis Functions for Ill-Posed Problems," vol. 195, p13-22, *Int. Assoc. of Hydrol. Sci.*

Deutsch, C. and Journel, A., *Geostatistical Software Library and User's Guide*, Oxford University Press, Oxford, 1997.

Foufoula-Georgiou E. and Kumar P. (eds.), *Wavelets in Geophysics*, Academic Press, Inc., 1994.

Gavalas, G.R., Shah, P.C., and Seinfeld, J.H., "Reservoir History Matching by Bayesian Estimation," *Soc. Petrol. Eng. J.*, 16(6):337-350, 1976.

Gill, P.E., Murry W., and Wright, M.H., *Practical Optimization*, Academic Press, New York, 1981.

Gómez-Hernández, J.J., Sahuquillo, A., and Capilla, J.E., "Stochastic Simulation of Transmissivity Fields Conditional to Both Transmissivity and Piezometric Data, 1. The Theory," *J. of Hydrology*, 203:162-174, 1998.

He, N., Reynolds, A.C. and Oliver, D.S., "Three-Dimensional Reservoir Description From Multiwell Pressure Data and Prior Information", paper SPE 36509, presented at the 1996 SPE Annual Technical Conference and Exhibition, Denver, October 6-9, 1996.

Huang, X. and Kelkar, M., "Integration of Dynamic Data for Reservoir Characterization in the Frequency Domain", SPE 36513 presented at the 1996 SPE Annual Technical Conference and Exhibition, Denver, October 6-9, 1996.

Jacquard, P. and Jain, C., "Permeability Distribution From Field Pressure Data", *Soc. Pet. Eng. Journal* (December 1965) 281-294.

Jansen, F.E. and Kelkar M.G., "Application of Wavelets to Production Data in Describing Inter-Well Relationships", paper SPE 38876, presented at the 1997 SPE Annual Technical Conference and Exhibition, San Antonio, October 5-8, 1997.

Kikani, J. and He, M.: "Multiresolution Analysis of Long-Term Pressure Transient Data Using Wavelet Methods", Paper SPE 48966 presented at the 1998 SPE Annual Technical Conference and Exhibition, New Orleans, September 27-30, 1998.

Landa, J.L., Kamal, M.M., Jenkins, C.D., and Horne, R.N., "Reservoir Characterization Constrained to Well Test Data: A Field Example", paper SPE 36511 presented at the 1996 SPE Annual Technical Conference and Exhibition, Denver, October 6-9, 1996.

Landa, J.L., "Reservoir Parameter Estimation Constrained to Pressure Transients, Performance History and Distributed Saturation Data", PhD dissertation, Stanford University, California, 1997.

Landa, J.L. and Horne, R.N., "A Procedure to Integrate Well Test Data, Reservoir Performance History and 4-D Seismic Data Into a Reservoir Description", paper SPE 38653, presented at the 1997 SPE Annual Technical Conference and Exhibition, San Antonio, October 5-8, 1997.

LaVenue, A.M. and Pickens, J.F., "Application of a Coupled Adjoint Sensitivity and Kriging Approach to Calibrate a Groundwater Flow Model," *Water Resour. Res.*, 28(6):1543-1569, 1992.

LaVenue, A.M., RamaRao, B.S., de Marsily, G., and Marietta, M.G., "Pilot Point Methodology for Automated Calibration of an Ensemble of Conditionally Simulated Transmissivity Fields, 2. Application," *Water Resour. Res.*, 31(3):495 -516, 1995.

Menke, W., *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, Orlando, Florida, 1984.

Mallat, S., "An Efficient Image Representation for Multiscale Analysis", in Proc. of Machine Vision Conference, Lake Tahoe, February 1987.

Mallat, S., "A Wavelet Tour of Signal Processing", Academic Press, San Diego, 1998.

Mallat, S., "Multiresolution Approximations and Wavelet Orthonormal Bases of $\mathbf{L}^2(\mathcal{R})$", *Trans. Amer. Math. Soc.*, 315:69-87, September 1989.

Mallat, S., "A Theory for Multiresolution Signal Decomposition: the Wavelet Representation", *IEEE Trans. Patt. Recog. and Mach. Intell.*, 11(7):674-693, July 1989.

McLaughlin, D. and Townley, L.R., "A Reassessment of the Groundwater Inverse Problem," *Water Resour. Res.*, 32(5):1131-1161, 1996.

Message Passing Interface Forum, *MPI-2: Extensions to the Message-Passing Interface*, University of Tennessee, Knoxville, Tennessee, 1997.

Meyer, Y., *Wavelets and Operators*, Cambridge University Press, 1992.

Grand Challenges 1993: High Performance Computing and Communications, A Report by the Committee on Physical, Mathematical, an Engineering Sciences, Washington, DC, Office of Science and Technology Policy.

Ogden, R.T., "Essential Wavelets for Statistical Applications and Data Analysis", Birkhauser Boston, 1997.

Oldenburg, D.W., McGillivray, P.R., and Ellis, R.G., "Generalized Subspace Methods for Large-Scale Inverse Problems," *Geophys. J. Int.*, 114(1):12-20, 1993.

Oliver, D.S., "Incorporation of Transient Pressure Data into Reservoir Characterization", *In Situ* (1994), Vol. 18, 243-275.

Oliver, D.S., "Multiple Realizations of the Permeability Field for Well Test Data", paper SPE 27970 presented at the University of Tulsa Centennial Petroleum Symposium, Tulsa, Aug. 29-31, 1994.

Oliver, D.S., "Multiple Realizations of the Permeability Field from Well-Test Data," *Soc. Petrol. Eng. J.*, 1(2):145-154, 1996.

Oliver, D.S., Reynolds, A.C., Abacioglu, Y., and Bi, Z., "Integration of Production Data into Reservoir Models," *Petroleum Geoscience*, (to appear in 2001).

Omre, H., Tjelmeland, H., and Wist, H.T., "Uncertainty in History Matching – Model Specification and Sampling Algorithms," Technical Report Statistics No.6/1999, Department of Mathematical Sciences, Norwegian University of Science & Technology, Trondheim, Norway.

Panda, M.N., Mosher, C., and Chopra, A.K.: "Application of Wavelet Transforms to Reservoir Data Analysis and Scaling", Paper SPE 36516 presented at the 1996 SPE Annual Technical Conference and Exhibition, Denver, October 6-9, 1996.

Parker, R., *Geophysical Inverse Theory*, Princeton University Press, Princeton, New Jersey, 1994.

Polak, E., *Computational Methods in Optimization*, Academic Press, New York, 1971.

Peaceman, D.W., *Fundamentals of Numerical Reservoir Simulation*, Elsevier, Amsterdam, 1977.

Phan, V.Q., "Inferring Depth-Dependent Reservoir Properties from Integrated Analysis Using Dynamic Data", MS report, Stanford University, California, 1998.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P., *Numerical Recipes In C*, 2nd ed., Cambridge University Press, New York, 1992.

Rao, C.R., *Linear Statistical Inference and Its Applications*, Wiley, New York, 1973.

Reynolds, A.C., He, N., Chu, L., and Oliver, D.S., "Reparameterization Techniques for Generating Reservoir Descriptions Conditioned to Variograms and Well-Test Pressure Data", SPE 30588, presented at the 1995 SPE Annual Technical Conference and Exhibition, Dallas, October 22-25, 1995.

Saad, Y. and Schultz, M.H., "GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems," SIAM J. Sci. Stat. Compt., July 1986, Vo.7, No. 3, pp.856-869.

Shah, P.C., Gavalas, G.R., and Seinfeld J.H., "Error Analysis in History Matching: The Optimal Level of Parameterization," *Soc. Petrol. Eng. J.*, 18(6):219-228, 1978.

Tan, T.B. and Kalogerakis, N., "A Fully Implicit, Three-Dimensional, Three-Phase Simulator with Automatic History-Matching Capability", paper SPE 21205 presented at the 1991 SPE 11th Symposium on Reservoir Simulation, Anaheim, CA, February 17-20, 1991.

Tan, T.B., "A Computationally Efficient Gauss-Newton Method for Automatic History Matching", paper SPE 29100 presented at the 1995 SPE Symposium on Reservoir Simulation, San Antonio, TX, February 12-15, 1995.

Tang, Y.N., Chen, Y.M., Chen, W.H., and Wasserman, M.L., "Generalized Pulse-Spectrum Technique for 2-D and 2-Phase History Matching", *Applied Numerical Mathematics* (1989) Vol. 5, 529-539.

Tarantola, A. and Valette B., "Generalized Nonlinear Inverse Problems Solved Using the Least Squares Criterion", *Reviews of Geophysics and Space Physics* (May 1982) Vol. 20, 219-232.

Tarantola, A., *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier Science Publishers, Amsterdam, 1987.

Thiele, M.R., Batycky, R.P., and Blunt, M.J., "A Streamline-Based 3D Field-Scale Compositional Reservoir Simulator," paper SPE 38889 presented at the 1997 SPE Annual Technical Conference and Exhibition held in San Antonio, Texas, October 5-8, 1997.

Vasco, D.W., Yoon, S., and Datta-Gupta, A., "Integrating Dynamic Data into High-Resolution Reservoir Models Using Streamline-Based Analytic Sensitivity Coefficients," paper SPE 49002 presented at the 1998 SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, September 27-30, 1998.

Wang, P., "Integrating Resistivity Data into Parameter Estimation Problem", MS report, Stanford University, California, 1999.

Watson, A.T., Seinfeld, J.H., Gavalas, G.R., and Woo, P.T., "History Matching in Two-Phase Petroleum Reservoirs", paper SPE 8250 presented at the 1979 SPE 54th Annual Technical Conference and Exhibition, Las Vegas, NV, September 23-26, 1979.

Wen, X., Deutsch, C.V., and Cullick, A.S., "High Resolution Reservoir Models Integrating Multiple-Well Production Data," paper SPE 38728 presented at the 1997 SPE Annual Technical Conference and Exhibition in San Antonio, Texas, October 5-8, 1997.

Wen, X., Deutsch, C.V., and Cullick, A.S., "Integrating Pressure and Fractional Flow Data in Reservoir Modeling With Streamline-Based Inverse Method," paper SPE 48971 presented at the 1998 SPE Annual Technical Conference and Exhibition in New Orleans, Louisiana, September 27-30, 1998.

Wu, Z., Reynolds, A.C. and Oliver, D.S., "Conditioning Geostatistical Models to Two-Phase Production Data", paper SPE 49003, presented at the 1998 SPE Annual Technical Conference and Exhibition, New Orleans, September 27-30, 1998.

Yang, P.H. and Watson, A.T., "Automatic History Matching With Variable-Metric Methods", paper SPE 16977 presented at the 1987 SPE 62nd Annual Technical Conference and Exhibition, Dallas, TX, September 27-30, 1987.

Yoon, S., Datta-Gupta, A., Vasco, D.W., and Behrens, R.A., "A Multiscale Approach to Production Data Integration Using Streamline Models," paper SPE 56653 presented at the 1996 SPE Annual Technical Conference and Exhibition, Houston, Texas, October 3-6, 1999.