Bayesian Seismic Waveform Data Inversion: Parameter Estimation and Uncertainty Analysis

Wenceslau Peres Gouveia, Jr.

— Doctoral Thesis —
Geophysics

Center for Wave Phenomena
Colorado School of Mines
Golden, Colorado 80401
303/273-3557
This work is dedicated to the memory of my father Wenceslau Peres Gouveia and to the memory of Professor Jack K. Cohen.
ABSTRACT

When solving an inverse problem one has to address three basic questions:

1. How accurately are the data known? How can its uncertainties be quantified, so that one can define to what extent it is important to fit the data?

2. How accurately is the physical system modeled? Does the model include all the physical factors that contribute significantly to the data? If not, how can such modeling errors be incorporated into an inverse calculation?

3. Is it possible to incorporate information about the physical system that is independent of the data? If so, what is the influence of this external information on the inverse calculation?

This work addresses all the above questions in the context of the seismic waveform inversion problem, under the framework of Bayesian inference. In this approach the solution of the inverse problem is not a single model of the subsurface, but a probability distribution on the space of models. This (a posteriori) probability distribution consists of the product of two terms. The first one (the likelihood function) defines what it means for a model to fit the data, and the second one (the a priori distribution) allows the incorporation of information about the subsurface that is independent of the data into the inverse calculation. In this way uncertainties about the observed data and about the underlying subsurface can be considered a priori in the inversion procedure. These are so-called data and model uncertainties, respectively. Questions of resolution of the computed estimates are answered by the a posteriori probability distribution. Theoretically this step can be executed by Monte Carlo methods, without any approximations. However, due to the high computational costs involved in this procedure, mainly due to the forward modeling seismic operator and the high dimensionality of the problem, I use a Gaussian approximation to the posterior in the assessment of the uncertainties.

Here, this technique is demonstrated on synthetic and field seismic data sets. The a priori distribution is derived from well-log measurements, and data and modeling uncertainties are taken into consideration in the inverse calculation. Uncertainties associated with the inverse calculation are derived from the a posteriori probability distribution in three different ways: via one-standard deviation error bars on the estimates of the subsurface, pseudo-random realizations of the subsurface and marginal distributions. Finally, I compare the Bayesian approach with a more conservative inversion procedure, in which the solution to the problem is the simplest model (in the sense of smoothness) that is still able to fit the data to a given accuracy. This more conservative approach is usually denominated Occam inversion. This comparison affords a clear illustration of the differences between the Bayesian and the Occamist
points of view, highlighting the actual role of the apriori information in the Bayesian inversion calculation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Start Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td></td>
<td>vii</td>
</tr>
<tr>
<td><strong>Chapter 1</strong></td>
<td><strong>GENERAL INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Preamble</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Previous Research</td>
<td>2</td>
</tr>
<tr>
<td>1.3</td>
<td>Thesis outline</td>
<td>5</td>
</tr>
<tr>
<td>1.4</td>
<td>Contributions of this work</td>
<td>6</td>
</tr>
<tr>
<td><strong>Chapter 2</strong></td>
<td><strong>BAYESIAN INFEERENCE: AN OVERVIEW</strong></td>
<td>7</td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>7</td>
</tr>
<tr>
<td>2.1.1</td>
<td>According to Jeffreys</td>
<td>7</td>
</tr>
<tr>
<td>2.1.2</td>
<td>According to Tarantola and Valette</td>
<td>9</td>
</tr>
<tr>
<td>2.1.3</td>
<td>Discussion</td>
<td>11</td>
</tr>
<tr>
<td><strong>Chapter 3</strong></td>
<td><strong>PRIOR PROBABILITIES</strong></td>
<td>13</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>13</td>
</tr>
<tr>
<td>3.2</td>
<td>On Entropy</td>
<td>14</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Definition of entropy</td>
<td>14</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Extension to Continuous Probabilities</td>
<td>15</td>
</tr>
<tr>
<td>3.2.3</td>
<td>The Principle of Maximum Entropy</td>
<td>16</td>
</tr>
<tr>
<td>3.3</td>
<td>Computation of the Lagrange Multipliers</td>
<td>17</td>
</tr>
<tr>
<td>3.4</td>
<td>Gaussian <em>a priori</em> probability distribution from well logs</td>
<td>21</td>
</tr>
<tr>
<td><strong>Chapter 4</strong></td>
<td><strong>NONLINEAR BAYESIAN INVERSION. APPLICATION TO SYNTHETIC DATA.</strong></td>
<td>27</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>27</td>
</tr>
<tr>
<td>4.2</td>
<td>The Bayesian Inversion Procedure</td>
<td>28</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Model parameterization</td>
<td>30</td>
</tr>
<tr>
<td>4.2.2</td>
<td>The Objective Function</td>
<td>31</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Resolution</td>
<td>32</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Numerical Calculations</td>
<td>34</td>
</tr>
<tr>
<td>4.3</td>
<td>Occam Inversion</td>
<td>48</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Resolution</td>
<td>51</td>
</tr>
</tbody>
</table>
Chapter 5  NONLINEAR BAYESIAN INVERSION. APPLICATION TO FIELD DATA. ............ 59

5.1 Introduction ................................................. 59
5.2 Regional Geology ......................................... 59
5.3 The Sorrento Data: Surface Seismic Data and Well Logs ................. 60
5.4 A priori Uncertainties ...................................... 65
  5.4.1 Data Uncertainties ...................................... 69
  5.4.2 Model Uncertainties ...................................... 81
5.5 Explaining the recorded data .................................. 83
  5.5.1 Discussion ................................................. 92
5.6 Uncertainty Analysis ......................................... 97
  5.6.1 The Occam approach ...................................... 112
5.7 General discussion .......................................... 116
5.8 Interpreter’s corner .......................................... 117

Chapter 6  CONCLUSIONS ............................................ 125

6.1 Computational cost ........................................... 129
6.2 Future work .................................................. 129
  6.2.1 Data uncertainties ....................................... 129
  6.2.2 Two- and Three-dimensional models ....................... 130
  6.2.3 Breaking through the Gaussian hypothesis .................. 130
6.3 Developed software ........................................... 131

References ......................................................... 133

Appendix A  THE FRECHÉT DERIVATIVES OF THE DISPLACE-
MENT FIELD ..................................................... 137

Appendix B  DISTRIBUTED-COMPUTER IMPLEMENTATION .... 143

Appendix C  APPENDIX C: LINEARIZING THE INVERSE PROB-
LEM .......................................................... 147

  C.1 Bayesian Inversion .......................................... 147
  C.1.1 Occam inversion .......................................... 147

Appendix D  APPENDIX D: RESOLUTION OF THE FORWARD MOD-
ELING OPERATOR ............................................... 151
ACKNOWLEDGMENTS

I am thankful to Eliane, wife and forever girlfriend, for her unlimited patience with me during these four years of graduate school. Her love, courage, and friendliness gave me the strength that I needed to accomplish my degree. I owe you a lot, sweetheart, and I was very lucky on the day you agreed to dance that music with me, and we began to date. The music is still playing, and loud. I am grateful, also, to you little Bernardo, who dropped into our lives as a gift, who jumped on my books and who, whenever I was not looking, turned off my computer. Small as you are, you have already taught me that there is more to life than just work and convinced me that, without you, this whole thing would be meaningless. Thank you, Mom and Dad, for trusting me no matter what and for teaching me the value of education. This thesis is dedicated to my father, my best friend, who will be always present in my memories. I also want to dedicate this work to Professor Jack K. Cohen, a good friend of all who knew him.

I am greatly indebted to my advisor, John Scales, and my committee members, Ken Larner, Ilya Tsvankin, Norm Bleistein, Jack Cohen, Tim Cross and Kadri Dagdelen for the guidance and stimulating scientific discussions, that happened many times during the years of graduate school. Special thanks go to John Scales, for introducing me to the world of inverse theory, which is the main subject of my dissertation. He kept my motivation high by always having his door opened to me for countless discussions on this and on many other subjects. One can learn a lot from John, and I certainly did. I am grateful to Ken and Norm for their (very!) thorough review of this manuscript, making it much more readable. I also want to express my gratitude to Tim Cross, for accepting the impossible task of teaching concepts of stratigraphy to me, an electrical engineer. Even more, he showed me that it is possible for people with completely different backgrounds to work efficiently together. This is a valuable lesson, especially in the interdisciplinary field of geophysics. Thanks are also due to Professor Tom Davis, Steve Roche and Robert Benson for providing me access to the Sorrento Data that I used in this thesis, and for helping me with their expertise in many of the data processing issues that were relevant to my work.

I am indebted to the many good friends I made at the Colorado School of Mines: Tariq Alkhalifah, Craig Artley, Lydia Deng, Bryan Devault, Tagir Galikeev, Bruce Mattocks, Fabiola Williams, Fernando Moraes, Andreas Rüger, Gilein Steensma, Timo Tjan, Antonio Ramos, Alberto Villarreal and many others. Thank you for standing by me during these CSM years. It has been an unique experience to have friends from all over the world. I am also grateful to Doug Hart, from Western Geophysical Company, and John Anderson, from Mobil Oil Company, for the constant encouragement on my research. Doug was also very kind in sharing with me his knowledge on optimization theory, and also for giving me the opportunity to work at
Western for two summers. To my special friends Jo Ann Fink, Barbara McLennon
and Marge Hoage: a sincere hug with tenderness. Finally, thanks to Professor Paulo
Osório and Eduardo Faria for being so helpful in the beginning of my career as a
geophysicalist.

I gratefully acknowledge the support of the sponsors of the Consortium Project on
Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena,
Colorado School of Mines, the Army Research Office under grant DAAH04-95-1-0173
and the National Science Foundation under grant DMS-9506603.
Chapter 1

GENERAL INTRODUCTION

1.1 Preamble

The purpose of solving an inverse problem is to make quantitative statements about physical systems from recorded data. Such statements must address the uncertainties of the parameter estimates of the physical system under consideration. This requires taking into account three distinct questions:

1. How accurately are the data known? How can its uncertainties be quantified, so that one can define to what extent it is important to fit the data?

2. How accurately is the physical system modeled? Does the model include all the physical factors that contribute significantly to the data? If not, how can such modeling errors be incorporated into an inverse calculation?

3. Is it possible to incorporate information about the physical system that is independent of the data? If so, what is the influence of this external information on the inverse calculation?

Here, I propose a practical methodology for addressing the above questions as they arise in the problem of seismic-waveform data inversion. I use the theory of Bayesian inference to estimate from surface seismic waveform data P-wave, S-wave impedance and density profiles, and their respective uncertainties. Information about the subsurface that is independent of the seismic data, namely well-log measurements, is also used in the inversion calculation. In the following chapters, I describe this procedure in detail and expose the theoretical and practical difficulties related to its application to surface seismic data.

To learn about the subsurface is, of course, the driving force behind acquisition, processing and interpretation of any geophysical data. Several procedures have been developed to extract information about the subsurface from different kinds of geophysical data, such as seismic, gravity, magnetic, electrical and electromagnetic data. More specifically with respect to seismic data, the main approach used in this study, those techniques must handle distinct acquisition geometries, a large variety of components that are considered noise (e.g., ambient noise, ground-roll, multiples and so on), sizeable amount of data (tens of millions of traces are acquired in a standard three-dimensional survey), etc.

Such procedures can be divided in two main categories. The first includes so-called data processing techniques (e.g., Yilmaz 1987; Dobrin and Savit 1988; and Telford et al. 1990), which are, indeed, the approach used in routine processing within
the exploration industry. Those techniques can be thought of as sequence of data pre-
conditioning steps that aim at tackling one specific issue at a time, until a description, or image, of the subsurface is attained. Often, the uncertainties associated with the resulting subsurface images are not assessed by these data-processing techniques.

The second category, to which the work developed here belongs, includes so-called inversion or inference procedures. Such approaches aim at explaining the data with an estimated model of the subsurface and also aim at quantifying the uncertainties associated with the estimates. This is carried out as a two-step procedure. In the first, one attempts to explain the observed data with an idealized model of the subsurface. Such a model can be obtained in different ways. For example, Cohen et al. (1985) and Bleistein et al. (1987), among others, derived wave-equation-based inversion operators that estimate subsurface reflectivity information from seismic data. In an alternative approach (Tarantola, 1987), which is adopted here, an initial model of the subsurface is refined until it fits the observed data with synthetic data generated with a forward modeling operator. In that approach, the degree of reliability of a given model or models that could explain the observed data is quantified (Tarantola 1987 and Parker 1994). Often, results from processing algorithms are used to pre-condition data for inference calculations. This is done either to improve the data quality, by attenuating some of its noise components, or by using subsurface models derived from processing algorithms as initial models that are refined by the inference procedure.

The possibility of being able to systematically incorporate diverse types of information and make quantitative statements about resolution of the subsurface estimates is the main driving force of inference or inversion procedures. Probability theory is used in this approach, since errors (noise) of the observed data and uncertainties related to possible subsurface models can be characterized by probability distributions, as will be discussed in detail in the next chapters.

There are, however, two different interpretations of probabilities. In one, the “frequency” interpretation, probabilities are defined as frequency of occurrence, determined from repeated experiments. In the other, the “Bayesian” interpretation, probabilities are used to quantify states of informations about a physical system. An interesting discussion on the distinction between these views can be found in Tarantola (1987) and Scales and Smith (1995).

The work done here adopts the Bayesian interpretation of probabilities. In the following, I review previous work that have been done in this arena and present the outline of this thesis, describing its objectives and original contributions.

1.2 Previous Research

Making inferences from data is one of the fundamental goals of science. The immediate purpose of acquiring, processing and interpreting geophysical data is to estimate properties of the Earth’s subsurface. Such an inference procedure can be understood as one that refines our understanding about a space of parameters used to represent a physical model under consideration, the subsurface in this study. As an
example, one might begin the inversion calculation with the assumption that the true subsurface parameters lie within a given interval, which represents one’s initial state of information. As the inversion is carried out, such a knowledge about the subsurface is refined, in the sense that the width of the intervals is reduced. In Bayesian theory, these states of information are associated with probability distributions. For example, lack of information about the Earth density at a given depth can be modeled by a broad probability distribution. This probability distribution, specified in the density domain, incorporates the range of density values one thinks of as reasonable for that depth.

The works of Jeffreys (1939) and Jaynes (1957) were fundamental in formalizing the mathematical methodology of Bayesian data inference. Although I will provide an overview of this methodology in the next chapter, let me now review its basic principles.

In Bayesian inversion the solution to an inverse problem is not a single parameterized model of the subsurface, but a probability distribution on the space of models $\mathbf{m}$. This distribution, the so-called a posteriori probability, which I will refer to throughout as $\sigma(\mathbf{m})$, consists of the product of two probability density functions. The first, the likelihood function $L(\mathbf{m})$, defines what it means for a model to fit the data. The likelihood function quantifies the misfit between the observed data $\mathbf{d}_{\text{obs}}$ and modeled or synthetic data generated by a suitable forward modeling operator $g(\mathbf{m})$. Thus, this function takes into account the errors (noise) in the observed data, as well as errors in the forward modeling procedure (for example, errors associated with modeling an elastic subsurface with an acoustic operator). The second probability density function, referred to as $\rho(\mathbf{m})$, incorporates into the inverse calculation information about the subsurface that is not contained in the data. In other words, $\rho(\mathbf{m})$ represents our ideas of model possibilities gained from some source other than the data being fitted. Thus, the a posteriori probability function is given by

$$
\sigma(\mathbf{m}) \propto \rho(\mathbf{m}) L(\mathbf{m}). \quad (1.1)
$$

In Bayesian inference, all statements on the resolution and uncertainties of the estimates of the subsurface parameters are derived from $\sigma(\mathbf{m})$. Such statements can be expressed in many ways; for example, as error bars on the parameter estimates, marginal a posteriori probability densities, or as pseudo-random realizations of the subsurface generated by sampling $\sigma(\mathbf{m})$. All of those will be explored in this work.

It is illustrative to consider the case where both the likelihood and the a priori probability distributions are Gaussian and defined by

$$
L(\mathbf{m}) \propto \exp \left[ -\frac{1}{2} (g(\mathbf{m}) - \mathbf{d}_{\text{obs}})^T \left( g(\mathbf{m}) - \mathbf{d}_{\text{obs}} \right) \right],
$$

$$
\rho(\mathbf{m}) \propto \exp \left[ -\frac{1}{2} \mathbf{m}^T C_M^{-1} \mathbf{m} \right]. \quad (1.2)
$$

This model assumes that the noise in the data is Gaussian, uncorrelated and identically
distributed, with unit variance. The \textit{a priori} probability \( \rho(\mathbf{m}) \) is, here, a zero-mean Gaussian with covariance matrix \( C_M \). The product of those two distributions yields the following expression for the \textit{a posteriori} probability \( \sigma(\mathbf{m}) \)

\[
\sigma(\mathbf{m}) \propto \exp \left\{ -\frac{1}{2} \left[ (g(\mathbf{m}) - \mathbf{d}_{\text{obs}})^T (g(\mathbf{m}) - \mathbf{d}_{\text{obs}}) + \mathbf{m}^T C_M^{-1} \mathbf{m} \right] \right\} \tag{1.3}
\]

The maximum \textit{a posteriori} (MAP) estimator, \( \mathbf{m}_{\text{map}} \), is governed as the model associated with the maximum value of \( \sigma(\mathbf{m}) \), which can be found by solving the following optimization problem

\[
\min_{\mathbf{m}} \left[ \|g(\mathbf{m}) - \mathbf{d}_{\text{obs}}\|_2 + \|C_M^{-\frac{1}{2}} \mathbf{m}\|_2 \right] , \tag{1.4}
\]

which is a standard least-squares procedure with a penalty term defined by the covariance matrix. Therefore, under the Bayesian framework, a least-squares calculation such as (1.4), follows when the uncertainties are Gaussian, expressed by the probability distributions in Equation (1.2).

The Bayesian approach was pioneered in geophysics by Tarantola and co-workers (Tarantola and Valette 1982 and Tarantola 1987). Tarantola’s book can be considered an encyclopedia in data fitting and parameter estimation. The problem formulation adopted there is general enough to be applied in a wide variety of applications, and many of those have received note in the technical literature. Those applications differ, basically, in two aspects. The first is related to their assumptions about the complexity of the subsurface: if it is an acoustic or elastic medium, and whether it can be represented as a one-, two- or three-dimensional model. The second concerns the physics of the wave propagation, modeled by the forward modeling operator \( g(\mathbf{m}) \): whether it is a linear or a non-linear procedure, or if it can be described by ray- or wave equation-based methods.

As examples of nonlinear Bayesian inversion procedures, Keys (1986), Dietrich and Kormendi (1990), Oldenburg (1984) and Pan et al. (1994), among others, estimated one-dimensional impedance and density profiles from waveform seismic data. Nonlinear Bayesian inversion methods applied to two-dimensional elastic media, using finite differences as the modeling procedure, have been studied by Mora (1987) among others. He built an optimization algorithm to compute the MAP estimate from an initial guess of the subsurface. He also addressed some of the practical complications that arise with this method due to lack of proper coverage of the subsurface. He has also shown how the MAP estimate can be improved when transmitted data (vertical seismic profiles in his work) were used in conjunction with surface seismic data. For the same problem, Crase et al. (1990) illustrated the sensitivity of the MAP estimate to the choice of the likelihood function with numerical examples. Four choices were investigated by the authors: the Gaussian, Laplace, Secant and Cauchy distributions. Their conclusions show the Cauchy distribution as the one providing estimates that are least sensitive to noise in the data. All these references lack assessment of uncertain-
ties associated with the observed data and forward modeling procedure. Moreover, the a priori probability distribution $\rho(m)$ was either uniform or built according to preconceptions of the subsurface, such as smoothness, which are not generally supported by any quantitative information about the subsurface.

The works of Scales and Tarantola (1994), Mosegaard and Tarantola (1995) and Mosegaard et al. (1995) illustrate nonlinear Bayesian calculations where the a posteriori probability is analyzed via Monte Carlo procedures. This is justified by the multimodal character of the a posteriori probability distribution used in their formulation of the inverse problem. As will be explained later, I make a Gaussian approximation of this distribution to allow use of less expensive techniques for its analysis. This approximation is motivated by the high computational cost of Monte Carlo procedures, making their application in the problem of multi-offset seismic waveform inversion impractical.

In the Bayesian inversion procedure done in this thesis, I assess the uncertainties in the observed data and in the forward modeling procedure, as well as the uncertainties in the a priori information about the subsurface. Once this is done, the a posteriori probability distribution is built and, to expedite its analysis, is approximated by a Gaussian distribution about a MAP model. As will be detailed in Chapter 4, this MAP model is computed by a local optimization technique applied to an initial model, chosen a priori. This Gaussian approximation is then analyzed, in order to draw conclusions about resolution and uncertainties of the estimates. Thus, the work presented here has features in common with some of those listed in the above review of the literature. However, the analysis of the uncertainties associated with the problem of waveform inversion, as will be discussed in the following chapters, is one of the new contributions presented in this dissertation.

1.3 Thesis outline

In the next chapter, I review the important points of Bayesian inversion by describing its methodology and how it can be applied to the specific problem of seismic waveform inversion. Chapter 3 describes how to construct a priori probabilities from well logs under the approximation that such measurements can be modeled by a Gaussian process. This a priori probability distribution will be used in the inverse calculations performed here. This approximation is also used for the likelihood function $L(m)$. Going beyond the Gaussian assumption is a difficult matter, as will be discussed in this chapter. A promising approach to do so is the maximum-entropy approach (Jaynes 1957 and 1968 and, for a tutorial, Gouveia et al. 1996). In Chapter 3, I show some preliminary results from the use of this technique to compute non-Gaussian a priori distributions.

Chapter 4 illustrates the application of Bayesian inversion to a realistic synthetic problem in which elastic surface seismic data are inverted for impedance and density. In this calculation I construct the a priori distribution from hypothetical well-log measurements. This chapter also shows a comparison of the results provided by the
W.P. Gouveia, Jr.

Bayesian procedure with those obtained with an alternative non-Bayesian procedure, the goal of which is to find the smoothest model that is still able to fit the data to a comparable accuracy. This “featurelessness” is achieved via Tikhonov regularization in a procedure named “Occam’s inversion.” In Chapter 5, I present results obtained with the Bayesian inversion of a field (land) seismic data set. This data set, provided to me by Prof. Thomas Davis (of the CSM Reservoir Characterization Project), was acquired at the Sorrento Basin near the Las Animas Arch, in Southeastern Colorado. I chose this data set because it is well understood that the geology down to the reservoir level is stratified with little structural complexity and that anisotropy is not a significant factor at Sorrento Basin (RCP Phase V, 1995). These characteristics make the modeling algorithm used in this thesis (the elastic isotropic version of the reflectivity method of Fuchs and Müller 1971) a suitable modeling algorithm for this problem. Finally, the last chapter addresses conclusions and future research directions.

1.4 Contributions of this work

Here, I accomplish a Bayesian inference calculation in which I estimate subsurface parameters from surface seismic waveform data. This is done with a thorough analysis of the uncertainties associated with the inverse problem. Such uncertainties are related to errors in the observed data, to the forward modeling procedure and also to the seismic data-independent information about the subsurface used to build the \( \rho(\mathbf{m}) \). Specifically, this information is derived from a suite of well-log measurements. Once the uncertainties are quantified they are used in the \( \sigma(\mathbf{m}) \), which is analyzed in order to assess the uncertainties and resolution of the subsurface estimates.

I test the inversion procedure proposed here on realistic synthetic and field data sets, described in Chapters 4 and 5, respectively. In both cases, the analysis of the \( \sigma(\mathbf{m}) \) distribution is expressed in terms of error bars on the parameter estimates, pseudo-random realizations of the subsurface sampled from this probability and marginal \( \sigma(\mathbf{m}) \) probability densities.

The extent to which this uncertainty analysis is carried out in the context of seismic waveform inversion is a new contribution of this work. Moreover, the construction of \( \rho(\mathbf{m}) \) probabilities from well logs measurements is also an important, and new, contribution to the field of inverse theory. This is in contrast to what is done in so-called regularized inversion procedures, wherein assumptions are made, such as smoothness, in order to stabilize the technique. One of the main objectives of this thesis is to honor the true concept of the \( \rho(\mathbf{m}) \): to encapsulate data-independent information in the form of a probability distribution and incorporate it into the inverse problem.

Other authors have attempted to apply Bayesian inversion procedures to seismic waveform data before. To my knowledge, however, it is the first time that an inverse calculation is performed along the Bayesian methodology with the formalism of quantifying data and model uncertainties developed here.
Chapter 2

BAYESIAN INFERENCE: AN OVERVIEW

2.1 Introduction

Two distinct schools of thought can be recognized in the literature on Bayesian inference. In one, the more classical approach popularized by the work of Jeffreys (1939), all inference statements are derived from the basic product and addition rules of probability. The other, a more recent interpretation introduced in Tarantola and Valette (1982), is based on the postulate that a state of information about a parameter space (for example, the space of subsurface models \( m \)), has a probability distribution as its more general description. Therefore, according to their view, Bayesian inference is a procedure in which such states of information are successively refined as data and further information on the underlying subsurface become available.

The work presented here falls under the framework of Tarantola and Valette (1982), which will be reviewed in this chapter. However, for completeness, in the next section I present an overview of the methodology adopted by Jeffreys. As has been demonstrated (Tarantola 1987; Scales and Smith 1995; and Moraes 1996), a connection can be established between these two schemes, using the Bayes theorem.

2.1.1 According to Jeffreys

There are only two basic rules for manipulating probabilities (Brethorst, 1990): the product rule and the sum rule. All other probability rules, and therefore inference statements, can be derived from these. Such basic rules are as follows. Consider that \( A, B \) and \( C \) are three arbitrary propositions. The product rule is given by

\[
P(A, B|C) = P(A|C) P(B|A, C),
\]

where \( P(A, B|C) \) is the joint conditional probability that “\( A \) and \( B \) is true given that \( C \) is true”. It is important to mention that for Bayesians who work in accordance with Jeffreys’ methodology, all probabilities are conditional. In this sense, probabilities refer to the likelihood of a given proposition (\( A \) and \( B \)) given that another one (\( C \)) is true.

The sum rule of probability theory is given by

\[
P(A + B|C) = P(A|C) + P(B|C) - P(A, B|C),
\]

where \( P(A + B|C) \) is the probability that “\( A \) or \( B \) is true given that \( C \) is true”. When
W.P. Gouveia, Jr.

$A$ and $B$ are mutually exclusive propositions, the sum rule yields

$$P(A + B|C) = P(A|C) + P(B|C).$$  \hspace{1cm} (2.3)

The sum rule has a direct application to parameter-estimation problems in a technique called “marginalization.” Essentially this approach allows one to investigate a specific parameter while removing uninteresting (“nuisance”) parameters from consideration (Moraes, 1996). However, it is on the product rule that the major part of Bayesian inference relies. By rewriting Equation (2.1) as

$$P(B, A|C) = P(B|C) \, P(A|B, C),$$  \hspace{1cm} (2.4)

and equating the right hand side of Equations (2.1) and (2.4), one obtains the Bayes theorem

$$P(A|B, C) = \frac{P(A|C) \, P(B|A, C)}{P(B|C)}.$$  \hspace{1cm} (2.5)

Noticing that $P(B|C)$ is a normalization factor for $P(A|B, C)$, since it is obtained by integrating $A$ out in $P(A|B, C)$. That is

$$P(B|C) = \sum_{A} P(A, B|C)$$
$$= \sum_{A} P(A|C) \, P(B|A, C),$$  \hspace{1cm} (2.6)

Equation (2.5) is often expressed as

$$P(A|B, C) \propto P(A|C) \, P(B|A, C).$$  \hspace{1cm} (2.7)

Equation (2.7) defines the \textit{a posteriori} probability function on the event $A$. This probability allows inference statements on event $A$ taking into account events $B$ and $C$. By associating event $A$ with a given parameterization $\mathbf{m}$ of the subsurface, event $B$ with the geophysical data $\mathbf{d}_{\text{obs}}$ and event $C$ with the \textit{a priori} information $(\mathcal{I})$, we can rewrite this equation as

$$P(\mathbf{m}|\mathbf{d}_{\text{obs}}, \mathcal{I}) \propto P(\mathbf{d}_{\text{obs}}|\mathbf{m}, \mathcal{I}) \, P(\mathbf{m}|\mathcal{I}),$$  \hspace{1cm} (2.8)

which is essentially Equation (1.1), with $P(\mathbf{m}|\mathbf{d}_{\text{obs}}, \mathcal{I})$ as the \textit{a posteriori} distribution $\sigma(\mathbf{m})$, $P(\mathbf{d}_{\text{obs}}|\mathbf{m}, \mathcal{I})$ as the likelihood $L(\mathbf{m})$ and $P(\mathbf{m}|\mathcal{I})$ as the \textit{a priori} distribution $\rho(\mathbf{m})$. Note that the notation is deficient in the sense that the same symbol $P()$ is used for the likelihood and for the \textit{a priori} probabilities, although different distributions can be associated with each one of those. However, this is a common notation used in Jeffreys’ approach to the inference problem.
2.1.2 According to Tarantola and Valette

The basic premise that governs the reasoning advocated by Tarantola and Valette (1982) is the postulate that a probability measure is the most general description of any state of information on a finite space (either model or data space). To clarify matters, consider the example of complete ignorance and complete certainty about a given N-dimensional space. In the latter this probability density would be a delta function, centered at the only possible value that the space can take on. In the former this probability density would be uniform over the whole space.

This premise motivates the definition of a joint probability function \( \sigma(d,m) \), on the space of data \( d \) and models \( m \). Such a distribution incorporates what is known a priori about data and subsurface parameters, and also the physical relationship between them, specified as a forward modeling operator. In Tarantola and Valette’s approach, both are characterized as probability distributions, resulting in the following expression for \( \sigma(d,m) \):

\[
\sigma(d,m) = \frac{\rho(d,m) \Theta(d,m)}{\mu(d,m)}, \tag{2.9}
\]

where \( \rho(d,m) \) is the a priori probability distribution on data and models and \( \Theta(d,m) \) incorporates the physical relationship between those two spaces. More specifically, the former takes into account the uncertainties on the observed data and on the subsurface parameters; in the latter, the errors in the modeling procedure are considered. In other words, the inability to model all features of the physical system (in this case the Earth) is handled by \( \Theta(d,m) \). \( \mu(d,m) \) is the so-called null or reference state of information, which is required in the derivation of Equation (2.9) (Tarantola 1987, page 52). This function represents the state of least information about a parameter space. In this dissertation, as well as in the most practical applications, \( \mu(d,m) \) is simply a uniform distribution on the space of data and models.

Equation (2.9) is, in fact, derived from the definition of “conjunction of states of information” introduced by Tarantola and Valette. They have shown that, if \( P_1(x) \) and \( P_2(x) \) are two probability density functions representing two different states of information about the parameter space \( X \), a new (or updated) state of information can be derived as (Tarantola 1987, page 155)

\[
P_{12}(x) = \frac{P_1(x) P_2(x)}{\mu(x)}. \tag{2.10}
\]

\( P_{12}(x) \), the new state of information constructed from what was previously known about the space \( X \) is encapsulated as the probability distributions \( P_1(x) \) and \( P_2(x) \). Again, \( \mu(x) \) is the reference state of information. This function can be thought of as a normalization factor in Equation (2.10). This can be seen by considering the case where one of the probabilities, say \( P_1(x) \), represents the reference state of information itself, i.e. \( P_1(x) = \mu(x) \). Thus, \( P_{12}(x) = P_2(x) \), indicating that the updated state of information has not changed with respect to \( P_2(x) \), as expected. A thorough discussion
on the reference state of information \( \mu \) can be found in the first chapter of Tarantola’s book.

Two hypotheses are considered in Tarantola (1987) to simplify Equation (2.9). In the first it is assumed that the \( a \ priori \) information and the non-informative states of information on data and models, are separable. In other words, data and models are independent with respect to the non-informative and \( a \ priori \) distributions. This is expressed as

\[
\begin{align*}
\rho(d, m) &= \rho_d(d) \rho_m(m), \\
\mu(d, m) &= \mu_d(d) \mu_m(m),
\end{align*}
\]

(2.11)

where \( \rho_d(d) \) and \( \rho_m(m) \) are the \( a \ priori \) probability distributions and \( \mu_d(d) \) and \( \mu_m(m) \) represent non-informative states of information, for data and models respectively.

In the second hypotheses, the joint probability distribution \( \Theta(d, m) \) is written as

\[
\Theta(d, m) = \Theta(d|m) \mu_m(m).
\]

(2.12)

Notice that the non-informative state of information was used in the above expression. Thus, no preference towards any particular model is considered in the probability distribution \( \Theta(d, m) \). The opposite would be the case if any other probability would have been used in Equation (2.12), instead of \( \mu_m(m) \).

Using Equations (2.11) and (2.12) in Equation (2.9), and subsequent marginalization in the data domain, it is straightforward to derive the following expression for the \( a \ posteriori \) probability distribution \( \sigma(m) \):

\[
\sigma(m) = \int_d \sigma(d, m) \, dd = \rho_m(m) \int_d \frac{\rho_d(d) \Theta(d|m)}{\mu_d(d)} \, dd.
\]

(2.13)

On the right hand side of the above equation the \( a \ priori \) probability distribution is given by \( \rho_m(m) \), and the likelihood function is represented by the integral term. Equation (2.13) is equivalent to Equation (1.1), when one explicitly considers modeling errors, via \( \Theta(d|m) \), and uncertainties about the data via \( \rho_d(d) \), in the formulation of the problem.

In the special case considered in this study, in which data and modeling uncertainties can be expressed by Gaussian distributions, I used the following expressions for \( \rho_d(d) \) and \( \Theta(d|m) \):

\[
\begin{align*}
\rho_d(d) &\propto \exp \left[ (d - d_{obs})^T C_{OBS}^{-1} (d - d_{obs}) \right], \\
\Theta(d|m) &\propto \exp \left[ (d - g(m))^T C_{MOD}^{-1} (d - g(m)) \right],
\end{align*}
\]

(2.14)

where \( C_{OBS} \) and \( C_{MOD} \) are the covariance matrices associated with data and modeling uncertainties, respectively. The assumption of Gaussian probability distributions is an approximation that honors the first two moments of the underlying process (in this
case, the data and modeling uncertainties). Higher-order moments are not taken into consideration in the calculation. In the following chapter, such a Gaussian hypothesis is discussed, as are the overwhelming difficulties of going beyond it.

The expression for the a posteriori probability distribution used in this dissertation is obtained by substituting the above expression into Equation (2.13). This, taking into consideration that \( \mu_d(d) \) is a constant factor, yields

\[
\sigma(m) \propto \rho_m(m) \exp \left[ (g(m) - \mathbf{d}_{\text{obs}})^T (C_{\text{MOD}} + C_{\text{OBS}})^{-1} (g(m) - \mathbf{d}_{\text{obs}}) \right].
\]  

(2.15)

In Chapters 4 and 5, I will show applications of Bayesian inversion on synthetic and field data sets. Notice, that in the synthetic calculation, the same forward modeling operator is used to generate the synthetic observed data and to do the inversion procedure. Thus, \( C_{ME} \to 0 \) in this case (i.e., there are no errors in the modeling). This is not the situation though, in the inversion of the field data set, where both modeling and data uncertainties should be taken into account via the covariance matrices \( C_{ME} \) and \( C_{DU} \), respectively. The procedure used to estimate these covariance matrices is described in detail in these chapters.

2.1.3 Discussion

The a posteriori probability distribution \( \sigma(m) \), defined in Equation (2.15), obviously depends on the uncertainties of the a priori model and of the observed data and forward modeling operator. Such uncertainties, represented by the a priori probability distribution \( \rho(m) \), and the covariance matrices \( C_{OBS} \) and \( C_{MOD} \), will clearly influence all the statements of resolution provided by \( \sigma(m) \). As will be described later, these statements are in the form of error bars on the parameter estimates, pseudorandom realizations of the subsurface and marginal distributions of the a posteriori probability.

It is important to mention that, since not all uncertainties can be considered in the inverse calculation, such resolution statements are in fact optimistic. However, this will always be the case, either due to computational cost (for example, uncertainties due to out-of-plane reflections) or simply because it is not clear how to quantify a specific source of uncertainty (for example, errors in the acquisition geometry).

Besides the likelihood function, already defined in the previous section, the other relevant component in the Bayesian inference procedure is the a priori probability distribution \( \rho(m) \). This distribution allows the incorporation of information into the inverse calculation that is independent of the observed data. Here, \( \rho(m) \) will be built from well-log measurements (sonic and density logs), via Gaussian distributions. Two parameters are needed to completely characterize a Gaussian distribution: its mean and its covariance. Here, the mean of \( \rho(m) \) is the most likely model a priori \( \mathbf{m}_{\text{prior}} \), since in the Gaussian case it is the model at which the a priori probability distribution is maximum. This model will be taken as a smooth depth profile of the quantities that are the unknowns in the inverse problem (elastic impedances and density). As
will be described in the following chapter, such information can be taken from velocity analysis, travel-time tomography or simply by smoothing the well-log measurements. The latter option will be used here.

The covariance matrix is a measure of uncertainty about \( m_{\text{prior}} \), and will also be estimated from well log measurements. This allows the incorporation of the medium correlations into the inverse calculation in a manner consistent with the well log information. A comparison of this approach with another inversion scheme, in which well-log measurements are not used, is carried on in Chapters 4 and 5.

The assumption of Gaussianity, used in this work for both likelihood and a priori probability functions, implies the hypothesis that only the first two moments (mean and covariance) of the distribution are required for its specification. Higher-order moments could, in principle, be considered in a maximum-entropy problem to build more general probability distributions than the Gaussian. However, the dimensionality of the problem studied here, coupled with the finite length of the well logs, are major hindrances to achieving this. Those issues and a brief introduction to the theory of maximum entropy are explained in Chapter 3.

A final important point to consider is the issue of using non-parametric distributions, derived from histograms, in the construction of the a posteriori probability distribution. This was the option chosen in Scales and Tarantola (1994), in which they use a priori non-parametric distributions, computed from histograms derived from well logs. Although such distributions are in principle more general than parametric ones (for example, the Gaussian distribution), they are more difficult to analyze. This is due to issues related to multimodality of these distributions, which require for Monte Carlo procedures to derive measures of resolution or of uncertainty on the parameter estimates. This is in contrast to gradient-based techniques used in this study, as will be explained in detail in the next chapter. The use of Monte Carlo procedures increases the computational cost of the Bayesian inversion proposed in this dissertation to a level that would make its application impractical to the problem addressed. This is one more motivation for choosing Gaussian distributions in the construction of the a priori probability distribution \( \rho(m) \).
Chapter 3

PRIOR PROBABILITIES

3.1 Introduction

Here I discuss the construction of the a priori probability distribution $\rho(m)$ from a suite of well log measurements, more specifically $P$-wave and $S$-wave sonic and density logs. The motivation for doing so is to incorporate information about the subsurface into the inverse calculation consistently with well logs. This avoids unsupported preconceptions, such as that the subsurface is smooth or, in other words, that the subsurface parameters are largely correlated. Instead, the degree of correlation among these parameters will be dictated by the well log measurements. The smoothness assumption is indeed an important stabilization factor in regularized inverse procedures (Parker, 1994), but, in fact, it is not generally justified by the data. Therefore, such an assumption does not fit in the Bayesian framework developed in this dissertation, since one of its main objectives is to derive the a priori distribution $\rho(m)$ from quantitative information about the subsurface, other than the seismic data.

In this study I hypothesize that $\rho(m)$ can be described by a multidimensional Gaussian probability distribution, whose parameters (the mean and the covariance matrix) are estimated from the sonic and density logs. As I show in the next sections, according to the principle of maximum entropy, the assumption of Gaussianity honors the first two moments (the mean and covariance) of the underlying process (in this case the well-log measurements). Higher-order moments associated with these measurements are not taken into consideration by the Gaussian hypothesis. Nonetheless, by building this Gaussian a priori distribution from quantitative information about the subsurface, I will not resort to assumptions such as smoothness of the medium parameters in the inverse calculation. As mentioned above, those assumptions may not be justified in a Bayesian framework.

The principle of maximum entropy, due to the work of Shannon (1948) for discrete probabilities and extended to continuous probabilities by Jaynes (1957) and (1968), has already proved its value for constructing a priori probability distributions in the one-dimensional case (e.g., Mead and Papanicolau 1984 and Moraes 1996). In a multidimensional space, however, the principle of maximum entropy presents some challenges that have yet to be overcome. Some of those will be discussed in the following sections.

The main objective of this chapter is to describe the procedure I have used to construct the a priori probability distribution from well log measurements. To better grasp the concepts involved in assigning prior probabilities, an understanding of the principle of maximum entropy is helpful. The next section presents an overview on
W.P. Gouveia, Jr.

discuss this subject, which is a summarized version of Gouveia et al. (1996a).

3.2 On Entropy

In the previous chapter I reviewed the formalism introduced in Tarantola and Valette (1982), in which it was postulated that the most general description of a state of information about a finite parameter space is given by a probability density. In fact, this probability density quantifies the information we have about such a space.

Entropy, being the negative of information, is related to our lack of understanding, or uncertainty about a system. This concept has its roots in the field of thermodynamics and statistical mechanics, and although it has the formal definition of the ratio of heat transfer to temperature of a thermodynamical system (Feynman et al., 1963), it can be interpreted as a measure of uncertainty of its behavior (Gouveia et al., 1996a). It was Shannon’s great insight (Shannon, 1948) that this quantitative measure of uncertainty could be applied generally to problems involving information, specifically in quantifying the amount of information (or entropy) of a communication channel in terms of the probabilities associated with the different symbols being transmitted.

The original motivation for studying entropy was to use the principle of maximum entropy to calculate conservative Bayesian priors from observations. “Conservative” in the sense that such prior distributions bring into the inversion calculation no more information than that given by the observations. For example, a well log contains measurements of a subsurface property acquired at \( n \) depth locations. If one regards the log as being a realization of a random process at those locations, then the joint \( n \)-dimensional probability distribution associated with this process quantifies the uncertainties associated with such measurements. In this study, this \( n \)-dimensional probability distribution is taken as an \emph{a priori} distribution in the inverse calculation. Theoretically it can be obtained by a straightforward application of the principle of maximum entropy. In fact, if we assume that the process is Gaussian, the task of estimating the parameters of this distribution according to this principle reduces to the classical problem of computing the first two moments (mean and covariance) from sampled data. However, if one wishes to get beyond the Gaussian assumption, the application of the maximum-entropy principle requires the use of higher-order-moment information. In the one-dimensional case Gouveia et al. (1996a) show examples of non-Gaussian \emph{a priori} distributions derived in this way. However, the challenges in higher dimensions are formidable.

3.2.1 Definition of entropy

Consider an experiment with \( N \) possible outcomes each occurring with a probability \( p_i \). Shannon (1948), in analogy with the original use of entropy in the field of thermodynamics, introduced the following definition of the entropy for such discrete probabilities:

\[
H [p] = - \sum_i p_i \log p_i.
\]  

(3.1)
In the above definition \( p \) is a one-dimensional distribution associated with the frequency of outcomes of the possible events. Therefore, \( p \) is not affected by the correlation of the outcomes of such events. For example consider the following two sequences

\[
0010110110, \\
0000011111,
\]

where the second one has its outcomes more correlated than the first. Such a correlation, however, is being neglected if one assumes that both sequences can be modeled by a one-dimensional probability distribution with two equally likely outcomes; 0 and 1. Therefore, according to (3.1) the entropy of both sequences is the same. To account for such correlation the multidimensional probability distribution \( p_{ij...n} \) has to be used, which entropy is

\[
H[p] = - \sum_i \sum_j ... \sum_n p_{ij...n} \log p_{ij...n}.
\]  

(3.2)

In fact, Equation (3.1) is a particular case of this definition, when the samples of the sequence are uncorrelated. Gouveia et al. (1996a) show that the entropy decreases as the correlation length of the sequence increases. Thus, (3.2) yields a larger entropy for the first sequence than for the second.

As will be shown, Bayesian multidimensional priors could in principle be computed from entropy considerations. Those distributions will be derived from subsurface in-situ measurements, which are definitely correlated. Therefore, Equation (3.2) is the one used in this study. However, given that the subsurface parameters can take on any value within a specific range, this definition has to be extended to continuous probabilities.

### 3.2.2 Extension to Continuous Probabilities

Several approaches have been proposed to extend the entropy definition introduced above to the continuous case (Jaynes 1957; Rietseh 1977; and others). Such approaches amount basically to a limiting process to convert the discrete summation in (3.1) or in (3.2) to an integral. The derivation of the continuous entropy is described in detail in Gouveia et al. (1996a), where it is shown that the entropy of a continuous random variable \( x \) can be defined as

\[
H[p(x); q(x)] = - \int_a^b p(x) \log \frac{p(x)}{q(x)} \, dx.
\]  

(3.3)

This equation actually defines what is known as relative entropy. \( q(x) \) is a normalizing function refered to as Jaynes’ invariant measure. This function is required in the continuous case to make the entropy measure \( H[] \) invariant to coordinate transformations. It cannot matter which coordinate system the probability density \( p(x) \) is defined as far as its information content defined by (3.3) is concerned. For instance,
if information is available about a parameter space \( x \) in the form of the probability distribution \( p(x) \), the same information must be available about the parameter space \( y = \log x \) in the form of the probability distribution \( \tilde{p}(y) \). \( p(x) \) and \( \tilde{p}(y) \) are related through the Jacobian of the coordinate transformation from \( x \) to \( \log x \). As it is shown in Tarantola (1987), the function \( q(x) \), by transforming to the new coordinate system \( y \) via the same Jacobian, is responsible to make the relative-entropy measure (3.3) invariant to coordinate transformations.

In an alternative interpretation of (3.3) (e.g., Kolmogorov 1956; Kullback 1959; Jaynes 1968; and Tarantola 1987), \( H[p(x); q(x)] \) defines the entropy of \( p(x) \) with respect to \( q(x) \). In other words, (3.3) quantifies how informative \( p(x) \) is with respect to a reference state of information defined by \( q(x) \). In fact, this function is referred to as the reference state of information. This interpretation is often the one used under the Bayesian framework, in which the inversion calculation is formulated as a procedure that quantifies and refines successive states of information as more knowledge about the subsurface becomes available.

### 3.2.3 The Principle of Maximum Entropy

Consider the problem of inferring a probability distribution from sample moments (e.g., mean and variances) computed from a single realization of the unknown distribution. In general there are an infinite number of probability distributions consistent with these moments. Among them, the one that maximizes the entropy (3.3) is the most conservative choice for the unknown distribution. “Conservative” in the sense that this distribution, by maximizing the entropy, is the one that contains the least amount of information about the underlying process and yet satisfies the numerical moments. Thus, we face a constrained optimization problem: find a distribution that maximizes entropy and is consistent with a set of constraints expressed in terms of the moments of this distribution. Specifically, the first constraint is to require that the probability distribution integrates to one

\[
\int_D p(x) \, dx = 1, \tag{3.4}
\]

and the remaining constraints are obtained by equating the sample moments \( \mu_k \) with the analytic central moments of \( p(x) \)

\[
\langle x^k \rangle = \int_D x^k \, p(x) \, dx = \mu_k, \quad k = 1, ..., K, \tag{3.5}
\]

where \( D \) is the domain of integration. Finding the maximum-entropy distribution that is consistent with the above set of constraints can be cast as an unconstrained optimization problem by introduction of Lagrange multipliers. This problem can be analytically solved for the maximum-entropy distribution with the result (Gouveia et
al., 1996a)
\[
p(x) = Z^{-1} q(x) \exp \left[ - \sum_{k=1}^{K} \lambda_k x^k (x) \right],
\]
(3.6)

with
\[
Z \equiv \exp(\lambda_0) = \int_{\mathbb{R}} q(x) \exp \left[ - \sum_{k=1}^{K} \lambda_k x^k (x) \right] dx,
\]
(3.7)

where \( \lambda_k \) are the Lagrange multipliers associated with the constraints, and \( Z \) is a normalization factor that assures that (3.4) is satisfied. Therefore, the determination of the maximum-entropy distribution requires the calculation of the Lagrange multipliers.

### 3.3 Computation of the Lagrange Multipliers

Several approaches have been presented for the computation of the Lagrange multipliers in a maximum-entropy calculation (e.g., Mead and Papanicolaou 1984 and Jumarie 1990). Those procedures take into consideration only one-dimensional probability densities. When multidimensional probability distributions are sought the situation becomes much more complicated and, as far as I am aware, no methodology has been proposed in the literature to fully tackle this problem.

The main difficulty with implementing such an algorithm is related to the high cost of multidimensional integration. Usually the determination of the Lagrange multipliers is done in an iterative fashion, by an optimization technique. This procedure starts with a trial set of multipliers that is systematically perturbed until the moments that are derived from the maximum-entropy probability match the (sample) moments computed from the available data. Therefore, in the multidimensional case, multidimensional integrations have to be carried out to compute the moments of the maximum-entropy probability distribution at each iteration. The dimension of the problems studied in this dissertation (ranging from 50 to 600) makes the use of this approach very difficult.

Here I propose an alternative algorithm in which multidimensional integrations are not required. In this procedure, which is an extension of the approach described in Jumarie (1990) to the multidimensional case, the constraints for the maximum-entropy calculation are \( m-th \) order multidimensional central moments computed from sample data. Specifically, these constraints are defined as:

\[
\langle x_i \rangle = \mu_i, \ m = 1, \quad \langle x_i x_j \rangle = \mu_{ij}, \ m = 2, \quad \langle x_i x_j x_k \rangle = \mu_{ijk}, \ m = 3, \quad \langle x_i x_j x_k x_l \rangle = \mu_{ijkl}, \ m = 4 \ldots
\]

\( i, j, k, l \ldots \in [1, n], \)
(3.8)

where \( x_i \) is the \( i-th \) component of the random vector \( \mathbf{x} \), \( \mu_{i...j} \) are the sample mo-
ments computed from the data, and \( n \) is the dimension of the problem. Considering these constraints, the maximum-entropy distribution defined in Equation (3.6) can be rewritten as

\[
p(x) = Z^{-1} q(x) \exp \left( -\lambda_i x_i - \lambda_{ij} x_i x_j - \lambda_{ijk} x_i x_j x_k - \ldots \right).
\] (3.9)

The proposed algorithm is based on the integration by parts of the multidimensional moment integrals

\[
\langle x_i x_j \ldots x_n \rangle = \int_V (x_i x_j \ldots x_n) p(x) \, dx.
\] (3.10)

To simplify the exposition of the algorithm, consider that the problem at hand is to find the maximum-entropy distribution \( p(x) \) that satisfies central moments up to only the second-order computed from sample data. Then, the maximum-entropy distribution is defined as:

\[
p(x) = Z^{-1} q(x) \exp \left( -\lambda_i x_i - \lambda_{ij} x_i x_j \right),
\]

\[
\langle x_i \rangle = \int_V x_i \, p(x) \, dx,
\]

\[
\langle x_i x_j \rangle = \int_V x_i x_j \, p(x) \, dx.
\] (3.11)

As an important remark, it is often mentioned in the literature that Gaussian distributions are maximum-entropy distributions when only moments up to the second-order are taken into consideration. This is just true when the reference state of information \( q(x) \) in (3.11) is defined as a uniform probability distribution. In that case \( p(x) \) is in fact a Gaussian distribution whose mean \( x_0 \) and covariance matrix \( C \) can be related to the Lagrange multipliers by (Gouveia et al., 1996a)

\[
\lambda_0 = \frac{n}{2} \log(2\pi) + \frac{1}{2} \log \det C + \frac{1}{2} x_{0i} C^{-1}_{ij} x_{0j},
\]

\[
\lambda_i = -x_{0j} C_{ij}^{-1},
\]

\[
\lambda_{ij} = \frac{1}{2} C_{ij}^{-1}.
\] (3.12)

For the second-order moment problem considered above, cumbersome but straightforward algebra shows that if we perform an integration by parts in the third expression of Equation (3.11) in the variable \( x_{i+1} \) we obtain

\[
\langle x_i \rangle = \lambda_{i+1} \langle x_i x_{i+1} \rangle + \sum_{j=1}^{n} \lambda_{i+1 j} \langle x_i x_{i+1} x_j \rangle + \lambda_{i+1 i+1} \langle x_i x_{i+1}^2 \rangle, \quad i = 1, \ldots, n-1. \] (3.13)

If we repeat the process, this time with the integration by parts being on the variable
Multidimensional Bayesian Inversion

\[ x_k, \text{ with } k \neq i \text{ and } k \neq j, \text{ we obtain:} \]

\[
\langle x_i \ x_j \rangle = \lambda_k \langle x_i \ x_j \ x_k \rangle + \sum_{l=1}^{n} \lambda_{lk} \langle x_i \ x_j \ x_k \rangle + \lambda_{kk} \langle x_i \ x_j \ x_k^2 \rangle, \quad i = 1, \ldots, n. \tag{3.14}
\]

Equations (3.13) and (3.14) form a linear system of equations that in principle could be solved for the Lagrange multipliers. Notice that no numerical integration is required in this procedure as it is in other techniques. The price we pay is the need for computation of higher moments. In this case of a second-order multidimensional moment problem, we have to compute moments up to the fourth-order from sample data.

To illustrate the algorithm proposed here, consider the following example. Figure 3.1 shows a correlated Gaussian process of unit variance, zero mean, and exponential correlation with correlation length of five samples. This figure also displays the analytic autocorrelation of the sequence and the autocorrelation computed numerically from its samples. Without loss of generality let us consider a uniform distribution for \( q(\mathbf{x}) \) and that we know the mean of the process in advance, zero in this case. We then want to compute the Lagrange multipliers \( \lambda_{ij} \) of the maximum-entropy distribution \( p(\mathbf{x}) \) defined in Equation (3.11), using the numerical moments computed from the samples (Figure 3.1b). Since the Lagrange multipliers are related to the elements of the (known) exponential covariance matrix of the underlying process via Equation (3.12), we can assess the results provided by the proposed algorithm.

Figure 3.2 shows the set of Lagrange multipliers obtained by solution of the linear system of equations defined in (3.13) and (3.14) in two different circumstances: when the calculation is carried out with the numerical moments computed from the samples, and when it is done with the analytic, thus exact, moments of the true distribution \(^1\). Clearly, this formulation is not capable of providing an accurate estimate for the Lagrange multipliers when numerical moments computed from the samples are used in the algorithm. However, when analytic expressions for all moments involved in the calculation are used, we obtain good results.

The conclusion that can be drawn from this experiment is that the proposed formulation is highly sensitive to the inaccuracy of computed moments. The inaccuracy in sample moments is well known in statistics (Priestley, 1981), and is attributed to the finite length of the input sequence. More accurate moment-estimation procedures, which perhaps require much larger data sets than the one considered in the above example, have to be used in conjunction with this technique such that Bayesian \textit{a priori} distributions can be derived from sample data via the maximum-entropy principle. Furthermore, the degree of accuracy of the estimated moments required to compute maximum-entropy distributions has yet to be investigated.

\(^1\text{Analytic expressions for high-order moments of a multidimensional Gaussian distribution are available in Miller (1964).}\)
Fig. 3.1. (a) Correlated Gaussian process. This process was generated by correlating a Gaussian multidimensional random variable with an exponential covariance, for which the correlation length is five samples. (b) The analytic correlation of this sequence and the correlation computed from its samples.
3.4 Gaussian a priori probability distribution from well logs

The fact that Gaussian distributions are maximum-entropy distributions that satisfy moments up to the second-order of the underlying process (here the well log measurements) justifies their use to incorporate prior information about the subsurface into a Bayesian inverse calculation. Going for higher-order-moment information in the construction of such a priori probabilities is, of course, of interest. However, this is a difficult achievement. As discussed before, the maximum-entropy procedure used here to accomplish this seems to be too sensitive to errors in the sample moments. Furthermore, non-entropy-based methods, just as the one investigated in Scales and Tarantola (1994), can result in multimodal a priori probability distributions which require Monte Carlo procedures in the inverse calculation, thus raising its computational cost.

Therefore, I will use a Gaussian distribution to incorporate the information on the subsurface parameters provided by the well log measurements. This distribution is given by

$$
\rho(m) = \sqrt{\frac{(2\pi)^{-\frac{M}{2}}}{\det C_M}} \exp \left[ -\frac{1}{2} (m - m_{\text{prior}})^T C_M^{-1} (m - m_{\text{prior}}) \right],
$$

where $M$ is the number of model parameters, $m_{\text{prior}}$ is the mean of this distribution and $C_M$ is its covariance matrix. The main-diagonal elements of $C_M$ represent the variance of the well-log fluctuations about $m_{\text{prior}}$, which is defined as the general depth trend of the well-log measurements. Large variance values are associated with weaker a priori information about specific subsurface parameters. Thus, for those
parameters, the likelihood function will have a larger weight in the inverse calculation than for subsurface parameters associated with smaller \textit{a priori} variances. The off-diagonal elements quantifies the extent to which the well-log fluctuations about $\mathbf{m}_{\text{prior}}$ are correlated.

Both $\mathbf{m}_{\text{prior}}$ and $C_M$ will be estimated from the well-log measurements. $\mathbf{m}_{\text{prior}}$, being some long-wavelength component of the log, is computed with a chosen running average applied to the set of measurements. As will be discussed later, $\mathbf{m}_{\text{prior}}$ is also used to incorporate the kinematics of the seismic-data reflection events, which are defined by this long-wavelength trend of the subsurface parameters, into the inverse calculation. The procedure to estimate the model covariance matrix $C_M$ is illustrated in Figure 3.3 for a $P$-wave sonic log, which is the same log used in the Bayesian inversion of field data described in Chapter 5. The procedure basically amounts to computing the autocorrelation of the fluctuations of the well log about $\mathbf{m}_{\text{prior}}$ within a sliding window, which is to account for the fact that the variance of the well log changes with depth. The double-arrow line in the figure shows the length of the autocorrelation window. Once the correlation $C(\tau)$ is computed for a window $\mathcal{W}$, the model covariance matrix is given by

$$
C[i,j] = C(j-i) = \frac{1}{N+1} \sum_{k=i}^{i+N} L(k) L(k+j-i).
$$

(3.16)

$N$ is the length of the sliding window $\mathcal{W}$ centered at depth $i\,dz$ and $L(k)$ is the well-log sample at depth $k\,dz$. $dz$ is the depth discretization interval. Figure 3.4 shows the covariance matrix generated from this process. Note that the amplitudes along the main diagonal of this matrix correlate with the variance of the fluctuations of the well log about $\mathbf{m}_{\text{prior}}$. For instance, the large fluctuations at depths around 950 m are expressed in the covariance matrix by the large variances at indexes around $i,j = 70$. Notice that, the indexes of the covariance matrix are the same as the ones of the model vector $\mathbf{m}$ used to parameterize the subsurface. Specifically, such indexes are defined as $(z_1 - z_0)/dz$, where $z_1$ is the depth of a given layer, $z_0$ is the initial depth of the target zone and $dz$ is the depth discretization level used in the inverse calculation. In the case of Figure 3.3, $z_0 = 250$ m and $dz = 10$ m. Thus, the index value of 70 for $z_1 = 950$ m.

It is important to mention that the procedure used here to estimate the parameters of the \textit{a priori} probability distribution $\rho(\mathbf{m})$ implicitly relies on the assumption of ergodicity (Priestley 1981 and Gouveia 1996). A random process is said to be ergodic when a single realization can be used, as opposed to several ones, to estimate the parameters of its underlying probability distribution. Since a single set of well-log measurements is used to compute the parameters of $\rho(\mathbf{m})$, the well-log data were assumed to be samples from an ergodic process. If more well logs were used in this procedure, the ergodicity assumption could be relaxed. This, however, was not
attempted in this study. In the next chapter I illustrate the use of these model covariances computed from a set of hypothetical well-log measurements in the Bayesian inversion of synthetic surface seismic data.
Fig. 3.3. Illustration of the technique to compute the a priori covariance matrix from well-log measurements. (a) shows a P-wave sonic log (dark line) and $m_{\text{prior}}$ (gray line), derived from that profile with a running-average procedure of length 200 m. (b) shows the fluctuations of the well log about $m_{\text{prior}}$, which are used in the computation of the model covariance matrix. The vertical double-sided arrow shows the length of the autocorrelation window. The two horizontal arrows indicate the bounds of the target depth zone, within which the covariance matrix is computed.
Fig. 3.4. Model covariance matrix computed for the $P$-wave sonic log illustrated in Figure 3.3.
W.P. Gouveia, Jr.
Chapter 4

NONLINEAR BAYESIAN INVERSION. APPLICATION TO SYNTHETIC DATA.

4.1 Introduction

This chapter is a condensed version of Gouveia (1996) and Gouveia and Scales (1996b). In it, I will describe the Bayesian inversion scheme and show its application to synthetic seismic data. These data, a shot gather generated for a 50-layer elastic model, are inverted for elastic impedances and densities, in the presence of Gaussian additive noise, band-limited to the frequency of the signal. In this calculation, hypothetical well logs are used to construct the \textit{a priori} probability distribution \( \rho(m) \).

As seen above, the solution to the inverse problem under the Bayesian framework is the \textit{a posteriori} probability \( \sigma(m) \). Several distinct approaches have been used to derive estimates of the subsurface parameters from this distribution. For instance, Mora (1987) and Crase et al. (1990) compute the maximum \textit{a posteriori} estimate by finding, via optimization techniques, the model at which \( \sigma(m) \) is maximum. Thus, they offer a single model as the solution of the inverse problem. Furthermore, they have not quantified either data or model uncertainties associated with the inverse calculation.

At another extreme, Scales and Tarantola (1994), Mosegaard and Tarantola (1995) and Mosegaard et al. (1995) resorted to Monte Carlo procedures to globally sample \( \sigma(m) \) in order to investigate the existence of models that, by being associated with significant maxima of this distribution, are considered high probability estimates of the subsurface.

The work developed here falls somewhere in between of the above approaches. In the formulation of the probability distribution \( \sigma(m) \) I have taken into consideration data uncertainties, i.e., errors (noise) in the data and in the forward modeling procedure. Uncertainties of the subsurface parameters have also been considered via an \textit{a priori} probability distribution derived from well-log measurements. However, due to the high computational cost of the seismic forward modeling operator, a global exploration of \( \sigma(m) \), which requires many forward modeling calculations, is not practical. Therefore, I use optimization techniques to find the maximum \textit{a posteriori} model \( (m_{\text{map}}) \), and then perform a local analysis of \( \sigma(m) \) in the vicinity of this model. Such an analysis is accomplished by approximating \( \sigma(m) \) with a Gaussian distribution (referred to as \( \sigma_g(m) \)) about \( m_{\text{map}} \), which is obtained by linearizing the forward modeling procedure about this model. Error bars on each of the subsurface estimates are derived from the covariance matrix of this Gaussian approximation, which is a function of the forward modeling operator and the \textit{a priori} uncertainties. Moreover, since analytic
procedures are available to sample multidimensional Gaussian probability functions (Parker, 1994), it is possible to generate a number of pseudo-random realizations by sampling \( \sigma_g(m) \). Examining the distribution of models obtained in this way, instead of only error bars, contributes to the assessment of the uncertainties of the inverse problem. As an important remark, notice that since a local optimization technique is used to estimate \( m_{\text{map}} \), this model preserves the long-wavelength depth component of the subsurface parameters and, as a consequence, the correct kinematics of the reflection events in the seismic data, which is incorporated into the calculation via \( m_{\text{prior}} \). Thus, this local uncertainty analysis is valid on a region of the model space in which subsurface models yield the correct kinematics of the reflection events.

This chapter is organized as follows. First, I introduce the inverse problem formulated within the framework of Bayesian inference. Then, I investigate the shape of the objective function used in this formulation for a simple case by displaying some of its contour plots within selected planes through the volume in parameter space. Following that I show numerical examples of the inversion of synthetic seismic data generated for a 50-layer elastic model, in which \textit{a priori} information on the subsurface was derived from synthetic well-log measurements. Assessment of the resolution of the subsurface estimates is then carried out by analyzing the Gaussian approximation to the \textit{a posteriori} probability distribution. Moreover, I investigate the sensitivity of the model \( m_{\text{map}} \) to the initial guess used in the optimization procedure, which is related to presence of local minima in the objective function used in the calculation.

Note that the inverse calculation developed here attempts to estimate the elastic parameters of the subsurface, using a model that consists of a sequence of layers of 10 m thickness. Considering that the seismic wavelength for this case is on the order of 200 m, the inversion is being done at a scale significantly below the seismic resolution. On the other hand, the \textit{a priori} distribution is built from the fine scale well-log measurements. Therefore, it can very well bring into the solutions of the inverse problem subsurface features that are not resolved by the seismic data. In other words, pseudo-random models generated by sampling the \textit{a posteriori} distribution might present features that are not required to fit the seismic data.

It is then appropriate to ask what is the most featureless (i.e., most smooth) model that is consistent with the data and with whatever other deterministic information that is available. This featurelessness is achieved via Tikhonov regularization with a discrete second-difference matrix. The resolution of subsurface-parameter estimates is implicitly determined by the length scale over which smoothing is possible while still fitting the data. This technique is called Occam inversion by Constable et al. (1987). After exploring the Bayesian inversion I perform an Occam calculation on the same data set and contrast the results at the end of this chapter.

4.2 The Bayesian Inversion Procedure

As discussed in previous chapters, in the Bayesian approach for parameter estimation, the solution of the inverse problem is the \textit{a posteriori} probability density
function $\sigma(m)$ defined on the space of models. Here, I will consider Gaussian probability densities for both the likelihood and a priori probability distributions, which are given by

$$L(m) \propto \exp \left[ -\frac{1}{2} (g(m) - d_{\text{obs}})^T C_D^{-1} (g(m) - d_{\text{obs}}) \right],$$

$$\rho(m) \propto \exp \left[ -\frac{1}{2} (m - m_{\text{prior}})^T C_M^{-1} (m - m_{\text{prior}}) \right].$$  \hspace{1cm} (4.1)

Most of the terms in the above equations were defined in the previous chapters, but for convenience those definitions will be repeated here. $g(m)$ is the forward seismic modeling operator that generates synthetic data for a given model $m$, $d_{\text{obs}}$ is the observed seismic data and $C_D$ is the covariance matrix of the noise in the data. $m_{\text{prior}}$ is the mean of $\rho(m)$, and $C_M$ is the covariance matrix describing the uncertainties in this model. The procedure founded on the Gaussian hypothesis used to construct the a priori probability distribution from well-log measurements discussed in Chapter 3, is applied to the synthetic calculation developed in this chapter.

Throughout, I will work with the assumption that the underlying model is elastic, isotropic and laterally homogeneous. The reflectivity method of Fuchs and Müller, (1971) therefore provides a suitable model operator in this scenario. The fact that this algorithm yields an analytic solution to the full elastic wave field is advantageous in the computation of the Frechét derivatives, which are the derivatives of the forward modeling operator with respect to the model parameters, as discussed in Appendix A. Moreover, the reflectivity method allows a high degree of parallelism, which is exploited in this work. Details on the parallel implementation of the Bayesian inversion are presented in Appendix B.

The first objective of the Bayesian inversion procedure is to find the model (or models) associated with significant maxima of $\sigma(m) \propto L(m) \rho(m)$, which is invariably done by some optimization technique. Note that to find the maximizers of $\sigma(m)$ is equivalent to minimizing over the space of models $m$, the objective function

$$\Theta(m) = \frac{1}{2} \left[ (g(m) - d_{\text{obs}})^T C_D^{-1} (g(m) - d_{\text{obs}}) + (m - m_{\text{prior}})^T C_M^{-1} (m - m_{\text{prior}}) \right],$$  \hspace{1cm} (4.2)

which I accomplish by a nonlinear version of the conjugate gradient method (Fletcher, 1987). This method is a local, iterative, optimization procedure that works by updating an iterate $m_n$ in the following way

$$m_{n+1} = m_n + \alpha \phi_n.$$  \hspace{1cm} (4.3)

Here, $\phi_n$ is a search direction, which is a function of the gradient of the objective function $\Theta(m)$ at iterations $n$ and $n - 1$, and $\alpha$ is the step length. This parameter is determined by a line-search procedure (Dennis and Schnabel, 1987). The gradient of
\( \Theta(m) \) is

\[
\nabla \Theta(m) = \Gamma(m) C_D^{-1} (g(m) - d_{\text{obs}}) + C_M^{-1} (m - m_{\text{prior}}).
\]

In this expression, \( \Gamma(m) \) is the Frechét derivative matrix, whose components are
defined as \( \partial g(m)/\partial m_i \). These derivatives are often obtained via the Born approxi-
mation of the displacement field. This procedure casts the computation of the Frechét
derivatives as a forward modeling problem with special boundary and initial condi-
tions (Tarantola, 1987). This approach was not taken in this work. Rather, the
Frechét derivatives are computed by analytical differentiation of the reflectivity inte-
gral. This yields a recursive algorithm similar to, but more expensive than, the forward
modeling procedure. In theory, however, this procedure is more accurate since no ap-
proximations are made in the computation of the derivatives as is done with the Born
approximation.

The optimization of the objective function (4.2) is carried out until a model iterate
fits the observed data with error no worse than the noise level in the data. This is
accomplished by defining the stopping criterion used in the optimization procedure
as the data residual normalized by the noise level in the data, which is quantified by
the data covariance matrix \( C_D \). Therefore, a model \( m \) is defined as the maximum a
posteriori model if it satisfies the following inequality

\[
(g(m) - d_{\text{obs}})^T C_D^{-1} (g(m) - d_{\text{obs}}) \leq 1,
\]

in the course of the optimization of \( \Theta(m) \).

### 4.2.1 Model parameterization

Different parameterizations are of course possible for the subsurface model \( m \). Nu-
merous studies have been done to understand the information on subsurface elastic
parameters available in seismic waveform data (Tarantola 1987; Jannane et al. 1989;
and others) and therefore choose an adequate parameterization. The following ones
were investigated in such studies:

- Lamé parameters \( \lambda \) and \( \mu \), and mass density;
- \( P \)-wave and \( S \)-wave velocities, and mass density;
- \( P \)-wave and \( S \)-wave impedances, and mass density, and
- \( P \)-wave impedance, Poisson’s ratio, and mass density.

The basic conclusion is that seismic data contain information on the long-wavelength
components of velocities and on short-wavelength of impedances. The fact that long-
wavelength components of velocities influence the traveltimes of the propagating waves
has as a direct consequence a non-quadratic objective function for the parameters de-
scribing the background velocities. Gradient-based methods provide little hope for
solving such problems, and global optimization techniques might be required. The objective function for the parameters describing the short-wavelength components of impedances (and also velocities), however, is approximately quadratic and, in principle, could be tackled by gradient-based optimization algorithms.

Another important issue is the coupling among different parameters. Ideally, in an inverse problem, the model parameters should be independently resolvable by the data. In a recent work, Debski and Tarantola (1995) studied the level of coupling among the model parameters in the waveform-based seismic inverse problem, for different choices of parameterization. Their main conclusion is that it is possible to interpret seismic amplitudes equally well with small values of velocity and large values of density, or with large values of velocity and small values of density. In other words these parameters trade off. The level of coupling is considerably lower when $P$-wave and $S$-wave impedances and density, or $P$-wave impedance, Poisson’s ratio and density are used as model parameters.

The above results and also numerical simulations done by Gouveia (1996) led to my choice for parameterization of the model space in terms of impedances and density. Note that the computational cost of the algorithm is about the same as the one when velocities and density are used as model parameters. As a final remark, since a gradient-based optimization method will be used for the minimization of the objective function that arises in the Bayesian formulation of the inverse problem, no attempt will be made to estimate the long-wavelength component of the subsurface parameters. We trust that this information is incorporated into the inverse problem via the initial model $m_{\text{prior}}$, which should yield the correct kinematics of the reflections recorded in the seismic data.

Once the minimizer of $\Theta(m)$ is computed, I carry out an analysis of the uncertainties of this model; however, before describing the procedure used in the uncertainty analysis, in the next section I study this objective function. This study provides insight into the shape of this function by displaying its contour plots for a simple inverse problem.

### 4.2.2 The Objective Function

Here we study the convexity of the objective function $\Theta(m)$, when the $a$ priori probability function $\rho(m)$ is taken to be uniform. Therefore, what is being investigated is the resolution provided by the (seismic) data, via the likelihood function $L(m)$.

For the model illustrated in Figure 4.1 (the same model was used in Pan et al. (1994)), I show contour plots of the objective function $\Theta(m)$ in Figure 4.2. Such plots are obtained by fixing all subsurface parameters but two, which are allowed to change within a pre-specified range. ($\pm 50\%$ of their true values, in this case.) In this specific two-layer problem, I plot the contours associated with the parameters of the two thin layers delimited by the arrows in Figure 4.1, which I refer to as layer 1 (the shallower one) and layer 2 in Figure 4.2.

Due to small $S$-wave velocity contrast in this model, the contour plot associated
Fig. 4.1. Elastic layered model associated with the contour plots displayed in Figure 4.2.

with the S-wave velocities is basically flat for all practical purposes. The other two, associated with P-wave velocities and density, exhibit valleys characterizing the range of models that are mapped to similar values by the objective function. Those models form essentially the null space of the likelihood function. For larger size problems it can be expected that the size of the null space will increase. The whole purpose of a priori information, encapsulated in \( \rho(m) \), is to bring additional information into the calculation, such that models that fit the data up to a pre-specified threshold can be differentiated according to some criterion. This is the motivation for incorporating a priori information into the inverse procedure.

4.2.3 Resolution

In the Bayesian calculation, all questions of resolution and uncertainties of the estimates are addressed via the posterior probability density \( \sigma(m) \). Notice that, even assuming Gaussian data errors and a priori information, for a sufficiently nonlinear forward operator \( g(m) \), one may have to resort to Monte Carlo integration or sampling methods to extract confidence sets or other measures of resolution from \( \sigma(m) \). Due to the high computational cost of generating synthetic seismic data, these sampling procedures are impractical to apply for the size of inverse calculations done in this dissertation. Therefore, I sidestep this issue by assuming that it is reasonable to make a Gaussian approximation to \( \sigma(m) \) in a neighborhood of \( m_{\text{map}} \). This is equivalent to linearizing the forward modeling operator about this model. Therefore, with this approximation, I assess the resolution and uncertainties of the parameter estimates
Fig. 4.2. Contour plots of the objective function. Those plots are obtained by changing a pair of subsurface parameters (one from layer 1 and the other from layer 2) systematically and keeping the other constants at their correct values. The range within which a parameter is allowed to vary is defined by ± 50% of its true value.
with a local analysis of the \textit{a posteriori} distribution about \(m_{\text{map}}\) instead of a global analysis of this function, as it is aimed by Monte Carlo procedures. Such a Gaussian approximation to \(\sigma(m)\) is defined as

\[ \sigma_g(m) \propto \exp \left( (m - m_{\text{map}})^T C^{-1}_{M'} (m - m_{\text{map}}) \right), \]  

(4.6)

where \(C_{M'}\) is the \textit{a posteriori} covariance matrix which is given by (Tarantola, 1987)

\[ C_{M'} = \left[ G^H C_D^{-1} G + C_M^{-1} \right]^{-1}, \]  

(4.7)

and \(G\) is the linearization of \(g(m)\) about \(m_{\text{map}}\). In the numerical examples I will extract information from the above equations in two different ways: first, via the square-roots of the main-diagonal elements of \(C_{M'}\), which are the unit standard-deviations error bars on the MAP model parameters. By examining only these error bars as a measure of resolution, however, one is left with a limited view of the uncertainties in the calculation, since they ignore correlations among the parameters. Therefore, I also generate a tour of pseudo-randomly simulated models sampled according to Equation (4.7) and \(m_{\text{map}}\). Along with the error bars, study of the distribution of models sampled in this way provides a more comprehensive picture of which features in the model are well resolved and which are not.

Notice that since the prior covariance matrix \(C_M\) appears in Equation (4.7), it is clear that the information extracted from the fine-scale features in the well-log measurements influence the resolution, to a greater or lesser extent depending on the relative weights associated with \(G^H C_D^{-1} G\) and \(C_M^{-1}\). When the well logs are not used in the calculation, the covariance matrix of the estimates shows a larger correlation among parameters (hence lower resolution), as will be discussed in the comparison between the Bayesian and Occam approaches for seismic waveform inversion.

### 4.2.4 Numerical Calculations

We now consider the problem of estimating \(P\)- and \(S\)-wave impedance and density profiles from a synthetic surface-seismic data set (Figure 4.3), generated for the elastic velocity model illustrated in Figure 4.4. The seismic data were corrupted by Gaussian noise, band limited to the bandwidth of the data. The well logs illustrated in Figure 4.4 were obtained by sampling a Gaussian distribution with an exponential covariance having a correlation length of five samples. An exponential covariance is defined such that the correlation of the underlying process decreases exponentially with the lag. Specifically, at the lag equal to the correlation length the correlation is reduce to \(\exp^{-1}\) times the value of the correlation at lag zero. This figure shows only the target zone for the inversion, which consists of 50 layers each of 10-m thickness. The data-misfit window, i.e. the time window over which the observed and synthetic data are compared, used in the inversion calculation is 1.5 s long and starts at 0.5 s.

Estimated model covariance matrices, one for \(P\)-wave impedance, one for \(S\)-wave
Fig. 4.3. Vertical component of the displacements generated for the elastic model of Figure 4.4. (a) Noise free data. (b) Data corrupted by Gaussian noise band-limited to the frequency bandwidth of the signal.

Fig. 4.4. 50-layer elastic model. The medium for depths less than 1 km is homogeneous with \( P \)-wave velocity, \( S \)-wave velocity and density equal to 3 km/s, 2 km/s and 1 g/cm\(^3\), respectively.
impedance and one for density are illustrated in Figure 4.5. As mentioned, those matrices define the uncertainties in the model \( m_{\text{prior}} \), and are used in the Gaussian a priori distribution \( \rho(m) \).

The departures from an exponential covariance matrix, manifested as side bands in the matrices shown in Figure 4.5, are artifacts due to the finite length of the well log. Such errors in the estimation of moments from sample data have also been illustrated by Figure 3.1 in Chapter 3 for a Gaussian process.

Since in this case the well logs in Figure 4.4 are stationary processes, the window over which the autocorrelation was computed to estimate the covariance matrix, as described in Chapter 3, is the entire well log. \( m_{\text{prior}} \) was obtained from the true model, smoothed with a uniform operator having a length of 100 m. It is important to note that the data shown in Figure 4.3 were generated with the same forward modeling operator used in the inversion procedure. Therefore, the only errors present in the data are the artificial Gaussian noise; no modeling errors exist in the following calculations.

Similar to that for the estimated model covariance matrix, I construct the data covariance matrix \( C_D \) by computing the autocorrelation of a single seismic trace containing only noise samples. The noise seismic trace and the covariance matrix \( C_D \) are shown in Figure 4.6.

**Results:** Figure 4.7 shows the MAP model, computed with 12 nonlinear conjugate gradient iterations, along with the true subsurface parameters and the model \( m_{\text{prior}} \). This model fits the observed data \( d_{\text{obs}} \) up to the noise level in the data, since it satisfies the stopping criterion for the optimization (4.5). The error bars in the figure are the square-roots of the main diagonal of the a posteriori covariance matrix (4.7). The following conclusions can be drawn from this experiment.

- The error bars include the true model, for most of the depth range within which the inversion was attempted. Notice that since the MAP model fits the observed data in a average sense, it is not guaranteed that all true subsurface parameters will lie within the error bars.

- The P-wave impedance error bars are the smallest, reflecting the fact that P-wave impedance is the best resolved parameter in this experiment.

- The density error bars are the largest, indicative of the poor resolution of density from seismic amplitudes. Moreover, since the initial guess for the density profile is also within the error bars, it is as good a model as \( m_{\text{map}} \), given the data provided for the inversion. Therefore, the only information on density obtained from the inverse calculation was the error bars, representing a one standard-deviation confidence interval for the true density.

Recall that no linearization was done in the estimation of \( m_{\text{map}} \), since the optimization was performed with a nonlinear conjugated gradient equipped with a line
Fig. 4.5. Estimated model covariance matrices for $P$-wave impedance (a), $S$-wave impedance (b) and density (c). The impedances were obtained by multiplying the velocity and density profiles displayed in Figure 4.4.
W.P. Gouveia, Jr.

Fig. 4.6. Noise seismogram (a) and the data covariance matrix (b). For visualization purposes the data covariance matrix is plotted just for the first 0.5 s of the data-misfit window.

search. In Appendix C, I discuss the results obtained by linearizing the inverse problem about \( \mathbf{m}_{\text{prior}} \). This yields a linear system of equations that is solved iteratively by a conjugate-gradient technique. As will be shown in the appendix, the linearization errors are not negligible, what result in a poor MAP estimate of the subsurface when compared to the one computed with the nonlinear optimization.

**On the resolution of the estimates:** Let us now consider the resolution provided by the Bayesian technique by examining the *a posteriori* covariance matrix \( (4.7) \) about \( \mathbf{m}_{\text{map}} \). Also, we will sample the Gaussian approximation of the *a posteriori* probability distribution \( \sigma_g(\mathbf{m}) \), Equation (4.6), thus generating pseudo-random subsurface models that are consistent with the *a priori* information and observed data. The resolution of the forward modeling operator \( g(\mathbf{m}) \) itself can be assessed with a SVD analysis. This is done in Appendix D.

The *a posteriori* covariance matrix is displayed in Figure 4.8. Direct inversion of the quantity \( G^H C_D^{-1} G + C_M^{-1} \) can be numerically unstable due the presence of small eigenvalues (Gouveia, 1996). What Figure 4.8 actually shows is the pseudo-inverse of \( G^H C_D^{-1} G + C_M^{-1} \), in which the smallest (normalized) eigenvalue allowed to enter in the calculation was 0.001. This matrix can be understood as a 3x3 block matrix. The main diagonal blocks are the autocorrelations of the *P*-wave impedance, *S*-wave impedance and density. The off-diagonal blocks represent cross-correlations among different parameters.
Fig. 4.7. The MAP model estimated from the data set shown in Figure 4.3. The true model and the initial guess used for the nonlinear optimization calculation are also shown.
Fig. 4.8. A posteriori covariance matrix for the problem of elastic waveform inversion. The labels P, S and \( \rho \) indicate \( P \)-wave, \( S \)-wave impedance and density, respectively.
The pseudo-random subsurface models sampled from $\sigma_g(m)$ are displayed in Figure 4.9. The sampling was done by taking the inner product of the lower triangular part of the $LU$ decomposition of $C_M$ with an uncorrelated Gaussian pseudo-random sequence (of zero mean and unit variance) and adding the result to $m_{\text{map}}$ (Parker, 1994). As expected the $P$-wave impedance shows the smaller variation. The density realizations show the largest variation, emphasizing that little information on density is available from seismic amplitudes. The $S$-wave impedance realizations are somewhere in between. Another way to display the pseudo-random models is shown in Figure 4.10. In this figure the realizations are shown side by side. Thus, larger lateral continuity implies a higher degree of confidence in the estimates of the subsurface parameter at a particular depth. Figure 4.10 illustrates one of the main purposes of the Bayesian inversion approach: to generate realizations of the subsurface that are consistent with the observed data and a priori information. The same sampling procedure can be applied to the a priori probability distribution $\rho(m)$. Such a priori realizations are shown in Figure 4.11. Any improvement in lateral coherence seen in Figure 4.10 when compared to Figure 4.11 is due to the information provided by the data, which gives rise to a narrowing of the a posteriori distribution relative to the a priori one. Notice that all models in Figure 4.11 show a long-wavelength depth component in common, which is given by $m_{\text{prior}}$. 

Multidimensional Bayesian Inversion

41
Fig. 4.9. Pseudo-random realizations of the \textit{a posteriori} probability distribution. Shown are \textit{P}-wave impedance (top), \textit{S}-wave impedance (middle) and density (bottom). Note that all realizations are shown in a blocked fashion.
Fig. 4.10. Random realizations of the *a posteriori* probability distribution. Lateral continuity implies on a high confidence of the estimates.
Fig. 4.11. Random realizations of the *a priori* probability distribution. No strong lateral continuity is seen on these plots.
Sensitivity of the MAP estimate: The gradient-based optimization procedure used in the calculation of the MAP model will converge to the local minimum closest to the initial guess. As has already been discussed, gradient methods are adequate for the problem of waveform data inversion given that the initial guess for the optimization problem provides the right kinematics of the reflection events. Therefore, such an initial guess has to incorporate an accurate representation of the long-wavelength component of the subsurface parameters.

Let’s now illustrate via synthetic experiments the sensitivity of the MAP model to the initial guess used in the optimization of the objective function (4.2). Two cases will be considered. In the first, the initial guess for the optimization is smoother than the one used in the MAP model estimation performed in the last section. In the second, for the overburden, I introduce an error of ±10% while for the target depth interval, I use the same initial guess as that in the previous section.

Figure 4.12 compares the initial guess used in this sensitivity analysis, obtained with a running average of 200 m of the well-log data, with the one used previously. Optimization starting with the less-detailed initial guess yields the MAP model shown by Figure 4.13. This model, although generally having larger errors relative to the true subsurface than did the MAP model obtained in the last section, still fits the data up to one standard deviation of the noise, but at the expense of a higher computational cost (roughly 50% higher). More conjugate gradient iterations are required to optimize the objective function (4.2), using a poorer model as the initial guess.

The MAP model is definitely sensitive to errors in the overburden, as is clearly seen in Figures 4.14 and 4.15. A shift of ±10% resulted in virtually no movement by the optimization process from the initial guess to the MAP solution, due to the small magnitude of the gradient vector at the initial model. Therefore, the optimization procedure is not able to provide a model that fits the data in accordance with the stopping criterion (4.5). In fact, the stopping criterion used in this calculation was that the module of the gradient of the objective function was smaller than a pre-specified threshold. This illustrates the presence of local minima in the objective function.

Such a small value for the gradient module is easily understood if one considers the expression for the gradient (4.4). Note that the term $m - m_{\text{prior}}$ is a zero vector at the initial guess ($m_{\text{prior}}$), and that the term $\Gamma(m_{\text{prior}}) C_D^{-1}(g(m_{\text{prior}}) - d_{\text{obs}})$ is the cross-correlation between the Frechét derivatives and the data residual at the first iteration. The error in the overburden parameters results in little correlation between those quantities. Notice that since $m_{\text{prior}}$ is a smooth model, the synthetic data $g(m_{\text{prior}})$ do not show much reflected energy. Therefore, such correlation is in fact mainly between the Frechét derivatives and the seismic data $d_{\text{obs}}$. This lack of correlation is thus expected since the Frechét derivatives will be shifted relative to the events of the seismic data. As a result, the elements of the gradient vector will have a small magnitude.

In the next section, I apply the Occam procedure to the inverse problem just described using the same model parameterization in terms of impedances and densit-
Fig. 4.12. Initial guesses used in the sensitivity analysis. The dashed line shows the initial guess used in the previous section. The thin line shows a smoother initial guess obtained with a 200-m running average. The true model is also shown.
Fig. 4.13. MAP model estimated from the data set shown in Figure 4.3 using the initial guesses shown in Figure 4.12. The true model is also shown.
ies, and compare the results with those obtained from the Bayesian procedure. This comparison will address not only the subsurface-parameter estimates but also the uncertainties associated with both inverse calculations.

### 4.3 Occam Inversion

The Occam strategy is to aim at a conservative interpretation of the seismic data by finding the smoothest model that satisfies a pre-specified data-fitting criterion. To do so, let us accept Equation (4.5) as being a useful measure of data fit irrespective of any Bayesian interpretation of the models. The Occam strategy is to find the smoothest model within the set that satisfies (4.5). One way to accomplish this is via the following constrained optimization problem:

$$\min_m \| Rm \|
\text{subject to : } (g(m) - d_{obs})^T C_D^{-1} (g(m) - d_{obs}) \leq 1,$$

(4.8)

where $R$ is a roughening operator, in this case a discrete second-order difference operator. Practically, this can be implemented as a weighted least-squares problem with a Lagrange multiplier (or regularization parameter) $\lambda$ to control the trade-off between model smoothness and data fit. Therefore, the following weighted regularized least-squares problem

$$\min_m \Gamma(m) = \frac{1}{2} \left[ (g(m) - d_{obs})^T C_D^{-1} (g(m) - d_{obs}) + \lambda (m - m_{prior})^T R^T R (m - m_{prior}) \right],$$

(4.9)

is solved for increasing values of $\lambda$ until Equation (4.5) can no longer be achieved. Possible forms for the roughening operator $R$ are

$$R = \begin{bmatrix}
-2 & 1 & 0 & 0 & \ldots & 0 & 0 \\
1 & -2 & 1 & 0 & \ldots & 0 & 0 \\
0 & 1 & -2 & -1 & \ldots & 0 & 0 \\
& & & & \vdots & & \\
0 & 0 & 0 & 0 & \ldots & 1 & -2 \\
\end{bmatrix},$$

(4.10)

and

$$R = \begin{bmatrix}
0 & 0 & 0 & 0 & \ldots & 0 & 0 \\
1 & -2 & 1 & 0 & \ldots & 0 & 0 \\
0 & 1 & -2 & 1 & \ldots & 0 & 0 \\
& & & & \vdots & & \\
0 & 0 & 0 & \ldots & 1 & -2 & 1 \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
\end{bmatrix}. $$

(4.11)
Fig. 4.14. MAP model estimated from the data set shown in Figure 4.3 using the initial guess obtained with the 100-m running average of the true log data, but with an error of +10% in the parameters of the overburden.
Fig. 4.15. The MAP model estimated from the data set shown in Figure 4.3 using the initial guess obtained with the 100-m running average of the true log data, but with an error of -10% in the parameters of the overburden.
Despite their similarity, there is an important difference between (4.10) and (4.11). While they both smooth the solution, (4.11) maps vectors whose elements are constant or linearly varying (with element index) into zero vectors, and therefore has the same null space as does the continuous second derivative operator. Thus, (4.11) will make (4.9) regular if and only if its null space does not overlap the null space of the forward modeling operator. This is easily seen by considering a model perturbation $\delta \mathbf{m}$ that happens to be in both the null space of $g(\mathbf{m})$ and the null space of $R$. Indeed,

$$
\begin{align*}
    g(\mathbf{m} + \delta \mathbf{m}) &= g(\mathbf{m}) \\
    R\delta \mathbf{m} &= 0
\end{align*}
\Rightarrow \Gamma(\mathbf{m} + \delta \mathbf{m}) = \Gamma(\mathbf{m}),
$$

what can be verified by direct substitution of $\mathbf{m} = \mathbf{m} + \delta \mathbf{m}$ in Equation (4.9). On the other hand, operator (4.10) is not singular, and will regularize (4.9), regardless the null space of $g(\mathbf{m})$. This is the reason why I use (4.10) in the Occam calculation.

This regularized nonlinear least-squares problem can be solved using the same optimization algorithm used in the Bayesian $\mathbf{m}_{\text{map}}$ calculation. Thus, it is simply a matter of repeating the calculation until the optimal regularization parameter $\lambda$ is found. The model $\mathbf{m}_{\text{occam}}$ associated with the largest value of $\lambda$ that still fits the data is the solution to this problem.

### 4.3.1 Resolution

In the Occam procedure the objective is to find the broadest average of the Earth structure that is consistent with the data. That way one can be confident that the features that are in the model are required to fit the data. In this procedure the resulting degree of smoothness is one measure of resolution. An alternative measure is obtained by propagating the data uncertainties, represented by $C_D$, into the model space. That provides a covariance matrix that quantifies the uncertainties of the estimate of the subsurface parameters taking into consideration not only the uncertainties in the data, but also the forward modeling and regularization operators. To do so, as in the Bayesian case, the assumption of linear forward modeling is required. This is done by linearizing $g$ about $\mathbf{m}_{\text{occam}}$ in Equation (4.9), what results in the following quadratic objective function for model perturbations about this model

$$
\Gamma_g(\delta \mathbf{m}) = \frac{1}{2} \left[(G \delta \mathbf{m} - \delta \mathbf{d})^T C_D^{-1} (G \delta \mathbf{m} - \delta \mathbf{d}) + \lambda \delta \mathbf{m}^T R \delta \mathbf{m} \right],
$$

where $\delta \mathbf{m}$ and $\delta \mathbf{d}$ are the model and data perturbations given by

$$
\begin{align*}
    \delta \mathbf{m} &= (\mathbf{m} - \mathbf{m}_{\text{occam}}), \\
    \delta \mathbf{d} &= (\mathbf{d}_{\text{obs}} - g(\mathbf{m}_{\text{occam}})).
\end{align*}
$$

51
The minimizer of this objective function $\Gamma_q(\delta \mathbf{m})$ is obtained by setting its derivatives with respect to $\delta \mathbf{m}$ equal to zero. This yields the following weighted, regularized linear system of equations,

$$\tilde{\mathbf{A}} \delta \mathbf{m} = \tilde{\mathbf{d}},$$

(4.15)

where,

$$\tilde{\mathbf{A}} = \begin{bmatrix} \frac{C^{-\frac{1}{2}}}{\sqrt{\lambda}} \mathbf{G} \\ 0 \end{bmatrix},$$

$$\tilde{\mathbf{d}} = \begin{bmatrix} \frac{C^{-\frac{1}{2}}}{\sqrt{\lambda}} \delta \mathbf{d} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \frac{C^{-\frac{1}{2}}}{\sqrt{\lambda}} (\mathbf{d}_{\text{obs}} - g(\mathbf{m}_{\text{occam}})) \\ \mathbf{0} \end{bmatrix}.$$  

(4.16)

The uncertainties of the estimate $\mathbf{m}_{\text{occam}}$ are given by the covariance matrix about $\delta \mathbf{m}$, which is the estimate obtained from Equation (4.15). Thus,

$$E \left[ \delta \hat{\mathbf{m}} \delta \hat{\mathbf{m}}^T \right] = E \left[ \tilde{\mathbf{A}}^T \tilde{\mathbf{d}} \left( \tilde{\mathbf{A}}^T \tilde{\mathbf{d}} \right)^T \right]$$

$$= \tilde{\mathbf{A}}^T E \left[ \tilde{\mathbf{d}} \tilde{\mathbf{d}}^T \right] \tilde{\mathbf{A}}$$

$$= \tilde{\mathbf{A}}^T C^{-\frac{1}{2}} \left( \mathbf{d}_{\text{obs}} - g(\mathbf{m}_{\text{occam}}) \right) C^{-\frac{1}{2}} \tilde{\mathbf{A}}^T,$$

(4.17)

where, $\tilde{\mathbf{A}}^T$ is the pseudo-inverse of $\tilde{\mathbf{A}}$, $\tilde{\mathbf{A}}^T$ is $\tilde{\mathbf{A}}^T$ without the last $m$ columns, and $m$ is the dimension of the model space. Since $G$ is obtained by linearizing $g(\mathbf{m})$ about $\mathbf{m}_{\text{occam}}$, $E \left[ \delta \mathbf{d} \delta \mathbf{d}^T \right] \approx C_D$, neglecting linearization errors. Therefore Equation (4.17) yields

$$E \left[ \delta \hat{\mathbf{m}} \delta \hat{\mathbf{m}}^T \right] \approx \tilde{\mathbf{A}}^T \tilde{\mathbf{A}}^T.$$  

(4.18)

4.3.2 Numerical calculations

The data set shown in Figure 4.6 were inverted for elastic impedances and densities using the Occam strategy. The initial guess used in the optimization problem (4.9) is the model $\mathbf{m}_{\text{prior}}$ and it is the same as that in the Bayesian calculation shown in Figure 4.7. As mentioned, this objective function will be minimized for increasingly larger values of $\lambda$ until the data can no longer fit. As in the Bayesian procedure, no linearization is performed in the following calculation. Appendix C shows the results provided by the Occam procedure when the forward modeling operator is linearized about $\mathbf{m}_{\text{prior}}$. Again, this appendix shows the poor Occam estimate of the subsurface resulted from such a linearization.
Multidimensional Bayesian Inversion

Results: The model associated with the largest regularization parameter $\lambda$ that still fits the data is shown in Figure 4.16. This figure also displays the true model and the initial guess. Visual comparison between this result and the one obtained from the Bayesian calculation, Figure 4.7, shows that $m_{\text{occam}}$ and $m_{\text{map}}$ are quite similar. This is more clearly shown in Figure (4.17), where both models are plotted together. However as discussed next, the uncertainties associated with these models are rather different.

Resolution of the estimates: The covariance matrix derived in Equation (4.17) can be thought of as a depth-dependent measure of resolution. This matrix, which is a function of the regularization parameter (hence of the degree of smoothness of the Occam model), as well as of the regularization matrix $R$ and the modeling operator, is shown in Figure 4.18. The greater width of its main diagonal band, compared with that in Bayesian posterior covariance (Figure 4.8), implies a larger correlation length or smoothness. Further, the relative resolution of the different parameters can be judged from the relative width of the main diagonal bands of the different blocks. In this particular example, $P$-wave impedances have the best resolution in general, followed by $S$-wave impedances and then densities. This difference in resolution among model parameters is in agreement with the Bayesian uncertainty analysis done in the previous section.

4.4 Bayes versus Occam

We have seen a synthetic example of a reflection seismic inverse problem carried out from both a Bayesian and an Occamist point of view. On the one hand, we have a Gaussian $a$ priori distribution of layered Earth models computed by estimating a covariance matrix from hypothetical well-log measurements. These measurements are performed at a resolution finer than the seismic wavelength and therefore contain information about features not resolved by the seismic data alone. With this information we compute the maximum $a$ posteriori model via a nonlinear optimization procedure. Making a Gaussian approximation to the posterior distribution about this MAP model allows one to compute the $a$ posteriori covariance, from which it is possible to derive both unit standard-deviation error bars and a tour of typical models pseudo-randomly sampled from this posterior.

For the Occam calculation, we optimize the nonlinear least-squares data misfit function, regularized with a discrete second-difference operator, for a sequence of increasingly large values of the smoothness. This gives the smoothest model that fits the data. While the length scale of the smoothing is a measure of resolution, we can derive a covariance matrix from the augmented linear system that gives more specific information on resolution. The width of the main-diagonal band of this covariance matrix is a depth-dependent measure of the resolution provided by the data alone.

The result is that, for the problem studied here, the Bayesian MAP and Occam models are fairly similar, but the uncertainties associated with the inverse calculation
Fig. 4.16. The Occam model estimated from the data in Figure 4.3 using a non-linear, regularized optimization procedure. The true model and the initial guess are also displayed.
Fig. 4.17. Comparison of the models obtained from Bayesian and Occam inversions. Note that the models are displayed in a non-blocked way to facilitate the comparison.
Fig. 4.18. Covariance matrix obtained for the Occam inversion. The labels P, S and \( \rho \) indicate P-wave, S-wave impedance and density, respectively.
can be quite distinct. Models pseudo-randomly simulated from the Bayesian *a posteriori* distribution have features that are not required to fit the data since they are influenced by the data-independent prior information. In contrast, by construction, the Occam inversion does not generate an Earth model *per se*, but rather an estimate of the coarsest average of the Earth that is consistent with the data. Since the *a priori* information used in the Bayesian approach can be of higher resolution than what the data itself can provide, as is the case for the well log, its incorporation into the calculation increases the resolution of the Bayesian estimates, and affords an illustration of the differences between this and the Occamist points of view.

It is important to remark that the conclusions drawn here cannot be thought of as general since the extent that the well log will influence the subsurface-parameter estimates depends on the relative significance of $G^TC_D^{-1}G$ and $C_M^{-1}$ in the formulation of the *a posteriori* distribution $\sigma(\mathbf{m})$. Therefore, results might be different for different levels of data and model uncertainties, which are quantified by the covariance matrices $C_D$ and $C_M$.

In the next chapter, I discuss the application of the Bayesian inversion approach just described to a field surface seismic data set. In the calculation for that application, the *a priori* data uncertainties will be restricted not only to the errors in the observed data, but will also incorporate modeling uncertainties.
W.P. Gouveia, Jr.
Chapter 5

NONLINEAR BAYESIAN INVERSION. APPLICATION TO FIELD DATA.

5.1 Introduction

Now, let’s turn to application of the Bayesian inversion procedure to a field surface seismic data set. These data, provided by Prof. Thomas Davis, of the CSM Reservoir Characterization Project, is a small subset of the vertical-component data extracted from a nine-component survey (RCP Phase V, 1995) acquired at the Sorrento Basin (near the Las Animas Arch, Southeast Colorado). Although the inversion software developed as part of this dissertation is able to handle transverse components of the displacement field, the shear data were judged too noisy to be used for the testing of this methodology.

The Sorrento Field reservoir rock consists of Morrowan fluvial sediments, which filled incised valleys during a series of pulses of rising sea level. The geology is stratified down to the reservoir level with subtle structural features. The assumptions of lateral homogeneity of the subsurface which I make here, are consistent with such a geologic scenario.

This chapter is organized as follows. In the next section, I briefly describe the regional structural setting and stratigraphy of the Sorrento field. Following that I present the surface seismic data and well logs used in the inversion for elastic impedances and density. All components of the inversion, such as data and model covariances, as well as the a priori models, are presented next. Then I show the MAP estimates and associated uncertainties resulting from the calculation. Finally, I conclude this chapter with discussion of the results.

5.2 Regional Geology

Sorrento Field is located at the eastern flank of the Denver Basin on the northwestern plunge of the Las Animas Arch. Reservoir rocks in this field are fluvial sandstones of Morrowan age that filled incised valleys, in a sequence of sea-level rises. According to the geologic model developed in Mark (1995), these fluvial sandstones represent stacked sequences resulting from a series of sea-level rises and falls, that alternately deposited sediments and partially removed them by erosion. Mississipian carbonates underly the Morrow interval, and the overlying Pennsylvanian and Permian sections are made up of thick zones of interbedded carbonates, shales and evaporites.

Both compressional seismic data and regional gravity and magnetic data show that northwesterly and northeasterly faults dominate the structural configuration of
W.P. Gouveia, Jr.

Sorrento Field today (RCP Phase V, 1995). Structure features are very subtle, however, and little geological dip is present down to the reservoir level.

It is generally acknowledged that the Morrowan reservoirs are poorly detected by compressional waves, due to low \(P\)-wave impedance contrasts (Rampton, 1995). Ideally, the inversion performed in this work should consider the \(S\)-wave as well as \(P\)-wave pre-stack data. The \(S\)-wave data set, however, is of poor quality with little reflection energy, which is the reason why the inversion described in the following sections was applied to the compressional data only.

5.3 The Sorrento Data: Surface Seismic Data and Well Logs

The intended source and receiver locations for the Sorrento data acquisition are shown in Figure 5.1. The black dots represent vibrator-source locations and empty dots represent the receiver location. The circle in the figure encompasses the five shot gathers used in the calculation, for which lines one to eight are active. Table 5.1 summarizes the acquisition parameters used in the survey, which are completely described in RCP (1995).

Among the eight lines recorded for each shot, five were chosen on the basis of their relatively good signal-to-noise ratio. Those are shown in Figures 5.2 through 5.6. These gathers contain the data used in the inversion calculation described in the next section. Field and residual statics correction along with band pass filtering were the only processing steps applied to the data, which was re-sampled to 4-ms sampling interval. The selected shot gathers surround a well denoted as MULL#14 (indicated by an arrow in Figure 5.1), in which \(P\)-wave and \(S\)-wave dipole sonic and density cased-hole logs have been acquired. As described in Chapter 3, I used such well logs, which are displayed in their raw form in Figure 5.7, to construct the \textit{a priori} model covariance matrices. These well logs went through a couple of pre-processing steps prior to being used in the estimation of the model covariance matrices. First, notice that the well logs do not provide information on the near surface (shallower than 106 m for the \(P\)-wave sonic, 765 m for the \(S\)-wave sonic and 110 m for the density log). Therefore, a linear extrapolation of the \(P\)-wave velocity and a constant density of 2.36 g/cm\(^3\) were used in the missing interval. For the \(S\)-wave velocity the approach taken was based on deriving the missing information from the correlation between the \(P\)-wave velocity and the \(S\)-wave velocity logs, for an interval .3-km thick, beginning at the depth of .8 km. A cubic polynomial, relating \(P\)- and \(S\)-wave velocities was derived from this correlation and used to compute the \(S\)-wave velocities for the shallow layers.

After interpolating the data for the near surface, I used a 25-sample (equivalent to 7.6 m) running median filter to eliminate spikes in the log measurements. The well logs were then blocked, with a weighted-average procedure (Leaney and Ulrych, 1987), such that within the target interval for the inversion (chosen to be from 0.25 km to 1.25 km) the layers are 10-m thick\(^1\). The final well logs are shown in Figure 5.8.

\(^1\)A regularly sampled depth interval is necessary to estimate the model covariance matrices.
Receicer location

● Source location

Fig. 5.1. Intended source and receiver positions for the 3D Sorrento Survey. The shot gathers within the circle will be denoted shots 1 to 5, starting from the leftmost one. The arrow indicates the location of the well MULL #14.
<table>
<thead>
<tr>
<th>Survey layout</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Receiver interval</td>
<td>67 m</td>
</tr>
<tr>
<td>Receiver line spacing</td>
<td>335 m</td>
</tr>
<tr>
<td>Receiver line orientation</td>
<td>North-South</td>
</tr>
<tr>
<td>Number of receiver lines</td>
<td>13</td>
</tr>
<tr>
<td>Source interval</td>
<td>67 m</td>
</tr>
<tr>
<td>Source line spacing</td>
<td>268 m</td>
</tr>
<tr>
<td>Receiver line orientation</td>
<td>East-West</td>
</tr>
<tr>
<td>Number of source lines</td>
<td>13</td>
</tr>
<tr>
<td>Recording</td>
<td></td>
</tr>
<tr>
<td>Number of active lines</td>
<td>8</td>
</tr>
<tr>
<td>Active receiver groups per line</td>
<td>42</td>
</tr>
<tr>
<td>Sample rate</td>
<td>2 ms</td>
</tr>
<tr>
<td>Number of swaths</td>
<td>12</td>
</tr>
<tr>
<td>Compressional source</td>
<td></td>
</tr>
<tr>
<td>Number of vibrator units</td>
<td>2</td>
</tr>
<tr>
<td>Number of sweeps</td>
<td>8</td>
</tr>
<tr>
<td>Sweep length</td>
<td>12 s + 3 s listen</td>
</tr>
<tr>
<td>Sweep frequencies</td>
<td>8-142 Hz, nonlinear mode</td>
</tr>
<tr>
<td>3-Component receivers</td>
<td></td>
</tr>
<tr>
<td>Array</td>
<td>3 receivers inline</td>
</tr>
<tr>
<td>Separation</td>
<td>0.9 m</td>
</tr>
</tbody>
</table>

Table 5.1. Sorrento data-acquisition parameters.
Fig. 5.2. Shot record 1 recorded by groups on lines 1, 2, 4, 5 and 6.

Fig. 5.3. Shot record 2 recorded by groups on lines 1, 2, 4, 5 and 6.
Fig. 5.4. Shot record 3 recorded by groups on lines 1, 2, 4, 5 and 6.

Fig. 5.5. Shot record 4 recorded by groups on lines 2, 3, 5, 6 and 7.
Next, I discuss the calculation of the data and model covariance matrices as well as the \textit{a priori} model $m_{\text{prior}}$, computed from both the seismic data and the well logs presented in this section.

\textbf{5.4 \textit{A priori} Uncertainties}

As in the synthetic calculation described in the previous chapter, the estimation of the MAP model is accomplished by reducing the value of the objective function (4.2) until the stopping criterion (4.5) is satisfied. Both equations are rewritten here for convenience.

\[
\Theta(m) = \frac{1}{2} \left[ (g(m) - d_{\text{obs}})^T C_D^{-1} (g(m) - d_{\text{obs}}) + (m - m_{\text{prior}})^T C_M^{-1} (m - m_{\text{prior}}) \right],
\]

\[
(g(m) - d_{\text{obs}})^T C_D^{-1} (g(m) - d_{\text{obs}}) \leq 1.
\]

The model $m_{\text{prior}}$, obtained from a running average of length 200 m applied to the well-log measurements, is shown in Figure 5.9. This degree of averaging was considered adequate for capturing the general trend of impedances and density as a function of depth. The construction of the data covariance matrix $C_D$ and the \textit{a priori} model covariance matrix $C_M$ is described below.
Fig. 5.7. *P*-wave and *S*-wave dipole sonics and density cased-logs acquired at the well MULL#14.
Fig. 5.8. Well logs of Figure 5.7 after interpolation, median filtering and blocking. The target depth interval for the inversion is 1-km thick and goes from 0.25 km to 1.25 km. The dashed line in the plot indicates the beginning of the target zone.
Fig. 5.9. The \textit{a priori} model $m_{\text{prior}}$ used in the inversion of the Sorrento data.
5.4.1 Data Uncertainties

The data covariance matrix $C_D$ in Equation (5.1) and (5.2) defines the uncertainties in the data, which can be classified as observational and modeling uncertainties. As was discussed in Chapter 2, I will make the assumption that both uncertainties are modeled by Gaussian probability distributions given by (Equation (2.14) repeated here)

$$
\rho_d(d) \propto \exp \left[ (d - d_{\text{obs}})^T C_{\text{OBS}}^{-1} (d - d_{\text{obs}}) \right],
$$

$$
\Theta(d|m) \propto \exp \left[ (d - g(m))^T C_{\text{MOD}}^{-1} (d - g(m)) \right],
$$

where $C_{\text{OBS}}$ and $C_{\text{MOD}}$ are the covariance matrices associated with the observational and modeling uncertainties, respectively. As discussed in Chapter 2 those probabilities are defined on the data space $d$, and characterize our uncertainties about the recorded data and the synthetic data generated by the forward modeling operator $g(m)$. It is shown by Tarantola (1987) that, under the Gaussian hypothesis, the data covariance matrix $C_D$ consists of the sum of individual covariances $C_{\text{OBS}}$ and $C_{\text{MOD}}$. This result will be used here to incorporate into the calculation both types of data uncertainties. Specifically, I considered the following ones:

- ambient noise
- near-surface heterogeneities
- scaling factor between field and synthetic data
- model discretization errors

The first two fall into the category of data uncertainties while the others should be regarded as modeling uncertainties. Therefore, the final data covariance matrix $C_D$ is given by

$$
C_D = C_{\text{OBS}} + C_{\text{MOD}}
$$

$$
C_{\text{OBS}} = C_{D}^{\text{AN}} + C_{D}^{\text{ST}}
$$

$$
C_{\text{MOD}} = C_{D}^{\text{SC}} + C_{D}^{\text{DS}},
$$

where the superscripts $\text{AN}$, $\text{ST}$, $\text{SC}$ and $\text{DS}$ refer to the covariance matrices associated with ambient noise, residual statics, scaling factor and discretization errors, respectively. The addition of the covariance matrices $C_{D}^{\text{AN}}$ and $C_{D}^{\text{ST}}$ to form $C_{\text{OBS}}$ and of the covariance matrices $C_{D}^{\text{SC}}$ and $C_{D}^{\text{DS}}$ to form $C_{\text{MOD}}$ is justified because the uncertainties due to the ambient noise and near surface, as well as the uncertainties due to the scaling factor and model discretization can be considered independent (Priestley, 1981). The formal definition of the covariance matrices mentioned above is given by

$$
C_{D}^{\text{AN}} = E \left[ (d - d_{\text{obs}}) (d - d_{\text{obs}})^T \right],
$$
\[ C_D^{ST} = E \left[ (\mathbf{d} - \mathbf{d}_{\text{obs}})(\mathbf{d} - \mathbf{d}_{\text{obs}})^T \right], \]
\[ C_D^{SC} = E \left[ (\mathbf{d} - g(\mathbf{m}))(\mathbf{d} - g(\mathbf{m}))^T \right], \]
\[ C_D^{DS} = E \left[ (\mathbf{d} - g(\mathbf{m}))(\mathbf{d} - g(\mathbf{m}))^T \right]. \]

The elements along the main diagonal of these matrices are the variances of the fluctuations about the mean of the probability distributions in (5.3), and the off-diagonal elements show to what extent these fluctuations are correlated. Such fluctuations are due to the specific noise under consideration. Note that the mean of those distributions are defined by the observed data \( \mathbf{d}_{\text{obs}} \) in \( \rho_d(\mathbf{d}) \) and the synthetic data \( g(\mathbf{m}) \) in \( \Theta(\mathbf{d} | \mathbf{m}) \). For the model \( \mathbf{m} \) used in the generation of the synthetic data at which \( \Theta(\mathbf{d} | \mathbf{m}) \) is centered I chose the well logs shown in Figure 5.8, since it is representative of the subsurface in the Sorrento Field. The geometry of the Sorrento data shown in Figure 5.2 through 5.6 was used in the generation of this synthetic data. Note that strictly speaking (Tarantola, 1987) the covariance matrix \( C_{MOD} \) associated with modeling errors should be independent of the model \( \mathbf{m} \). Therefore the underlying hypothesis in the approach presented here is that this matrix can be estimated for a single model (defined by the well-log measurements) and used throughout in the inversion process. An alternative estimation procedure is to estimate this covariance matrix for an ensemble of models and select the one associated with the largest uncertainties, thus using a worst-case covariance matrix in the inversion. This, however, was not attempted in this work.

The calculation of the covariance matrices is similar for all items listed above in the sense that in all cases, the calculation boils down to the estimation of a sample covariance matrix from an ensemble of \( N \) “noise” seismograms \( \mathbf{s}_i \). The noise seismograms should represent fluctuations about the mean of the probability distributions defined in Equation (5.3). As I will show shortly, these fluctuations are computed by generating realizations of the data \( \mathbf{d} \) in accordance with the uncertainty being studied. The noise seismograms is then obtained by subtracting the appropriate mean, \( \mathbf{d}_{\text{obs}} \) in the case of observational uncertainties and \( g(\mathbf{m}) \) in the case of modeling uncertainties, from those realizations. Once this is done, one estimator of the covariance matrix from \( N \) of those seismograms is given by (Priestley, 1981)

\[ \hat{C} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{s}_i \mathbf{s}_i^T. \]

According to Equation (5.1) the dimension of the (square) data covariance matrix \( C_D \) should be the same as the data-misfit vector \( g(\mathbf{m}) - \mathbf{d}_{\text{obs}} \). Therefore, the noise seismograms \( \mathbf{s}_i \) should be of the same length as the data-misfit window, 1 s in this calculation, resulting in a covariance matrix of dimensions 251x251 (the time-sampling interval is 0.004 s). The construction of those seismograms, for each of the uncertainties considered in this inverse calculation, is discussed next.
**Ambient noise:** Here, the noise seismograms $s_i$ represent samples of the ambient noise in the data. These are constructed from time windows over the Sorrento data prior to the first arrivals within which no significant source-generated signal has been recorded. For this purpose, I have used the samples of the first 0.1 s of the data plotted in Figures 5.2 through 5.6 to form several 1-s noise seismograms $s_i$. This was done by simply concatenation of those samples from adjacent seismograms of the Sorrento data, starting from the leftmost trace of shot 1 shown in Figure 5.2 to the rightmost trace of shot 5 shown in Figure 5.6. Those seismograms are thought of as fluctuations about the observed data due to ambient noise. Thus, expression (5.6) can be used to estimate a sample ambient noise covariance matrix. The noise seismograms and the covariance matrix are illustrated in Figure 5.10. Note that this matrix is essentially band diagonal, which width shows to what extent the ambient noise is correlated, about 20 samples (0.08 s) in this case. The covariance matrix presents some events outside this main diagonal (e.g., the ones perpendicular to the main diagonal at indexes 60, 125 and 150). These artifacts are associated with population effects on the estimation of the sample covariance, since a finite number of noise seismograms is being used in this process, but also with the fact that source-generated noise is also present in the 0.1 s window used to extract samples of the ambient noise from the seismic data. This type of noise, which shows in the noise seismograms as laterally correlated events (e.g., see traces around 150 at time .6 s and traces around 450 at times .3 s and .6 s in Figure 5.10), is a short coming of the approach used in the estimation of the data covariance matrix. Ideally, traces containing just ambient noise samples should have been used, but those were not available in the seismic data set.

**Near-surface heterogeneities:** In the inversion of the Sorrento data the near-surface model, which includes depths down to 250 m, was derived from the well logs and kept fixed throughout the inversion process. This model, consisting of the five layers above the horizontal dashed lines shown in Figure 5.8, obviously represents an oversimplification of the complexity associated with the true near-surface. Statics corrections (field, residual and refraction statics) are kinematic procedures that aim at reducing the influence of near-surface anomalies in the reflected signal. Although these procedures are generally quite effective they are not error-free. Moreover, they do not account for the distortions caused by the near surface heterogeneities on the amplitudes of the reflected signal.

Accounting for all uncertainties associated with the near-surface is obviously a difficult, if not impossible, task. Some of these uncertainties, however, can be taken into account. Specifically I considered errors in the residual-statics-estimation procedure. This was done by corrupting the Sorrento data (Figures 5.2 through 5.6) with random surface consistent statics and then computing the data difference between the original and corrupted Sorrento data sets. These data residuals quantify the fluctuations about Sorrento data due to uncertainties in the residual-static values. Therefore, in accordance to the probability distribution $\rho_d(d)$ in Equation (5.3) and the definition of $C_D^{T}$
FIG. 5.10. (a) Noise seismograms; (b) Ambient noise covariance matrix.
in (5.5), this residual data set can be used to estimate the data covariance matrix $C_D^{ST}$ associated with such uncertainties. In this approach the residual data set was built from five realizations of the random-static values, which were limited to ± 8 ms and were generated by sampling an uniform random number generator. Given that the residual statics in the Sorrento data are in the order of 20 ms, 8 ms represents an error of 40% in the residual-static estimation procedure. Once the data residuals are obtained, a time window of length 0.5 s that encompasses the main reflection events of these data was used to form the noise seismograms $s_i$, again by simply concatenating such samples in the same way it was done for the covariance matrix associated with ambient noise. Equation (5.6) was then applied to compute the sample data covariance matrix. Both noise seismograms and data covariance matrix are illustrated in Figure 5.11. $C_D^{ST}$ indicates that the perturbations about the reflection events created by the static-time shifts have a higher frequency content than the ambient noise, due to the smaller width of its main diagonal band when compared to the one of $C_D^{AN}$. This difference in frequency content is expected since the reflection data itself is higher frequency than the ambient noise. The two off-diagonal bands in the covariance matrix are artifacts since adjacent traces of the data residuals are used to construct the noise seismograms. These off-diagonal bands which represent the correlations between those adjacent traces will not be relevant in the construction of the final covariance matrix. As will be shown later, this is because this matrix results from a summation of the data covariance matrices associated with each noise in consideration, which will emphasize the elements at the main diagonal band.

**Scaling factor between seismic and synthetic data:** An inversion procedure based on data fitting generally requires a scaling, or calibration, factor between synthetic and field data. This is often done by minimizing the least-squares misfit between the nearest-offset field seismic traces with synthetic seismograms computed for a nearby well log, for a given time window. Alternative approaches exist (Crase et al., 1990) in which the scaling factor is an unknown in the inverse calculation. For simplicity, I have adopted the former strategy in which I have used a reflectivity modeling algorithm with a point-force excitation to generate the synthetic seismograms, based on the well log MULL#14 from the Sorrento Field (Figure 5.8). The source signature consists of a zero-phase wavelet, band limited from 2 to 60 Hz, and computed from the inverse Fourier transform of a Hanning window (Oppenheim and Schafer, 1989). I consider this source signature accurate enough for the purpose of reproducing the field data, as will be shown later in this chapter.

To estimate the optimum scaling factor I selected near-offset traces from line 3 of shot 1 (Figure 5.1), considered to be sufficiently free of source-generated noise so that the scaling factor can be estimated in a reasonable fashion. The offsets of those traces are 303 m, 317 m and 345 m, and a time window from 0.48 s to 0.71 s was used in the process. The optimum scaling factor is chosen to be the one that minimizes the data least-squares misfit within a pre-specified interval. Figure 5.12 shows the field traces (odd trace number) and the synthetic traces (even trace number) scaled by this
Fig. 5.11. (a) Noise seismograms; (b) Near-surface covariance matrix.
optimum factor. This figure also shows the data misfit plotted against the scaling factor and the source signature used in the synthetic data generation.

This procedure is subject to uncertainties related to the well-log measurements and also to the choice of the wavelet used in the synthetics. Moreover, the data-misfit curve shown in Figure 5.12.c is relatively flat at the minimum such that a 20% change in the scaling factor results on a change of only 3% of the minimum misfit value. These issues motivate the incorporation of the uncertainties related to this scaling factor into the inverse calculation. As with the other covariance matrices, this is done by computing a covariance matrix from the fluctuations about the center of the distribution \( \Theta(d|m) \) in Equation (5.3) due to possible values for the scaling factor. As mentioned, the center of this distribution is taken to be the synthetic data generated for the well log scaled by the optimum factor, which is a reasonable choice since the well log is representative of the subsurface at Sorrento Field. Such fluctuations are obtained by the difference of this data set and the same synthetic data generated for the well log, but now scaled by different factors. Four factors were chosen such that errors up to 20% are considered in the optimum scaling factor. Specifically those factors were defined as .80, .85, .90 and .95 times the optimum one. Thus, four different data sets were used to construct data fluctuations from which the noise seismograms were derived. This was done by concatenating the samples of a 0.5 s time window that includes the main reflection events of this data residual. The concatenation was done in the same way as before, i.e, samples within the window of adjacent traces were used to form the noise seismograms. Figure 5.13 shows the noise seismograms and the covariance matrix obtained from these data via Equation (5.6). Note that the lower amplitude noise seismograms represent small fluctuations about the mean of \( \Theta(d|m) \) obtained when these fluctuations are computed with scaling factors close to the optimum one. The resulting covariance matrix shows a similar banded structure than the previous covariance matrices. The off-diagonal bands in this matrix are related to the correlation of the adjacent traces of the data perturbations used to construct the covariance matrix. Those are the same type of artifacts obtained when computing the covariance matrix associated with the near-surface heterogeneities. For reasons already mentioned, such artifacts will not be significant in the final covariance matrix. Note that the covariance matrix associated to uncertainties in the scaling factor shows a similar correlation length than that of the data fluctuations associated to near-surface heterogeneities. This can be expected since the synthetic and recorded data have the same frequency content, approximately.

**Model discretization errors:** An important *a priori* assumption made in the inversion of the Sorrento data set is that the subsurface can be described by a laterally homogeneous medium. Once this is done it remains to be determined how many layers should be used in the parameterization of the subsurface model. I have opted for a model parameterization consisting of 100 layers, each with constant thickness of 10 m. This choice was driven by computational reasons, and others are of course possible.

The purpose of a model covariance matrix associated with discretization errors
Fig. 5.12. (a) Near-offset field traces (odd trace number) and synthetics (even trace number) scaled by “optimum” factor. (b) Normalized data misfit as a function of the scaling factor. (c) The zero-phase wavelet used for generation of synthetic data.
Figure 5.13. (a) Noise seismograms; (b) Scaling-factor covariance matrix.
is to incorporate into the inverse calculation errors due to the parameterization of a continuous medium by a discrete model. Along the same methodology used previously, such a model covariance matrix will be estimated from noise seismograms that should represent fluctuations about the mean of $\Theta(d|m)$ which, again, was defined as the synthetic data generated for the well log (Figure 5.8) scaled by the optimum factor obtained from the data-misfit curve shown in Figure 5.12.c. The fluctuations about this mean are derived from the difference of this data set and the synthetic data generated for the full well log (Figure 5.7). Such fluctuations quantify to what extent the amplitudes of the synthetic data generated for the discretization employed in this inverse calculation can change if a (much) finer discretization level is adopted. The samples within a window of 0.5 s over the main events of these data difference were concatenated, as in the same way as before, to form the noise seismograms shown in Figure 5.14. This figure also shows the discretization-error covariance matrix $C_D^{DS}$, which is characterized by the same banded structure as seen in the other covariance matrices. The width of the main diagonal band of $C_D^{DS}$ shows a comparable correlation length than the one associated with the scaling-factor and near-surface related errors.

**Discussion:** As mentioned in the beginning of this section, under the Gaussian hypothesis, the final data covariance matrix is the sum of the covariance matrices associated with each one of the sources of uncertainties (Tarantola, 1987). Taking into consideration uncertainties related to ambient noise, near-surface heterogeneities, scaling factor and discretization errors I obtained the final covariance matrix illustrated in Figure 5.15. The elements of its main diagonal define the variances of the total data errors, and the width of its main diagonal indicates to what extent those are correlated. The off-diagonal artifacts, mainly observed in the covariance matrix related to uncertainties associated with the near surface and scaling factor are not significant in the final data covariance matrix since the main-diagonal-band elements were emphasized by the summation process. This covariance matrix is employed in the definition of the stopping criterion (5.2) and also in the construction of the a posteriori covariance matrix $C_M$ defined by Equation (4.7). Note that this matrix should be regarded as an average covariance matrix, since it was built from considerations about the entire Sorrento data set used in the inversion. Although possible, I have not attempted to estimate an individual data covariance matrix for each of the gathers shown in Figures 5.2 through 5.6.

In order to give some insight on the relative magnitude of the individual uncertainties for this particular calculation. The average variance (obtained by averaging all elements of the main diagonal of the covariance matrix) for the uncertainties associated with model discretization is the largest among all the ones under consideration. Specifically, the average variance of the model-discretization uncertainties is about four times larger than the average variance of the uncertainties related to ambient noise and near-surface heterogeneities, and roughly six times larger than the average variance of the uncertainties related to the computation of the scaling-factor. This illustrates the importance of considering discretization errors in an inference calculation, what,
Fig. 5.14. (a) Noise seismograms; (b) Discretization-error covariance matrix.
as far as I am aware, has not been attempted before.

It is important to remark that the procedure employed here to quantify the data uncertainties assumes that the noise seismograms represent realizations of the underlying distributions \( \rho_\delta(\mathbf{d}) \) and \( \Theta(\mathbf{d}|\mathbf{m}) \) that model the specific noise in consideration. Those noise seismograms were constructed under the following hypotheses:

- The window of .1 s used to characterize the ambient noise truly represents realizations of this type of noise, which correlation should be no larger than the length of this window. Source-generated events, if present in the noise seismograms, should not be representative in the estimation of the respective data covariance matrix.

- The distortions imposed by the near-surface on the reflection in the recorded data can be modeled by surface consistent residual statics. Thus, errors on the estimation of residual statics, assumed to be limited to \( \pm 8 \) ms, were used to quantify the covariance matrix associated with uncertainties on the data due to near-surface heterogeneities. Since the magnitude of the residual statics in the Sorrento data are in the order of 20 ms, 8 ms represents a maximum error of 40% in the residual-static-estimation procedure.

- The magnitude of the error in the scale factor is limited to 20% of its optimum value.

- The full well log measurements (Figure 5.7) can be used to quantify the uncertainties associated with the discretization employed in the calculation.

At this moment it is not clear how to carry out a sensitivity analysis of the resulting covariance matrices to the above assumptions, without having to repeat the whole inverse calculation for different choices in the construction of the data covariance matrices. This implies a computational cost much higher than the inversion step and, although this sensitivity analysis is an important issue, it has not been attempted in this work.

Moreover, the hypothesis of Gaussianity that underlies all the previous covariance computations is obviously opened to questions. As emphasized in Chapter 3, by modeling data uncertainties with Gaussian distributions, moments of the likelihood function higher than two are neglected. In this chapter, I also investigated approaches based on the maximum-entropy principle that could in theory be used to construct non-Gaussian likelihood functions. Such approaches are, however, of little practical value due to the difficulties related to the high dimensionality of the problem (in this case the dimension of the data space) and to the estimation of high-order statistics from sample data.

Finally, it is important to emphasize that not all possible uncertainties associated with the inverse problem have been accounted for in the Bayesian methodology proposed here. Examples are, among others, uncertainties related to the presence of anisotropy, absorption, lateral velocity variations and three-dimensional structural
5.4.2 Model Uncertainties

In the Gaussian assumption the model uncertainties about the a priori model $m_{\text{prior}}$ are defined by the model covariance matrix $C_M$. I applied the procedure described in Section 3.4 of Chapter 3 to compute this matrix from the well logs shown in Figure 5.8. This results in three covariance matrices: one for the $P$-wave impedance, one for the $S$-wave impedance and one for the density profile. The $P$-wave impedance and $S$-wave impedance covariance matrices were estimated from impedance logs, formed by the product of the $P$-wave velocity and density in the first case, and $S$-wave velocity and density in the second. The model covariance matrices are shown in Figure 5.16. It is important to (re-)emphasize that the a priori model covariance matrices define our prejudices about the subsurface, without taking into account the surface seismic data. For instance, Figure 5.17 shows the a priori

\footnote{The units used for impedances and density are $\text{km g s}^{-1} \text{cm}^{-3}$ and $\text{g cm}^{-3}$, respectively.}
Fig. 5.16. Model covariance matrices (a) P-wave impedance, (b) S-wave impedance and (c) density.
standard deviations given by the square roots of the main-diagonal elements of the three model covariances in Figure 5.16. Such standard deviations define an interval about \( \mathbf{m}_{\text{prior}} \) within which the true subsurface parameters are assumed to lie with a given degree of confidence. Learning about a particular parameter from the surface seismic data means reducing and/or relocating these intervals, thus refining the \textit{a priori} state of information about this specific parameter. If the data do not provide any information about a parameter, the \textit{a priori} knowledge about this parameter will remain unchanged by the inverse calculation. This is easily understood by looking at the expression of the \textit{a posteriori} covariance matrix when a Gaussian approximation is used for the \textit{a posteriori} probability distribution \( \sigma(\mathbf{m}) \) (Equation (4.7), repeated here)

\[
C_{M'} = \left[ G^H C_D^{-1} G + C_M^{-1} \right]^{-1}.
\] (5.7)

If the data are not informative the term \( G^H C_D^{-1} G \) will go to zero and \( C_{M'} = C_M \). As the data become more informative this term will increase, thus reducing the uncertainties about the subsurface parameters. In the next section I discuss estimation of MAP models for the Sorrento data, followed by analysis of the \textit{a posteriori} uncertainties associated with the inverse calculation.

5.5 Explaining the recorded data

The MAP model is derived from an iterative optimization process applied to the objective function (5.1). The optimization is implemented with the same algorithm used in the synthetic calculation: an initial model is systematically perturbed until it fits the data well enough to satisfy the stopping criterion (5.2). The data-misfit window within which the synthetic and observed data are compared is 1 s long and starts at the time of 0.2 s, and encompasses the main reflection events of the gathers displayed in Figures 5.2 through 5.6. All large-amplitude first arrivals have been suppressed by muting prior to the inversion, since those events were attributed to reflections and refractions from the near-surface (including depths shallower than 250 m). Recall that the parameters of the near-surface layers are kept constant throughout the inversion process since they are outside the depth target zone, which is bounded by the depths of 250 m and 1250 m. \( \mathbf{m}_{\text{prior}} \), shown in Figure 5.9, was chosen as the initial model for the optimization. The synthetic data generated for this model (Figure 5.18) show little reflected energy, as expected due to the model’s smooth character. The wavelet used in the calibration of the field and synthetic data described before (Figure 5.12) was used to generate this gather and throughout in the inversion process. All 25 gathers shown in Figures 5.2 to 5.6 were inverted for a \( P \)-wave impedance, \( S \)-wave impedance and density profiles, yielding 25 MAP subsurface models. The initial guess was the same for all gathers, as were the data covariance and model covariance matrices.

Figures 5.19 through 5.23 show the MAP models derived from each of the shot gathers. All these models fit the observed data to the level of the noise in the data in the sense of Equation (5.2). The error bars shown in the figures are the square
Fig. 5.17. *A priori* standard deviations (a) $P$-wave impedance, (b) $S$-wave impedance and (c) density.
roots of the main diagonal elements of the \textit{a posteriori} covariance matrix (5.7), the computation of which will be discussed in the next section. The number of conjugate gradient iterations needed to compute the MAP models ranged from 2 (for lower quality gathers) to 8 (for gathers with better signal content).

The synthetic data computed for each of the MAP models are shown in Figure 5.24 through 5.28. Comparison of these gathers with the Sorrento data (Figures 5.2 through 5.6) indicates that the recorded data was explained to a good extent. The main reflection events of this data within the data-misfit window have been reproduced by the synthetic data generated for the MAP models. This is more clearly shown by the data residuals, which is the difference between the field and synthetic data for each of the MAP models. These residuals, shown in Figures 5.29 through 5.33 \footnote{The same amplitude scale is used in all seismic data plots throughout this chapter.}, are mainly associated with non-hyperbolic reflections (See the events at around .7 s in the second gather of Figure 5.30 and in the fourth gather of Figure 5.33.) and source-generated noise. (Examples are the third gather of Figure 5.31 and the second gather of Figure 5.32.) The former is due to mild lateral velocity variations present in the Sorrento Field. Such non-hyperbolic events cannot be modeled with a elastic isotropic reflectivity modeling algorithm since the reflectivity method does not take into account these variations in the the subsurface.

\begin{figure}[h!]
\centering
\includegraphics[width=0.5\textwidth]{fig5.18.png}
\caption{Synthetic shot gather generated for the initial model (Figure 5.9). The geometry of shot 1 line 4 (Figure 5.1) was used in the modeling.}
\end{figure}
Fig. 5.19. MAP model derived from shot gather 1 (Figure 5.2): (a) P-wave impedance profiles. (b) S-wave impedance profiles. (c) Density profiles. The numbers at the top of each profile are associated with the line numbers. The same plot organization is used for Figures 5.20 through 5.23.
Fig. 5.20. MAP model derived from shot gather 2 (Figure 5.3).
Fig. 5.21. MAP model derived from shot gather 3 (Figure 5.4).
Fig. 5.22. MAP model derived from shot gather 4 (Figure 5.5).
Fig. 5.23. MAP model derived from shot gather 5 (Figure 5.6).
Fig. 5.24. Synthetics computed for the MAP models estimated from shot gather 1.

Fig. 5.25. Synthetics computed for the MAP models estimated from shot gather 2.
5.5.1 Discussion

It is important to emphasize that, although all models shown in Figures 5.19 through 5.23 fit the seismic data in accordance with the stopping criterion (5.2), they are not “the” solution of the inverse problem. As will be discussed shortly, different models fit the observed data equally well. In the Bayesian framework the a posteriori probability distribution $\sigma(m)$ will be used to provide insight on the uncertainties associated with the waveform seismic data, via several confidence measures about the subsurface parameters. Such measures will take into account the a priori uncertainties on the data and on the subsurface parameters, given by the data and model covariance matrices $C_D$ and $C_M$ respectively, and the resolution provided by the seismic data. At this point, however, the following conclusions can be drawn by inspection of the models and their associated error bars shown in Figures 5.19 through 5.23.

- The error bars of the MAP models tend to increase with depth. Note, however, that these error bars are limited to the magnitude of the a priori error bars on the subsurface parameters, as discussed in Section 5.4.2. Figure 5.17 shows that the a priori error bars begin to decrease beyond some depth (around 1 km for the $P$-wave and $S$-wave impedance and around .9 km for the density), which is due to the fact that the well-log measurements fluctuate to a lesser extent about $m_{\text{prior}}$ at those depths, when compared to shallower ones. Therefore, the a posteriori error bars shown in the MAP figures will follow this trend, and this is the reason why they are smaller at larger depths (e.g., beyond 1.1 km) than
Fig. 5.27. Synthetics computed for the MAP models estimated from shot gather 4.

Fig. 5.28. Synthetics computed for the MAP models estimated from shot gather 5.
Fig. 5.29. Residuals associated with the MAP model obtained from shot gather 1.

Fig. 5.30. Residuals associated with the MAP model obtained from shot gather 2.
Fig. 5.31. Residuals associated with the MAP model obtained from shot gather 3.

Fig. 5.32. Residuals associated with the MAP model obtained from shot gather 4.
Fig. 5.33. Residuals associated with the MAP model obtained from shot gather 5.

at shallower depths.

- The $P$-wave impedance error bars are the smallest ones, followed by the $S$-wave impedance and density error bars. This is in accordance with what was obtained in the calculation using synthetic data developed in Chapter 4, and indicates the larger information content available in the vertical component of the recorded displacement field about $P$-wave impedance when compared to $S$-wave impedance and density.

- As seen, the MAP models are computed from an iterative optimization procedure, which perturbs an initial model ($m_{\text{prior}}$) until the seismic data is fit according to a pre-specified criterion defined by (5.2). Recalling that the seismic response of the initial model $m_{\text{prior}}$ lacks reflected energy, since this model is smooth, not many iterations will be required to fit a seismic gather of poor data quality, i.e., with little reflection data. Thus, in this circumstance, the MAP model will be smoother, or more featureless, when compared with MAP models derived from better-quality data.

In Sorrento, large-offset gathers present in general a smaller signal-to-noise ratio than do near-offset ones. Obviously, this is due to amplitude losses related to the propagation of the seismic wavefield through the subsurface, which will be larger for larger source-receiver offsets. Therefore, given that the noise level in the data quantified by the data covariance matrix $C_D$ is assumed to be the same
for all gathers, the number of iterations required to estimate the MAP models from low-quality larger-offset data is smaller than the one required to estimate the MAP models from better-quality near-offset data. Thus the former models will present a smoother character than the latter ones. In other words, for the vertical component of the displacement field recorded at Sorrento, larger-offset gathers tend to provide less information about the subsurface parameters than do nearer-offset ones.

For example, this is shown by contrasting the MAP model computed from line 4 of shot gather 1 with the MAP model derived from line 6 of the same shot (both shown in Figure 5.19). The former shows considerably less short-wavelength depth features than the latter, which is due to less signal content exhibited by the seismic data acquired at line 6, as compared to the data acquired at line 4 (see Figure 5.2). Another example is given by the MAP models derived from lines 1, 3 and 7 of shot gather 5. The one estimated from the data acquired at line 3 shows considerably more detail than the other two. Again, this correlates with the larger signal content of the gather acquired at this line when compared to the data acquired at lines 1 and 7 for this shot gather (see Figure 5.6).

As will be shown shortly, the lower resolution provided by larger-offset gathers about the subsurface parameters can be quantified by inspecting the Gaussian approximation to the *a posteriori* probability distribution. As indicated by Equation (4.7), the covariance matrix of this probability distribution is a function of the *a priori* data and model uncertainties quantified by the matrices $C_D$ and $C_M$, and a function of the Jacobian (or Frechet derivative) matrix $G$, obtained from the linearization of the forward modeling operator $g(m)$ about the MAP model. The matrix $G$, which quantifies the resolution provided by the seismic data alone, is the factor responsible for incorporating the specific data-acquisition geometry into the resolution analysis.

### 5.6 Uncertainty Analysis

All questions related to uncertainties associated with the MAP models presented in the previous section have to be addressed via the *a posteriori* probability distribution $\sigma(m)$. As discussed in Chapter 4, the complicated nature of this function due to the nonlinearity associated with the reflectivity integral requires Monte Carlo sampling procedures for this uncertainty analysis. Considering the dimensionality of the inverse problem and the high computational cost of the modeling algorithm this approach is not practical. Therefore, as was done in the previous chapter, I perform a linearization of the forward modeling operator about the MAP model, which yields a Gaussian approximation, $\sigma_g(m)$, to the *a posteriori* probability distribution. This approximation, given by Equation (4.6), is repeated below.

$$
\sigma_g(m) \propto \exp \left[ - \frac{1}{2} (m - m_{\text{MAP}})^T C_{MP}^{-1} (m - m_{\text{MAP}}) \right].
$$

(5.8)
For the Sorrento data there are 25 \textit{a posteriori} covariance matrices $C_{M'}$, each one associated with a given MAP model. A typical matrix, the one associated with the MAP model for line 4 of shot gather 1, is illustrated in Figure 5.34. This 3x3 block matrix shows the auto-correlation of the $P$-wave impedance, $S$-wave impedance and density on its main block diagonal, and the cross-correlations among different parameters in the off-diagonal blocks. The blocked structure of this matrix is pretty similar to that of the covariance matrices associated with the other MAP models. The energy in the off-diagonal blocks indicates the degree of coupling among different parameters. Since the \textit{a priori} model covariance is block diagonal (no cross-correlation among the subsurface parameters were considered in this problem), and the data covariance matrix is quasi-diagonal, the non-zero elements in the off-diagonal blocks of $C_{M'}$ are due to the Jacobian $G$ in Equation (5.7). In other words, the non-diagonal elements in the \textit{a posteriori} covariance are due to correlations of the synthetic seismic data generated when a given subsurface parameter is perturbed with the synthetic seismic data generated when another subsurface parameter is perturbed. The extent to which, both seismic data are different, i.e., uncorrelated, is related to the degree of data resolution for those two specific parameters. Specifically, if such data sets are exactly the same, the data provide no resolution about those two parameters.
Although the blocked structures of the data covariances are similar for different MAP models, the magnitudes of their elements are not. This is well demonstrated by Figures 5.35 through 5.39. These figures contrast the \textit{a posteriori} standard deviations (the error bars shown in Figures 5.19 through 5.23, derived from the square roots of the main diagonal of the respective \textit{a posteriori} model covariances) with the \textit{a priori} standard deviations (square roots of the main diagonal of $C_M$, shown in Figure 5.17). As discussed in Section 5.4.2 the magnitude of the \textit{a posteriori} error bars are limited to the magnitude of the \textit{a priori} error bars. Moreover, the extent to which the \textit{a priori} errors bars are reduced by the inverse calculation is a measure of the resolution on the subsurface parameters provided by the data. Figures 5.35 through 5.39 clearly show the larger resolution provided by the seismic data on the $P$-wave impedance, when compared with the $S$-wave impedance and density. In fact, the density \textit{a priori} and \textit{a posteriori} error bars are essentially the same, what is indicative of the lack of resolution provided by the Sorrento data on this parameter.

In these figures one can see a correlation between the reduction of the \textit{a priori} standard deviations and the data quality (or signal-to-noise ratio) of the shot gathers (Figure 5.2 through 5.6). Consider two specific shot gathers, situated at extremes in data quality. These gathers, shown in Figure 5.40, are the ones acquired by the geophone groups located at line 4 of shot 1 and at line 7 of shot 5 (Figure 5.1), respectively. In Figure 5.41, the same gathers are shown but now with the synthetic data generated for the respective MAP models. Notice that this figure shows the field and synthetic data as a single gather, with synthetic and field traces alternating so that the data fitting can be more clearly assessed. It is clear from these data that the seismic data acquired at line 4 of shot 1 is more informative about the subsurface than the one acquired at line 7 of shot 5, due to its higher signal content. This statement is quantitatively confirmed by Figure 5.42, which contrasts the \textit{a priori} and \textit{a posteriori} standard deviations for the two lines. (These curves are also shown in Figures 5.35 and 5.39.) Those plots show the larger resolution provided by the data acquired at line 4 of shot 1 on the $P$- and $S$-wave impedance parameters. Almost no resolution on $S$-wave impedance is provided by the data acquired at line 7 of shot 5, since \textit{a posteriori} and \textit{a priori} error bars are essentially the same. For the same reason, one can conclude that neither one of the gathers has no further information on the density of the subsurface than the one incorporated into the calculation via the \textit{a priori} distribution.

This difference in resolution is due to the Jacobian $G$ in the expression of the \textit{a posteriori} covariance matrix $C_M$, since the covariances $C_D$ and $C_M$ related to the \textit{a priori} uncertainties in the data and in the subsurface parameters are the same for all gathers used in the inversion. As emphasized in Section 5.5.1, the matrix $G$, being a measure of data resolution, quantifies the sensitivity of the recorded seismic data (taking into consideration its acquisition geometry) to changes in the subsurface parameters. Therefore, Figure 5.42 indicates the larger resolution provided by the data acquired at line 4 of shot 1 than the one of the data acquired at line 7 of shot 5, due to its higher signal content.
Fig. 5.35. Comparison of *a priori* and *a posteriori* standard deviations for the MAP models estimated from shot gather 1.
Fig. 5.36. Comparison of \textit{a priori} and \textit{a posteriori} standard deviations for the MAP models estimated from shot gather 2.
Fig. 5.37. Comparison of \textit{a priori} and \textit{a posteriori} standard deviations for the MAP models estimated from shot gather 3.
Fig. 5.38. Comparison of a priori and a posteriori standard deviations for the MAP models estimated from shot gather 4.
Fig. 5.39. Comparison of *a priori* and *a posteriori* standard deviations for the MAP models estimated from shot gather 5.
Fig. 5.40. Seismic gathers acquired by the receiver groups located at (a) line 4 of shot 1, and at (b) line 7 of shot 5. Recall that, in both gathers, the high-amplitude first-arrivals are outside the data-misfit window. The signal-to-noise ratio of the gather acquired at line 4 of shot 1 is higher than the one of the gather acquired at line 7 of shot 5.
Fig. 5.41. Synthetic data generated for the MAP models derived from the gathers displayed in Figure 5.40. The synthetic and field data are plotted in such a way that their traces alternate, to facilitate the comparison. Note that the data have been muted to suppress the first arrivals.
Fig. 5.42. *A posteriori* standard deviations computed for the MAP models estimated from the seismic gathers acquired at line 4 of shot 1 and at line 7 of shot 5 (Figure 5.40).
The standard deviations shown in Figures 5.35 through 5.39 provide a limited view of the uncertainties associated with the MAP estimates. By taking just them into account, one is neglecting the off-diagonal elements of the \textit{a posteriori} covariance matrix \( C_{M'} \). Sampling the \textit{a posteriori} distribution \( \sigma_{\rho}(m) \) to generate pseudo-random realizations of the subsurface incorporates all the information available in \( C_{M'} \). Moreover, examining those pseudo-random models displayed side by side can be useful in unveiling subsurface features that are resolved by the data. Features that are present in most of the realizations are associated with high-confidence estimates of the subsurface.

Next, let us apply the same sampling procedure used in the previous chapter, based on the \( LU \) decomposition of \( C_{M'} \), to the gathers shown in Figure 5.40. Again, those two gathers were chosen because they represent the extremes in data quality. Figures 5.43 and 5.44 show 100 pseudo-random subsurface realizations associated with these data. As expected, P-wave realizations tend to show the smallest degree of variation, followed by the S-wave realizations and then density. The arrows in Figure 5.43 show some of the short-wavelength depth features that can be distinguished among the P- and S-wave impedance realizations. Such features should be interpreted as well resolved by the seismic data. Note that the density realizations demonstrate that nothing could be learned from the data about this parameter. All realizations associated with the gather acquired at line 7 of shot 5 show small to no continuity, indicating that these surface seismic data provided almost no information about the subsurface parameters.

As was done in the previous chapter, it is interesting to assess the information content of the \textit{a priori} probability distribution \( \rho(m) \) not only by considering the standard deviations (as in Figure 5.17) but also by sampling this distribution to generate pseudo-random realizations of subsurface models. Some \textit{a priori} realizations are displayed in Figure 5.45. The only apparent feature that these realizations have in common is the long wavelength depth trend given by \( m_{\text{prior}} \). Therefore, short-wavelength depth features that are laterally continuous in Figures 5.43 and 5.44 must represent features resolved to some extent by the data.

Finally, a third mechanism to assess the uncertainties associated with the inverse calculation is the use of unidimensional \textit{a posteriori} marginal distributions. A unidimensional marginal distribution of a probability density \( p(x) \) of a vector random variable \( x \) is obtained by integrating this function in all dimensions but one. Specifically, the marginal of \( p(x) \) on the \( j-th \) component of \( x \) is defined as

\[
p(x_j) = \int_D p(x) \, dx_1 \, dx_2 \ldots dx_{j-1} \, dx_{j+1} \ldots dx_n,
\]

where \( n \) is the dimension of \( x \) and \( D \) is the domain of integration. \( p(x_j) \) characterizes the state of information about the parameter \( x_j \) given by the multidimensional probability distribution \( p(x) \). Therefore, it is a suitable tool for studying the resolution of individual subsurface parameters. Without the Gaussian approximation adopted in this work the marginal distributions would have to be constructed by histograms.
Fig. 5.43. Pseudo-random realizations of the subsurface generated for the seismic gather acquired at line 4 of shot 1 (shown in Figure 5.40.a). Some resolution is provided by the data for P-wave impedances and for S-wave impedances.
Fig. 5.44. Pseudo-random realizations of the subsurface generated for the seismic gather acquired at line 7 of shot 5 (shown in Figure 5.40.b). Little or no resolution is provided by the data for any of the subsurface parameters.
Fig. 5.45. Pseudo-random realizations computed from the \textit{a priori} probability distribution $\rho(m)$. 
obtained from Monte Carlo sampling of the \textit{a posteriori} probability distribution $\sigma(m)$. As discussed previously, such Monte Carlo approaches are not practical in the context of seismic waveform inversion, due to the many forward modeling calculations required by this method. For Gaussian multidimensional probability densities, however, there are analytic expressions for the marginals (Miller, 1964), which I used here to compute the \textit{a posteriori} marginals for some of the subsurface parameters estimated with the Sorrento data.

Specifically, these marginals were computed for the $P$-wave, $S$-wave and density parameters estimated from the seismic gathers shown in Figures 5.40, at four specific depths: 0.6, 0.8, 1.0 and 1.2 km. All marginals are shown in Figures 5.46, 5.47 and 5.48. There is also shown the marginals of the \textit{a priori} distribution at the same depths. Since both \textit{a priori} and \textit{a posteriori} probabilities quantify information about the subsurface, the former taking into account the well-log measurements and the latter considering the well logs and the seismic data, the extent to which both probabilities differ is a measure of the seismic-data resolution.

The following conclusions can be drawn from these plots:

- The more informative character of the $P$-wave marginals when compared to the $S$-wave and density marginals, shows that the $P$-wave impedance is the best resolved parameter of the calculation.

- The marginals derived from the data acquired at line 4 of shot 1 are more informative than the marginals derived from the data acquired at line 7 of shot 5. This is consistent with the little reflection energy available in the latter gather when compared to the former one.

- The \textit{a posteriori} marginals for the density estimates (Figure 5.48) are very similar to the \textit{a priori} ones. This is, again, indicative of the lack of information about density available in seismic waveform, compared to the \textit{a priori} information used in the calculation.

- In some situations (e.g., for the $S$-wave impedance-parameter estimate at depth 1.2 km (Figure 5.47) and for the density-parameter estimates at depths 0.6 and 0.8 km (Figure 5.48), both derived from the data acquired at line 4 of shot 1), the \textit{a posteriori} marginal probability density is basically a shifted copy of the \textit{a priori} probability. This illustrates that \textit{a priori} and \textit{a posteriori} standard deviations are essentially the same for those parameters, but the mean of these distributions are not ($m_{\text{prior}}$ and $m_{\text{map}}$, respectively). In other words, the inverse calculation only re-centered the \textit{a priori} distribution for those parameters.

### 5.6.1 The Occam approach

The MAP models for the Sorrento data, shown in Figure 5.19 through 5.23, were estimated by performing the inversion at a scale (10 m) which is well below the $P$-wave
Fig. 5.46. P-wave impedance marginals, at depths 0.6, 0.8, 1.0 and 1.2 km. (a) Marginals associated with the data acquired at line 4 of shot gather 1 (Figure 5.40.a). (b) Marginals associated with the data acquired at line 7 of shot gather 5 (Figure 5.40.b).
Fig. 5.47. $S$-wave impedance marginals, at depths 0.6, 0.8, 1.0 and 1.2 km. (a) Marginals associated with the data acquired at line 4 of shot gather 1 (Figure 5.40.a) (b) Marginals associated with the data acquired at line 7 of shot gather 5 (Figure 5.40.b).
Fig. 5.48. Density marginals, at depths 0.6, 0.8, 1.0 and 1.2 km. (a) Marginals associated with the data acquired at line 4 of shot gather 1 (Figure 5.40.a) (b) Marginals associated with the data acquired at line 7 of shot gather 5 (Figure 5.40.b).
seismic wavelength at the target depth zone (approximately 120 m wavelength). Since the \emph{a priori} distribution was derived from the well log, it incorporates features that are not constrained by the data. It is therefore pertinent to carry on the inversion without the \emph{a priori} distributions computed from the well logs, but instead, under the Occam framework, as explained in the last chapter. Again I used the second-order difference regularization matrix (Equation (4.10)), and sought the model associated with the largest regularization parameter that yields a data fitting according to the criterion (5.2). This approach, just applied to the gather acquired at line 4 of shot gather 1 (Figure 5.40.a), results in the Occam estimate $m_{\text{occam}}$ which is plotted besides the Bayes estimate (model 4 in Figure 5.19) in Figure 5.49. This figure shows in detail the significant difference between the Bayes and the Occam subsurface models. Several features present in the Bayes model, which were derived from the well-log measurements, are not seen in the Occam model. For example, compare both $P$-wave impedance profiles at depth 0.8 km, and both $S$-wave impedance and density profiles at the depths of 0.5 and 0.8 km. Notice that, the extent to which such models differ is larger than that in the synthetic calculation done in the previous chapter, indicating that the well log had a stronger influence in the inversion of the Sorrento data. Such influence is of course dependent on the relative contributions of the likelihood and the \emph{a priori} distribution in the objective function (5.1). Figure 5.50 shows the synthetic data computed for the Bayes and Occam models. The difference of those data sets are shown in Figure 5.51. Although differences exist, they are, however, within the noise level in the observed data.

5.7 General discussion

The Occam and the Bayes MAP models estimated in this study, by fitting the data to the same pre-specified criterion, afford a clear illustration of the non-uniqueness issue in seismic waveform inversion and brings insight on the distinctions between the two approaches. The Bayesian model, being the center of the \emph{a posteriori} distribution, is consistent with the data and the well-log measurements. The Occam model is derived solely from the seismic data. Since the goals of the two inverse calculations are different, it makes no sense to ask which subsurface model is more accurate. Further information has to be brought into the calculation, besides the seismic and well-log data, to narrow the spectrum of possible solutions to this inverse problem. One should be aware, however, that the differences in the models shown in Figure 5.49 are the implications of using \emph{a priori} distributions built from information about the subsurface that is of higher resolution than that which the data can provide. In this study, such a information was derived from well-log measurements of the subsurface acquired at a borehole nearby the receiver locations at which the seismic data were acquired. Note that I am making the assumption that the same \emph{a priori} distribution can be used in the inversion of all seismic gathers shown in Figures 5.2 through 5.6. Given that the geology at Sorrento Field is essentially vertically stratified with minor structure features, this assumption seems reasonable.
It is important to acknowledge that the Bayesian calculation developed here indicates that the inversion of the Sorrento data results in a large degree of uncertainty about the subsurface parameters. This is verified by the resolution measures provided by the \textit{a posteriori} distribution, such as the error bars, the pseudo-random realizations of the subsurface and the marginal distributions. As it has been discussed in Section 5.4.1, such measures depend on the assumptions made to construct the \textit{a priori} data and model covariance matrices. Nonetheless, they are consistent with the large difference between the Occam and Bayes models shown in Figure 5.49, and do illustrate the broad spectrum of subsurface models that can be considered solutions of the inversion of the Sorrento data.

As a final remark it should be pointed out that no lateral correlation among the MAP models estimated from the Sorrento data was incorporated into the inversion procedure described in this chapter. In order to do so a two-dimensional inversion should be carried out, allowing for the presence of lateral inhomogeneity in the medium in a way consistent with the seismic data. Nonetheless, the 25 MAP models, when sorted according to their spatial locations\footnote{The location of a given MAP model is defined as the common-mid-point coordinate given by the source location and the minimum-offset receiver location for the seismic gather from which the MAP model was estimated.} and displayed side by side, do show some degree of lateral correlation, as is shown in Figure 5.52, consistent with the stratification of the Sorrento field. However, drawing conclusions from this image in terms of the lateral continuity of events is not appropriate since the seismic gathers were inverted independently. The results obtained from the Bayesian inversion of the Sorrento data, as was performed here, should be interpreted for each gather by sampling the \textit{a posteriori} distribution, as done in Figures 5.43 and 5.44, or via the marginal probability distributions as the ones displayed in Figures 5.46, 5.47 and 5.48. The analysis of the MAP model alone provides an incomplete picture of the solution to the inverse problem. Although it is an important model since it is the center of the \textit{a posteriori} distribution $\sigma(\mathbf{m})$, it lacks any information on the uncertainties associated with the calculation.

5.8 Interpreter’s corner

The fact that two significantly different models, the Bayes and the Occam models displayed in Figure 5.49, fit the Sorrento seismic data to the same pre-specified accuracy, illustrates the ambiguous character of the seismic waveform data used in this calculation. Therefore, a seismic interpreter, when deriving information about these data, has to take into consideration the spectrum of subsurface models that are consistent with the data under consideration.

The Bayesian methodology provides one way to assess the uncertainties associated with seismic waveform inversion by offering as the solution of the inverse problem a probability distribution on the space of models, instead of a single subsurface model.
Fig. 5.49. Bayes and Occam estimates derived from line 4 of shot 1 (Figure 5.2). (a) $P$-wave impedance. (b) $S$-wave impedance. (c) density profiles.
Fig. 5.50. (a) Synthetic data generated for the Bayesian estimate. (b) Synthetic data generated for the Occam estimate. Both models are plotted in Figure 5.49.
Such a probability distribution encapsulates the information derived from the seismic and other sources of data about the subsurface parameters.

Specifically in this study, I used a Gaussian distribution to approximate the \textit{a posteriori} distribution $\sigma(m)$ about its maximizer (the model $m_{map}$). This approximation was done to expedient the procedure to extract confidence measures about the subsurface from $\sigma(m)$. Otherwise, such measures have to be derived by sampling this distribution via Monte Carlo procedures, thus raising the cost of the inverse calculation. Note that this Gaussian approximation characterizes the uncertainties about $m_{map}$. Since this model preserved the long-wavelength depth component incorporated into the inversion via $m_{prior}$, it yields the right kinematics of the reflections events in the data. Thus, the Gaussian approximation can be considered a local approximation of the \textit{a posteriori} probability distribution on a region of the model space, in which the subsurface models satisfy the kinematics of the reflections in the seismic data.

The \textit{a posteriori} probability distribution is in fact a device to which the seismic interpreter can address questions about the subsurface parameters. Obviously, the answers obtained from this distribution is in the form of probabilities derived from its integration on a given domain of the model space. An example of such questions is what is the probability that the $P$-wave impedance of a given layer is within a certain range. This probability is computed from the integration of the marginal \textit{a posteriori} distribution for this parameter within this range. Let $p(m_P)$ be this marginal distribution for the $P$-wave impedance $m_P$ under consideration. Thus, the
Fig. 5.52. MAP models estimated from the Sorrento Data (Figures 5.19 through 5.23) sorted and displayed side by side. (a) P-wave impedance, (b) S-wave impedance and (c) density profiles. The respective gray scale is shown at the side of each plot.
W.P. Gouveia, Jr.

probability $P$ that $m_P$ is within the range $[m_P^1, m_P^2]$ is given by

$$P = \int_{m_P^1}^{m_P^2} p(m_P) \, dm_P.$$  

(5.10)

In the case of Sorrento field the probability that the $P$-wave impedance at a depth of 0.6 km lies between the values of 8 and 9 km g s$^{-1}$ cm$^{-3}$ is about 0.82. This probability was derived by integrating the marginal shown in the top of the left column in Figure 5.46, which was estimated for the seismic gather acquired at line 4 of shot 1 (Figure 5.40.a).

Questions involving more than one parameter can also be addressed. For example, the probability $P$ that the $P$-wave impedance and density of a given layer are limited by a certain interval, such as $[m_P^1, m_P^2]$ and $[m_P^1, m_P^2]$, respectively. This probability is derived from the marginal distribution $p(m_P, m_\rho)$ associated with those two parameters, and it is given by

$$P = \int_{m_P^1}^{m_P^2} \int_{m_\rho^1}^{m_\rho^2} p(m_P, m_\rho) \, dm_P \, dm_\rho.$$  

(5.11)

Obviously, more elaborated questions involving more parameters of the subsurface have to be addressed by integrating the multidimensional marginals associated with these parameters. As one more example, the probability that a layer which top is at depth $n \, dz$, where $dz$ is the depth discretization interval used in the inversion, has a thickness of $3 \, dz$ according to the $P$-wave impedance is given by

$$P = P(m_P^1 \leq m_{P n} \leq m_P^2, m_P^1 \leq m_{P n+1} \leq m_P^2, m_P^1 \leq m_{P n+2} \leq m_P^2)$$

$$= \int_{m_P^1}^{m_P^2} \int_{m_P^1}^{m_P^2} \int_{m_P^1}^{m_P^2} p(m_P, m_{P n+1}, m_{P n+2}) \, dm_P \, dm_{P n+1} \, dm_{P n+2},$$  

(5.12)

where $p(m_P, m_{P n+1}, m_{P n+2})$ is the marginal probability density associated with $P$-wave impedances at depth $n \, dz, (n+1)dz$ and $(n+2)dz$. The limits of integration $m_P^1$ and $m_P^2$ define the range within which the $P$-wave impedance is allowed to vary inside the layer. The fact that the $a$ posteriori distribution is approximated by a Gaussian distribution facilitates the derivation of such marginals, which have analytic expressions in this case.

Equations (5.10), (5.11) and (5.12) are simple examples on how the $a$ posteriori probability $\sigma(\mathbf{m})$ can be used to answer some of the questions addressed by the seismic interpreter. One of the objectives of this work is to propose a methodology such that this probability is built honoring the uncertainties due to errors in the data and due to possible subsurface models.

In focusing the proposed methodology for practical applications, I have made several assumptions in the formulation of $\sigma(\mathbf{m})$. Specifically, those are the assumptions about Gaussian uncertainties and the ones underlying the procedures employed to estimate the data and model covariance matrices. Those were discussed in Section 5.4,
and will be emphasized next, in the conclusions.
Chapter 6

CONCLUSIONS

In this research I have developed a theoretical and computational strategy for making quantitative statistical inferences about subsurface elastic properties from seismic data and well-log measurements. The theoretical framework I have adopted is that of Bayesian inference, in which information about the subsurface is quantified in term of probabilities. By doing so, the uncertainties associated with the inverse calculation due to noise in the data and due to possible subsurface models are formally incorporated into the calculation. Those probabilities are combined to derive a final a posteriori probability distribution on the space of models, which is used to extract confidence measures on the subsurface-parameter estimates.

In this work, the uncertainties in the data were classified as observational and modeling errors (Tarantola, 1987), and have been characterized by the following Gaussian probability distributions (Equation (2.14) repeated here)

\[
\rho_d(d) \propto \exp \left[ (d - d_{\text{obs}})^T C_{\text{OBS}}^{-1} (d - d_{\text{obs}}) \right],
\]

\[
\Theta(d|m) \propto \exp \left[ (d - g(m))^T C_{\text{MOD}}^{-1} (d - g(m)) \right].
\]

(6.1)

In this approximation, observational and modeling errors are quantified by the covariance matrices of the probability distributions \(\rho_d(d)\) and \(\Theta(d|m)\), respectively. Those matrices are defined by the expected value of the fluctuations about the means of those distributions, due to the specific noise in consideration. The means are defined to be the recorded seismic data for \(\rho_d(d)\) and the synthetic data generated for the subsurface model derived from the well-log measurements (Figure 5.8) for \(\Theta(d|m)\). Note that, strictly speaking, the modeling uncertainties characterized by \(\Theta(d|m)\) should be independent of the model itself (Tarantola, 1987). Therefore, the use of the synthetic data generated from the well log as the mean for this distribution is in fact an approximation, considered reasonable because the well log is representative of the subsurface at Sorrento Field. A more rigorous way to quantify the uncertainties associated with modeling errors would be to estimate the covariance matrix \(C_{\text{MOD}}\) for an ensemble of subsurface models, and select the one that yields the largest uncertainty. This approach provides an upper bound on the modeling uncertainties that could have been used in the inverse calculation. For practical purposes, I have used instead the covariance matrix associated with modeling errors estimated for a typical model, defined by the well-log measurements.

The observational uncertainties quantified in this study were related to ambient noise and time distortions on the data due to near-surface heterogeneities. The first
0.1 s of the Sorrento data were used for the estimation of the data covariance matrix associated with ambient noise. Data uncertainties due to near-surface heterogeneities were quantified from data fluctuations about the recorded data induced by corrupting these data with artificial random residual statics. Those random statics were limited to ±8 ms and were obtained from an uniform random generator.

The modeling errors taken into consideration were associated with the scaling factor used to calibrate the synthetic data with respect to the field data, and with the discretization level adopted in the calculation. For the scaling factor I estimated a covariance matrix from data residuals obtained by the difference of the synthetic data scaled by an optimum factor and the same data, scaled by different factors. The factors were chosen to be 0.80, 0.85, 0.90 and 0.95 times the optimum one. Uncertainties associated with model discretization errors were obtained by deriving a covariance matrix from the data difference between the synthetic data generated from the well log discretized at the level used in the inversion, and the synthetic data generated from the full well log.

By adding all the covariance matrices derived from the specific types of noise considered in the calculation I have obtained a quantitative measure of the uncertainties in the data. Several assumptions have been made though, in the estimation of each one of the covariance matrices, which will, of course, influence the \textit{a posteriori} assessment of the uncertainties about the subsurface-parameter estimates. Those assumptions are as follows:

- The window of .1 s used to characterize the ambient noise is free of source-generated energy, and truly represents realizations of this type of noise.

- The distortions imposed by the near-surface on the recorded data can be modeled by surface consistent residual statics. Thus, errors on the estimation of residual statics, assumed to be limited to ±8 ms, were used to quantify the covariance matrix associated with uncertainties on the data due to near-surface heterogeneities. Since the magnitude of the residual statics in the Sorrento data are in the order of 20 ms, 8 ms represents a maximum error of 40% in the residual-static-estimation procedure.

- The magnitude of the error in the scale factor is limited to 20% of its optimum value.

- The full well log measurements can be used to quantify the errors associated with the discretization employed in the calculation.

- All noise seismograms, the construction of which was discussed in Section 5.4, represent realizations of the underlying distributions \( \rho_d(d) \) and \( \Theta(d|m) \) that model the specific noise under consideration.

The \textit{a priori} distribution \( \rho(m) \) incorporates information about the fine layering of the subsurface derived from a well log recorded in a borehole near the locations
where the seismic data were acquired. A Gaussian approximation was used for this
distribution, and its mean and covariance matrix were computed from the well-log
measurements. The mean (referred to as $\mathbf{m}_{\text{prior}}$) consists of a long-wavelength depth
component of the well-log measurements, and is used to incorporate the kinematics of
the seismic-data reflection events into the inverse calculation. Because I am using local
optimization routines in the computation of the MAP models, I am not attempting
to estimate the long-wavelength depth component of the subsurface. Under an optimization framework it is well known that global methods, like genetic algorithms and
simulated annealing, are required for this purpose.

The motivation to use such an $a$ priori probability is to incorporate information
about the subsurface parameters from a source independent of the seismic data, which
was taken to be the well log. The implicit assumption made in the construction of this probability is that the well-log fluctuations can be modeled as an ergodic
process. Thus, one single realization, i.e., the well-log fluctuations, of the underlying
probability suffices to estimate its covariance matrix. If more well logs were used in
the estimation procedure, the ergodicity assumption could be relaxed. This, however,
was not attempted in this study. Furthermore, note that the same $a$ priori distribution
was used in the inversion of all Sorrento data. Therefore, the hypothesis that the well
log is representative of the subsurface at the location where the seismic data were
acquired, is also made in this work.

Once the $a$ priori data and model uncertainties are defined, they are used to con-
struct the $a$ posteriori distribution. As seen, this $a$ posteriori was approximated by a
Gaussian distribution about its maximizer (the MAP model), which was computed by
a local optimization technique. Confidence measures about the subsurface estimates,
such as unit-standard deviation intervals, pseudo-random realizations of the subsur-
face and marginal distributions, were obtained from this Gaussian approximation of
the $a$ posteriori distributions. Without such an approximation Monte Carlo sampling
techniques would have to be used to estimate these confidence measures. The use
of those techniques is likely to be impractical for the problem of multi-offset seismic
waveform inversion for the foreseeable future due to their high computational cost.
Such Monte-Carlo based Bayesian inversion schemes have been applied to zero-offset
seismic waveform data (Scales and Tarantola 1994 and Mosegaard et al. 1996) and to
gravity data (Mosegaard and Tarantola, 1995).

The results obtained from the proposed methodology when applied to the Sor-
rento data indicate a fairly sizeable amount of uncertainty in the parameter estimates.
Specifically, $P$-wave impedance was the best resolved parameter, followed by $S$-wave
impedance and then density. In fact, very little was learned about $S$-wave impedance
and nothing was learned about density from the seismic data with respect to the $a$

priori probability distribution. Therefore, the $a$ posteriori distribution provided wide
confidence intervals on the subsurface parameters. Such large ambiguity in the data
is also verified by carrying out the seismic inversion under the Occam methodology,
in which the result of this calculation is the smoothest subsurface model that still fits
the data to a pre-specified accuracy. The Occam inversion was applied to one of the

seismic gathers of the Sorrento data and the results were contrasted with the ones provided by the Bayesian method. Occam and Bayes MAP models are quite different and are shown in Figure 5.49. The a priori distribution brought features from the well-log measurements into the Bayes MAP model. Such features are more significant at depths around 0.8 km for the P-wave, S-wave and density profiles. Specifically, for the S-wave impedance impedance and density profiles, significant features at shallower depths (around 0.5 km) present in the MAP model cannot be seen in the Occam model. This large difference in these two models is consistent with the larger degree of uncertainty associated with the inverse calculation, quantified by the a posteriori probability distribution. Such a large spectrum of possible subsurface models must be taken into consideration by the seismic interpreter when deriving information about the subsurface from seismic data. Additional data or information has to be brought into the inverse calculation so that the subsurface-parameter estimates can be better constrained.

The magnitude of the resolution and uncertainty measures derived in the approach of Bayesian inversion are of course subject to the assumptions made in quantifying the a priori uncertainties. Strictly speaking, a sensitivity analysis of the a posteriori uncertainties to the choices made in the construction of data and model covariances should be carried out. This has not been attempted in this work since is not clear how to accomplish this without having to redo the inversion calculation for different data and model covariance matrices. Furthermore, not all uncertainties have been considered in the problem, such as the influence of anisotropy and structural complexity on the seismic data. Nonetheless, the uncertainty analysis carried out here indicates the large ambiguity of the seismic data used in the inverse calculation, which is further illustrated by the differences in the subsurface models resulting from the Bayesian and Occam methodologies. This is the most important and new contribution of this work.

The works of Mora (1987) and Crase et al. (1990) among others, are examples of Bayesian calculations in which Gaussian probabilities were also used. Their covariance matrices $C_D$ and $C_M$, however, have played the role of weighting and regularization devices. Therefore, data and model uncertainties have not been properly considered in such calculations and, as a consequence, the significance of the a posteriori probability distribution and all the uncertainty statements derived from it are lost. The key contribution of this dissertation has been to make a serious attempt to quantify the uncertainties in a seismic waveform inversion problem, including uncertainties in the data and in the forward modeling procedure, and in the subsurface model parameter via a a priori probability derived from the well-log measurements. Moreover, I have analyzed a Gaussian approximation to the a posteriori distribution to derive resolution statements about the subsurface. This distribution can be used to address questions posed of the models, such as “What is the probability that the P-wave impedance at a given depth is within a certain range” or “What is the probability that a given layer has a certain thickness?”. The precise questions should be strongly influenced by the needs of the interpreter. What I have done in this work is to provide an efficient and rigorous framework for answering those questions.

128
6.1 Computational cost

The estimation of the MAP models is the most computationally intensive step in the Bayesian inverse calculation. This is due to the computation of the synthetic seismograms and the Frechét derivatives required in each iteration of the conjugate-gradient-based optimization procedure.

The inversion algorithm was designed to run on a multi-processor, share- or distributed-memory, computer environment. The code was thoroughly tested on a 4-processor SGI Power Challenge and on network of Pentium-based workstations.

The inverse calculations for the 300 unknowns of the Sorrento data were mainly executed using the Silicon Graphics machine. On this platform, the elapsed time required by each iteration of the conjugate-gradient optimization of a single MAP model is about 20 min \(^1\). No more than 12 iterations were required in the estimation of the MAP models for the Sorrento data. Such cost can, however, be further reduced by using more sophisticated optimization techniques such as a pre-conditioned Newton’s method (Dennis and Schnabel, 1987), which in principle provides faster convergence rates than the conjugate-gradient method. Furthermore, the availability of more powerful computational resources may make the application of inversion techniques as the one proposed in this study cost effective in practical situations.

6.2 Future work

I consider the following three issues the most important ones to be addressed in future research.

- The data uncertainties were limited to the errors in the data due to ambient noise and near-surface heterogeneities, and to the errors in the forward modeling procedure due to the scale factor between synthetic and field seismic data and due to model discretization errors.

- The subsurface was parameterized as a one-dimensional model.

- Data and model uncertainties have been modeled by Gaussian probabilities. Furthermore, the \textit{a posteriori} distribution was approximated by a Gaussian distribution about the model at which that distribution is maximum.

Each one of those issues is discussed next.

6.2.1 Data uncertainties

In this work data uncertainties were quantified by taking into consideration some of the data errors and some of the limitations in the forward modeling procedure. I have

\(^1\)This elapsed time was obtained with the computer fully dedicated to the inversion of the Sorrento data.
judged the specific sources of errors considered here relevant to the inverse calculation.
Moreover, the procedures used to quantify those errors were developed such that the
proposed methodology can still be applied to practical problems. It is clear though
that further research is required in the topic of quantifying data uncertainties. A
more thorough analysis on the identity of the important sources of data errors for the
inversion and how to assess them in a quantitative manner are important research
issues that have to be further investigated so that the Bayesian methodology can
practically (and formally) be applied to geophysical inverse problems.

6.2.2 Two- and Three-dimensional models

To allow a parameterization of the subsurface such that lateral heterogeneities in
two or three dimensions can be considered, more general forward modeling opera-
tors than the reflectivity method must be used. In the situation of minor geological
complexity, in which the ray-theoretical assumptions are valid, dynamic ray-tracing
is likely to be a good choice for the forward modeling operator. However, under
more complicated geological scenarios, finite-difference algorithms seem to be the only
method of choice. Obviously these are more general forward modeling operators than
the reflectivity method, and the fact that the number of unknowns in the inverse
problem increase exponentially with the number of dimensions considerably raise the
computational cost of the procedure proposed here.

Therefore, to extend this approach to more general subsurface models and still
make it applicable to field data, the inversion should not only concentrate on target
areas of the subsurface, thus reducing the dimensionality of the problem, but it should
also to resort to additional geological constraints as a priori information than a single
well log, to narrow the spectrum of possible subsurface models that can be considered
solutions to the inverse calculation. Those geological constraints can be in principle
derived from sources such as geostatistical stochastic simulations using spatial well-
log data or high-resolution stratigraphic correlations derived from multiple well-logs,
such as the one described in Cross et al. (1993). The incorporation of this type of
information has not yet been explored in a seismic waveform inverse calculation here,
or elsewhere, and is an open and exciting research topic.

6.2.3 Breaking through the Gaussian hypothesis

In this study seismic data-independent information about the subsurface was
obtained from well-log measurements and incorporated into the calculation via the
Gaussian a priori probability distribution $\rho(m)$. The probabilities used to charac-
terize modeling and data errors were also modeled by Gaussian distributions. By
making a Gaussian assumption for these probabilities, only moments up to the second
order of the underlying process are being considered in the inversion procedure. Thus,
higher-order moment information is neglected. The principle of maximum entropy de-
scribed in Chapter 3 can be used for constructing non-Gaussian a priori distributions
from sample data. An algorithm to that end is proposed in that chapter, however,
its sensitivity to errors in the estimation of statistical moments is too large for the
algorithm to be considered useful. Further research on more accurate moment estima-
tion techniques is required for the use of the maximum-entropy algorithm proposed
in this dissertation.

In addition to that, even if non-Gaussian probabilities are used in the modeling
of the data and model uncertainties, the analysis of the resulting (non-Gaussian) a
posteriori probability to assess the uncertainties associated with the inverse calculation
is not a trivial matter. Non-Gaussian probabilities require Monte Carlo sampling
techniques to provide confidence measures on the subsurface parameters (Mosegaard
and Tarantola, 1995) which, as mentioned several times in this dissertation, is not
practical for inversion of seismic-waveform multi-offset data.

The conclusion is that non-Gaussian Bayesian multidimensional inversion, in
which the non-Gaussianity can be justified from the data, is likely to be impractical
for seismic field data for the foreseeable future. This is due not only to its high
computational cost, but also due to the lack of theoretical and practical tools required
to construct multidimensional non-Gaussian probabilities from sample data.

6.3 Developed software

Several codes have been developed as part of this work. The source codes
and installation instructions can be downloaded from the CWP Linux WWW server
(http://www.cwp.Mines.edu/cwpcodes). All software was developed in ANSI-C, and
the ones that run in a distributed fashion use the PVM library version 3.3.x. The
main programs are the following:

• stratinv: Main inversion algorithm. It consists of a parallel implementation
  of the optimization procedure used in the minimization of the objective func-
  tion (4.2). Covariance or regularization matrices can be used in this approach.

• covadap: Implements estimation of the covariance matrix from well logs, using
  the technique described in Chapter 3.

• dataCovar: Computes a sample data covariance matrix from an ensemble of
  seismograms.

• svdutils: Implements several calculations that can be done with singular value
decomposition (SVD): solving a linear system by SVD truncation, computation
of resolution matrices and generalized inversion. It also computes the bias of
the MAP and Occam estimators as described in Gouveia and Scales (1996b).

• posteriori: Computes the a posteriori covariance matrix.

• sample: Samples a Gaussian distribution via the LU decomposition technique.

• logBlock: Blocks the well log with a weighted average based procedure.
W.P. Gouveia, Jr.

- **sudrefsvmaster**: Parallel elastic reflectivity modeling.
REFERENCES


W.P. Gouveia, Jr.


Multidimensional Bayesian Inversion


Mosegaard, K., Singh, S., Snyder, D. and Wagner, H. 1996 Monte Carlo analysis of seismic reflections from Moho and the W-reflector. Submitted to *J. Geophysical Research.*


W.P. Gouveia, Jr.


Yilmaz, O. 1987. *Seismic data processing* SEG.
Appendix A

THE FRECHÉT DERIVATIVES OF THE DISPLACEMENT FIELD

The Frechét derivatives, being the derivatives of the forward modeling operator with respect to the medium parameters, are of primary importance for an inversion procedure that relies on optimization techniques. In this study, numerical computation of these derivatives is unacceptable due to the high computational costs involved. For instance, the numerical differentiation of the displacement field in the Sorrento data demands 300 forward modeling procedures, one for each subsurface parameter. Thus, the need for an analytic differentiation technique.

Tarantola (1987) proposed one analytic approach that, based on the Born approximation of the displacement field, casts the computation of the Frechét derivatives as a forward modeling procedure with specific initial and boundary conditions. This was the method used in Mora (1987) and in Crase et al. (1990) in a finite-difference based inversion. Dietrich and Kormendi (1990) applied Tarantola’s ideas to the reflectivity method used in an elastic stratified medium inversion, the same type of medium studied in this work. They also investigated the accuracy of the Born approach and found out that it provides unsatisfactory results when the thickness of the layer at which the differentiation is being computed is larger than one wavelength of the wavefield.

Those limitations are not a hindrance for the application of this method to the Bayesian inversion procedure carried out here since the parameterization of the subsurface is done with layers of thickness below the seismic wavelength. Nonetheless, I propose an alternative approach for the computation of these derivatives, by just analytic differentiation of the reflectivity integral without resorting to any kind of approximation. This yields a theoretically more accurate differentiation of the displacement field, though more expensive than Tarantola’s approach.

Before describing the approach, I shall introduce the notation, which is close to the one adopted in Müller (1985). Consider a stack of \( n \) layers, each one described by its \( P \)-wave impedance, \( S \)-wave impedance, density and thickness. The interface \( j + 1 \) separates the layer \( j \) from the deeper layer \( j + 1 \). The reflectivity matrix at the bottom of layer \( j \), \( MB_j \), consists of the ratios of the amplitudes of the upgoing and downgoing waves at this depth. Considering only the \( P \) and \( SV \) propagation modes (since for the far field the \( SH \) mode is decoupled), \( MB_j \) is a 2X2 matrix which elements are

\[
MB_j = \begin{bmatrix}
P-wave_u & S-wave_u \\
P-wave_d & S-wave_d
\end{bmatrix},
\]

where the superscripts \( u \) and \( d \) represent the upgoing and downgoing wavefield, respectively. Along the same lines, the reflectivity matrix at the top of the layer \( j \), \( MT_j \),

\[137\]
is given by the ratio of the upgoing and downgoing wavefields at this level.

As shown in Müller (1985), the displacement field at the surface is basically a function of $MB_0$, the reflectivity matrix at the bottom of the first layer. The term “basically” is used here because the near-surface boundary conditions and the thickness of the first layer obviously enter in the calculation of the displacement components. Since in this inversion work the parameters of the first layer are kept fixed, differentiation of the wavefield itself amounts to differentiation of the reflectivity matrix $MB_0$. This matrix is computed in a recursive manner according to the formula (Müller, 1985)

$$MB_j = R_{j+1}^d + T_{j+1}^u \left[I - E_{j+1} MB_{j+1} E_{j+1} R_{j+1}^u \right]^{-1} E_{j+1} MB_{j+1} E_{j+1} T_{j+1}^d.$$  (A.2)

The significance of the matrices in this expression is given below:

- $R_{j+1}^u$: Reflection coefficient matrix for interface $j + 1$ when the incident wave is an upgoing plane wave traveling in layer $j + 1$.
- $R_{j+1}^d$: Reflection coefficient matrix for interface $j + 1$ when the incident wave is a downgoing plane wave traveling in layer $j$.
- $T_{j+1}^u$: Transmission coefficient matrix for interface $j + 1$ when the incident wave is an upgoing plane wave traveling in layer $j + 1$.
- $T_{j+1}^d$: Transmission coefficient matrix for interface $j + 1$ when the incident wave is a downgoing plane wave traveling in layer $j$.
- $E_{j+1}$: A phase shift matrix that takes into account the plane-wave propagation within layer $j + 1$.
- $I$: The 2X2 identity matrix.

The dependence of the displacement field on the medium parameters (elastic impedances and densities) is given by the reflection and transmission coefficient matrices. The procedure developed here to compute the Frechét derivatives provides a recursive expression similar to the one in (A.2), but now for the derivatives of $MB_j$ with respect to any parameter of the medium, $\partial MB_j/\partial \text{parameter}_j$, where $\text{parameter}_j$ is either the $P$-wave impedance, $S$-wave impedance or density of the $j$–th layer. Inspection of Equation (A.2) shows that $MB_j$ is dependent only on the parameters for layers $j$, $j + 1$ and $j + 2$. Thus, it is possible to derive the following expressions:
\[
\frac{\partial MB_j}{\partial \text{parameter}_{j+1}} = \frac{\partial R^d_{j+1}}{\partial \text{parameter}_{j+1}} \\
+ \frac{\partial T^u_{j+1}}{\partial \text{parameter}_{j+1}} H^{-1} MT_{j+1} T^d_{j+1} \\
+ T^u_{j+1} H^{-1} \left[ MT_{j+1} \frac{\partial R^u_{j+1}}{\partial \text{parameter}_{j+1}} \right] H^{-1} MT_{j+1} T^d_{j+1} \\
+ T^u_{j+1} H^{-1} \left[ MT_{j+1} \frac{\partial T^d_{j+1}}{\partial \text{parameter}_{j+1}} \right].
\] (A.3)

\[
\frac{\partial MB_j}{\partial \text{parameter}_{j+2}} = -T^u_{j+1} H^{-1} \frac{\partial}{\partial \text{parameter}_{j+2}} \left[ I - MT_{j+1} R^u_{j+1} \right] H^{-1} MT_{j+1} T^d_{j+1} \\
+ T^u_{j+1} H^{-1} \frac{\partial MT_{j+1}}{\partial \text{parameter}_{j+2}} T^d_{j+1}.
\] (A.4)

Where:

\[
H = \left[ I - E_{j+1} MB_{j+1} E_{j+1} R^u_{j+1} \right],
\]
\[
MT_j = E_j MB_j E_j.
\] (A.6)

Differentiation of the reflection and transmission coefficient matrices was performed analytically with the Mathematica software package. The derivation of Equations (A.5) is cumbersome, but straightforward. An important result from linear algebra used here was the differentiation of the inverse of a matrix \( A \) with respect to a parameter \( \alpha \), given by (Golub and Van Loan, 1989)

\[
\frac{\partial A(\alpha)^{-1}}{\partial \alpha} = -A(\alpha)^{-1} \frac{\partial A(\alpha)}{\partial \alpha} A(\alpha)^{-1}.
\] (A.7)

Equations (A.5) were coded as part of the gradient calculation of the objective function (4.2). Moreover these expressions were also used, in a separate code, to study the
resolution of the forward modeling operator for a given geological model. This is illustrated in Figure A.1. The top of this figure displays the $P$-wave impedance Frechét derivatives for the synthetic model studied in Chapter 4 (Figure 4.4) at the MAP model (Figure 4.7). Each of the 50 gathers represents the derivative of the vertical component of displacement with respect to a perturbation of the $P$-wave impedance for a given layer. Larger amplitudes in these gathers is indicative that data resolution will be larger for a specific subsurface parameter. These figures show that the resolution does not necessarily decrease monotonically with depth for this specific model. In fact lower amplitudes in these gathers correlate with larger values of the velocities in the model shown in Figure 4.4, e.g. around depths 1.45 km and 1.28 km, which are associated with layer indexes of 22 and 28, respectively. This is expected because the wavelength of the wavefield is larger at these depths due to the higher medium velocities. In the bottom of Figure A.1 the $P$-wave impedance Frechét derivatives are computed for a subsurface model that parameters linearly increase with depth. In this case the resolution of the forward modeling operator decreases monotonically with depth.
Fig. A.1. $P$-wave Frechét derivatives (a) for the subsurface model illustrated in Figure 4.4, and (b) for a subsurface model which parameters increase linearly with depth.
W.P. Gouveia, Jr.
Appendix B

DISTRIBUTED-COMPUTER IMPLEMENTATION

The inversion algorithm used here employs as the forward modeling procedure the reflectivity method of Fuchs and Müller (1971). In this procedure the displacement field is computed in the frequency domain. Since frequency components of the wavefield are independent of each other this algorithm allows a high level of parallelism. The same applies to the computation of the Frechét derivatives of the displacement field, also computed in the frequency domain.

The inversion code developed as part of this research has the modeling and the Frechét derivative computation steps implemented in a distributed fashion. This implementation is based on the simple idea of a master processor assigning frequency partitions to slave processors. This is illustrated in Figure B.1. As the slave processors finish the computation of their respective frequency partitions, they send the results back to the master processor, which transforms the data to the time domain. The code is implemented in standard ANSI C, and all the message-passing and synchronization routines are provided by the PVM library.

The implementation depicted in Figure B.1 is probably close to optimum for a homogeneous network of computers, on which the only running program is the distributed reflectivity algorithm. This is not usually the case. The typical scenario is to have computers with distinct computational power and with different levels of CPU utilization connected to the same network. This characterizes a heterogeneous computer environment. In this situation some strategy of load balancing is required in order to allocate more work to computers of higher computational capacity. The load balancing used here is simple, but it provided satisfactory results. It consists of splitting the total job into small tasks, the size of which (i.e., the number of frequencies to be computed) is specified by the user. Once the slave processors finish their computations for a specific frequency partition, they return their data to the master processor. If there are still more frequency components to be computed the master processor sends an additional task to the slave processor. This procedure is repeated until no more tasks are left. In this implementation, more powerful machines will tend to receive more tasks than do less powerful ones. Figure B.2 illustrates the sending of more frequency components to the more powerful machines.
W.P. Gouveia, Jr.

Fig. B.1. The distributed reflectivity algorithm. $[\omega]$ indicates the frequency interval assigned to each slave processor by the master. The “ball” represents the frequency components of the displacements or Frechét derivatives computed by the slave processors.

An important feature of the inversion code is that it is fully portable across different computer systems. The code was thoroughly tested on a network of LINUX Pentium-based workstations, on a SGI Power Challenge with four processors and on a network of IBM RS/6000s.
FIG. B.2. Illustration of the load-balancing strategy of the distributed implementation.
Appendix C

APPENDIX C: LINEARIZING THE INVERSE PROBLEM

In this appendix I solve the inverse problem described in Chapter 4 by linearizing the forward modeling operator about \( m_{\text{prior}} \). This results in a linear system of equations that is solved iteratively by a conjugate-gradient technique. The stopping criterion used in the calculation is defined in Equation (4.5), but now the linearized forward modeling operator is used in this expression instead of \( g(m) \).

C.1 Bayesian Inversion

The linearization of the forward modeling operator \( g(m) \) about the model \( m_{\text{prior}} \) results in a quadratic approximation for the objective function (4.2). The minimizer of this quadratic objective function can be found by solving the following linear system of equations

\[
(G^T C_D^{-1} G + C_M^{-1}) \delta m = G^T C_D^{-1} (g(m_{\text{prior}}) - d_{\text{obs}}),
\]

\[
m_{\text{map}} = m_{\text{prior}} + \delta m. \tag{C.1}
\]

\( G \) is the forward modeling operator \( g(m) \) linearized about \( m_{\text{prior}} \), and \( \delta m \) are the sought model perturbations. I used a standard conjugate-gradient procedure to solve this weighted, regularized linear system of equations. The result, initial guess \( m_{\text{prior}} \) and true model are shown in Figure 4.7. Comparison with the model obtained in the nonlinear approach (Figure 4.7), shows that the linearized result is considerably poorer in accuracy. Although a direct comparison is obviously not fair due to the large differences in computational cost of both procedures, the magnitude of the errors in the linearized solution makes this model not acceptable as a estimate of the subsurface. In this particular problem the nonlinear implementation costs approximately ten times more than does the linearized version. The main objective here is to illustrate the higher accuracy in the results provided by the nonlinear approach to the problem of extracting impedance and density information from seismic amplitudes.

C.1.1 Occam inversion

The linear system obtained in the Occam calculation has a form similar to that obtained in the Bayesian case. It can be written as

\[
(G^T C_D^{-1} G + \lambda R^T R) \delta m = G^T C_D^{-1} (g(m_{\text{prior}}) - d_{\text{obs}}),
\]

\[
m_{\text{occam}} = m_{\text{prior}} + \delta m. \tag{C.2}
\]

\[147\]
Fig. C.1. The MAP model estimated from the data in Figure 4.3 when linearization about the model $m_{prior}$ is performed. The true model and the initial guess are also displayed.
As done in the nonlinear calculation, the above linear system is solved for increasing values of the regularization parameter $\lambda$. $m_{\text{occam}}$ is the model associated with the largest regularization parameter that still fits the data in accordance with the stopping criterion (4.5). If this model, shown in Figure C.2, is compared with the one obtained in the nonlinear version of the Occam procedure (Figure 4.16), we see that the linearization errors are not negligible. As in the Bayesian case, the model obtained with the nonlinear optimization (Figure 4.16) is far superior to that the one obtained with the linearized inversion.
Figure C.2. The Occam’s model estimated from the data in Figure 4.3, using a linearized, regularized inversion procedure. The true model and the initial guess are also displayed.
Appendix D

APPENDIX D: RESOLUTION OF THE FORWARD MODELING OPERATOR

Here, I present a singular value decomposition (SVD) analysis of the forward modeling operator $g(m)$. Let $G = Q\Sigma V^T$ be the SVD decomposition of the linearized operator $G$ about some model, where $Q$ and $V$ are matrices whose columns are the eigenvectors of $GG^T$ and $G^T G$, respectively. $\Sigma$ is a diagonal matrix whose diagonal elements are the eigenvalues of $GG^T$ (Golub and Van Loan, 1989). Although any subsurface model would do for the linearization of the forward modeling operator $G$, I chose the model $m_{\text{map}}$ displayed in Figure 4.7. Considering that we have a linear system $Gm = d$, the solution model $m$ can be represented as

$$m = \sum_{i=1}^{N} \frac{q_i \cdot d}{\Sigma_{ii}} v_i,$$

where $q_i$ and $v_i$ are the $i$-th columns of the matrices $Q$ and $V$, respectively. $\Sigma_i$ is the $i$-th singular value of $G$. In general model eigenvectors $v_i$ associated with large values of $\sigma_i$ are well resolved. Figure D.1 displays the 150 model eigenvectors of $G$. This figure shows in its upper part the normalized singular values and, in its lower part, the coefficients $\frac{q_i \cdot d}{\Sigma_{ii}}$ of the expansion (D.1). In order to form the solution model $m$ in Equation (D.1), model eigenvectors with their respective weight are added until the superposition fits the data to within one standard deviation. The cutoff is indicated by the arrow in Figure D.1.

Just by taking into consideration the model eigenvectors and the singular values displayed in Figure D.1, the SVD analysis indicates that for both impedance and density profiles, long-wavelength (smoother) components can be in general better resolved than can short-wavelength (rougher) components from seismic amplitudes. This is shown more clearly by the curves plotted beside the spectrum of the eigenvectors, in which their roughness (quantified by the module of the vector obtained from dotting the operator $R$ defined in Equation (4.10) into the respective eigenvector) is plotted as a function of their number within the spectrum. This analysis should, however, be coupled with the coefficients displayed in the bottom part of Figure D.1. Notice that the first 50 eigenvectors of the spectrum make a very small contribution (due to the small value of their coefficients) to expansion (D.1). Therefore, this portion of the parameter spectrum is not well resolved by the data, though model eigenvectors are associated with relatively large singular values.

This SVD analysis indicates that, for the specific problem under consideration, the model eigenvectors within the interval limited by the indexes $[50, 90]$ of the eigenvector
W.P. Gouveia, Jr.

spectrum are the best resolved ones. According to the roughness plots, those represent long-wavelength depth components for P-wave and S-wave impedances, and small-wavelength depth components for density. The terms long and small are used here relative to the largest depth wavelength (or smallest roughness) shown by the spectrum of eigenvectors. Note, however, that such a distinction in terms of wavelength is not so clear from the SVD analysis, especially for density, since the roughness, although generally increasing with the eigenvector number within the spectrum, presents large fluctuations about this trend.
Fig. D.1. Singular Value Decomposition of the elastic waveform inversion. The arrow indicates the last model eigenvectors incorporated in the solution such that the data is fit to within one standard deviation of the noise. For each of the parameters the model eigenvectors are normalized to unit. Note that the singular values are plotted in a linear-log scale, and that the horizontal scale for the roughness plot is the same for all parameters.