Bayesian Methods applied to Reflection Seismology

Camila Carneiro da Silva Caiado

A Thesis presented for the degree of
Doctor of Philosophy

Department of Mathematical Sciences
University of Durham
England

February 2012

Dedicated to

my parents.

Bayesian Methods applied to Reflection Seismology

Camila Carneiro da Silva Caiado

Submitted for the degree of Doctor of Philosophy
October 2011

Abstract

Quantifying uncertainty in models derived from observed seismic data is an important issue in exploration geophysics. In this research we examine the geological structure of the subsurface of the Earth using controlled source seismology which consists of data recorded in time and the distance between acoustic sources and receivers. There are a number of inversion tools to map data into depth models, but a full exploration of the uncertainty of such models is rarely done because of the lack of robust strategies available for the analysis of large non-linear complex systems.

In reflection seismology, there are three principal sources of uncertainty: the first comes from the input data which is noisy and band-limited, the second is from the modeling assumptions used to approximate the physics of the problem in order to make the problem tractable, and the last is from the ambiguity in data and model selection. The latter is by far the hardest source of uncertainty to assess, not only are there a large number of models which are appropriate for a given seismic profile and still physically and geologically plausible, but also the judgement related to the acceptability of a model varies according to the expert handling the data. The fact that there are many possible solutions, depending on how the problem is treated, adds a new layer of uncertainty to the question.

Here we propose a Bayesian approach to assess the uncertainty in velocity models derived from seismic reflection data. We have developed a method used to identify and track seismic events called the Seismic Event Tracking algorithm. We then

created the BRAINS (Bayesian Regression Analysis in Seismology) class of models used to estimate velocities, travel times and depths with associated measures of uncertainty for each identified horizon. Since the experts' prior judgements and problem requirements vary according to the situation being analysed, the Bayesian methodology is the most appropriate to create a gray box that accepts the input of prior knowledge but that is also able to cope with vague or no prior information; here each model in the BRAINS class can be used at different stages of seismic processing, depending on the inputs necessary for the next step of modeling. Moreover, each estimate produced has an uncertainty model attached that can be explored before making a decision.

In order to investigate the robustness of the models proposed, we analysed a series of single and multigathered synthetic examples, some of which had attributes that differ from the modeling assumptions or carried ambiguities derived from the limitations of data recording. Finally, we analysed a 2D real data set part of a seismic survey acquired over the Naturaliste Plateau and Mentelle Basins off the south west coast of Australia. We show the efficiency of the BRAINS approach on real data and recover velocity and depth models with posterior depth standard errors of at most 0.4% relative to posterior depth means, and posterior RMS velocity standard errors of at most 1.7% relative of posterior RMS velocity means. We also observe that variations in interval velocities is higher with an average of 2.4% for the posterior interval velocity standard deviation and mean ratio which reaches a maximum of 23.7% in areas of high uncertainty.

Declaration

The work in this thesis is based on research carried out at the Department of Mathematical Sciences, Durham University, UK. No part of this thesis has been submitted elsewhere for any other degree or qualification and it all my own work unless referenced to the contrary in the text.

Acknowledgements

Here we have the stuff that nobody will read unless this piece of work becomes somewhat famous or someone looking for inspiration on what to write in their acknowledgements section stumble on it.

I could thank an extremely large number of people that supported me since birth but, in order to select only the significant ones, I would have to create a rating index to classify them according to their helpfulness level taking into account any pain they might have caused. Let's call this index the Friend rating index of (un)expected deeds and take this measure to be in the interval $[-1, 1]$ implying that the higher the score, the better, and I shall only mention a handful of high scorers.

For the uncountable hours of discussions in Bayesian analysis and seismology, I gratefully acknowledge my supervisors Michael Goldstein and Richard Hobbs. I thank Michael for keeping me in the Bayesian side and away from the darker and normalized frequentist world, and I thank Richard for the many hours spent clarifying the basics of seismology, simulating data, providing expert information and being an accessible simulator. I'm incredibly grateful to WesternGeco for funding this project and understanding that uncertainty in seismic data analysis exists and has to be dealt with. I would like to thank Rachel Chester who spent a month during the summer of 2011 learning how to use our models to investigate a real dataset. I also would like to thank Geoscience Australia for making the seismic survey used in our real data examples available.

During these past four years in Durham, few people were consistently helpful and understanding, namely, Walia Kani, Ian Davison, Angela Connor, Ashleigh Luke and Adrienne Tinker née Halpin. I'd like to thank them all for being great friends and for all the caffeine supplied. In that case, I should probably thank all

coffee shops in Durham as well, but never mind.

The highest scorer and the most persevering by far has to be my mother who tolerated me for 26 years and never gave up. The second highest scorer is my father who lost this battle a few years ago but was consistently fighting for the top of the chart until the end. These are the two variables to whom I dedicated this work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Uncertainty is a term widely used in all academic fields but defined in a slightly different way in each of them. It can be related to doubt, lack of confidence, changeability or questionability; in physical sciences, it is more often related to the lack of knowledge, degrees of variation or measurability. The most appropriate placement of the term is given by Claude Shannon, also known as the father of information theory, who stated that "information is the resolution of uncertainty". Quantum mechanics, with theories like Heisenberg's uncertainty principle and Schrödinger's paradox, and finance, with risk analysis and market resilience, transformed uncertainty into a popular subject.

The need to understand uncertainty is accepted by most but there is not a standard way of measuring it. In computer sciences, it is usually assessed by calculating the amount of information in a given dataset using entropies, measures of divergence and other quantifiers of information. However, it is usually assumed that the data contains all information and, consequently, explains all uncertainty. In classical physics, uncertainty is reduced to observational error and modelling errors are ignored. In quantum mechanics, probabilities are associated to different possible outcomes resulting in a more flexible strategy to quantify the variability of a model behind an experiment.

These classical approaches are closely related to the frequentist line of thought in statistics. In frequentist inference, the aim is to calculate the odds that an event occurs or not relying on the fact that a certain experiment can be repeated an infinite

number of times. This line of thought fails to predict rare events and, due to the fact that only well-defined random experiments are acceptable, it cannot handle the uncertainty behind the definition of the model in itself.

The Bayesian view of probability, however, allows incomplete knowledge of a model by accepting that a perfect experiment is rarely possible and, to get closer to this perfect model, it includes all possible sources of information and weights them according to their plausibility. Unlike the frequentist framework, Bayesian inference permits a constant update of information by modifying their level of belief and the level of confidence in the primary hypothesis.

Bayes' theorem is the core of Bayesian statistics guided not only by its simplistic formula but mainly by its principle of information update. The theorem states that the posterior probability of a certain event A given an event B is proportional to the product of the prior probability of A and the likelihood of B given A. Therefore, an update of a prior belief has an immediate impact on the final model and the magnitude of the impact of this update depends on the quality of the data and its robustness.

Inference methods based on Bayes' rule are most commonly known for their applications to problems involving small datasets or rare events where an expert might have a strong prior belief. However, in the past two decades, with the development of mass computing, Bayesian methods are becoming more popular in environments where a large number of variables exist, not all variables are known, simulations are too costly or the model itself is incomplete.

In Geophysics, most of the problems related to seismic surveying meet at least one of the conditions above but with the bonus that data is extremely abundant with a high-level of redundancy given the current pace of seismic data collection. A lot of effort has been put on data collection and data analysis but the field lacks tools for uncertainty analysis. Like all other applied scientists, geophysicists have a vast amount of prior information about their data that cannot be measured through traditional methods. The fact that a piece of knowledge cannot be easily expressed as a numerical value or a mathematical equation does not imply that it cannot be incorporated in the model. Here we will concentrate on seismic reflection

surveying but all the principles presented throughout this work can be applied to other problems in geophysics and applied sciences. Now I describe the experiment explored in this thesis and briefly introduce its principles.

Seismic reflection surveying is a common method used to investigate the geological structure of the Earth's lithosphere. The experiment consists of seismic sources and receivers at or close to the surface organised in a space-time grid that is usually regular. The most used sources are impulsive, such as dynamite or air guns. The seismic waves produced are recorded by geophones which detect ground vibrations on land and hydrophones that detect pressure changes in water. In order to avoid overlapping and mixed signals, the sources are fired a set time and distance apart and the amplitudes and travel times of the seismic waves generated are recorded by an array of receivers. On detonation of the seismic source, the impulse propagates in all directions and, when a seismic wave reaches a boundary between two layers where there is a difference in acoustic impedance, a fraction of its energy is reflected and the remaining energy continues through the boundary; the behaviour of a seismic wave at a boundary depends on the angle of incidence and can be explained by the Zoeppritz equations (Yilmaz, 2002) which accounts for the amplitudes of primary and secondary waves at each side of the boundary. After detonation, the time that a wave takes to travel from the source and back to the receiver is recorded and a change on the recorded amplitude can be seen. This process is fairly similar to diagnostic ultrasonography apart from the dimensions of the structures being analysed; for the same reason acoustic waves of sonographic devices are not efficient at penetrating bones and gases, seismic waves may present a particular behaviour when they reach a reservoir, basalt or a salt body. Later we will discuss the behaviour of seismic waves through different materials and shapes and the limitations of seismic reflection surveying.

The aim of seismic reflection surveying is to map the geology beneath the target area. The quantity of data available is usually not a problem since streamers (e.g. an array of hydrophone receivers) are capable of collecting data with both high spatial and temporal sampling. The quality of the recorded data varies according to the area and resources available but industry standards have minimised many of the sources

of error, for example the advent of GPS for navigation has improved the positioning accuracy. Most of the noise found in state of the art datasets are directly related to environmental variables like weather and the geological structure itself. As in the classic Bayesian applications, we are also looking for rare intermittent events in time but, similar to the current Bayesian trends, data is vastly available. However, the data do not directly give information about the geology and, to overcome this limitation, the data need to undergo several stages of processing, modelling and interpretation which rely on prior information held by experts. A critical part of the analysis is the determination of the seismic velocity in the subsurface which, in its simplest form, relates the measured times to the depths of the geological interfaces.

Briefly, our goal is to explore the data collected through seismic surveys by using recorded times and amplitudes in a given grid position to construct a velocity and depth model of the local subsurface with a fast, statistically and computationally robust method that also produces an uncertainty analysis of the problem. In Chapter 2, there is a short introduction to wave and ray theory to better understand seismic waves. In Chapter 3, we present the velocity models and data recording methods followed by a discussion on Bayesian models and prior information and, in Chapter 4, we introduce a method to filter seismic data, the tools that will later be used to process the data are described and computational issues are discussed. In Chapters 5 and 6, synthetic examples are analysed followed by a real data example in Chapter 7. In Chapter 8, we compare our Bayesian strategies to the methods currently in use and list topics that should be explored further.

Given that this work involves two distinct scientific fields, after Chapter 8 we have a short glossary of geophysical terms followed by a notation table; in Appendix A, we have a brief introduction to probability and statistics including an introduction to Bayesian inference. Finally, the algorithms and the interactive graphics user interface can be found at designed for this thesis are available upon request through the author and the University of Durham.

Chapter 2

Waves and Ray Theory

In this Chapter we introduce the principles of reflection seismology, starting with waves and their properties. Understanding behaviour, assumptions and approximations made in physical models is necessary when building the velocity models we will discuss in the next Chapter. We start by defining waves and then describe their properties.

Waves are periodic variations in time and space whose motion usually transports energy but not matter; any such quantities have speeds that depend on the medium of propagation and satisfy the so-called Wave Equation, a second-order linear partial differential equation which will be presented later in this Chapter. See Whitham (1974) for more details on wave theory and Yilmaz (2002) for an introduction to seismic data analysis. When referring to media of propagation, waves can be classified as mechanical or vacuum waves. As their names imply, mechanical waves depend on matter to propagate while vacuum waves do not. In this work, we are only interested in mechanical waves, i.e. seismic waves, but most of the theorems discussed here also apply to vacuum waves like electromagnetic radiation. In fact, some of the principles explained here are more intuitive when applied to optics so we will use light rays to form an analogy with seismic rays for illustration.

2.1 Wave Function Properties

As a mathematical concept, a wave can be described using a wave function of its displacement from its equilibrium position in time. We will list the most relevant properties of this function and use Figure 2.1, which represent two sinusoidal wave functions, as a visual aid. First, we define the amplitude at a certain time as the absolute value of the wave function at that point. Occasionally we will refer to signed amplitudes when describing the position of a point below or above the axis of equilibrium. We also define peak amplitude as being the absolute value of a non-zero local maximum or minimum; in Figure 2.1, the time of the peak amplitude of the blue sinusoid is represented by the green dashed line.

A function is said to be periodic with period p if its graph is invariant under translation in the axis of equilibrium by a distance of p; in our graph in Figure 2.1, the period of the wave function represented by the blue trace is the distance from the point $(0, 0)$ to the point $(p, 0)$. Unless stated otherwise, we will use seconds (s) as a unit for period. The frequency is defined as the inverse of the period and represents the number of cycles per second of a wave function. It is usually measured in Hertz, i.e. $1\text{Hz} = 1\ \text{s}^{-1}$ and in Figure 2.1, the period of both the red and blue curves are the same and therefore, their frequencies are both $1/p$Hz. For example, the Earth's geostationary orbit is around 23 hours and 26 minutes or 86160 seconds resulting in a frequency of approximately 1.16×10^{-5}Hz while the period of the BBC Tees transmission signal is 1.05×10^{-8} s giving a frequency of 95 MHz. In the first case, the period is more easily comprehended than the frequency, in contrast with the second one.

The phase shift between two curves is defined as the difference in phase between two equivalent points, i.e. the phase shift is the ratio of the time shift and the period times 2π. In Figure 2.1, $(q - p)$ is the time shift between the two sinusoids and $2\pi(qp^{-1} - 1)$, the phase shift. If the time shift between two curves is zero, they are said to be in phase and a set of points from distinct wave functions in the same phase is called a wavefront. The wavelength is the spatial period of the wave, i.e. the distance travelled by the wave during one period and it is usually measured in metres (m) so, in our example, the length of the path of the red curve from the origin

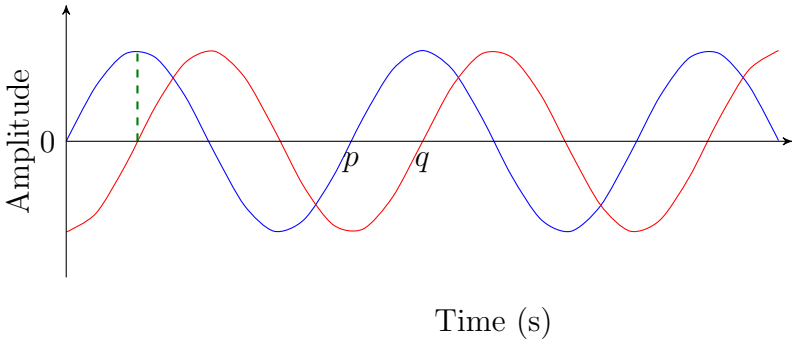

Figure 2.1: Representation of two sinusoidal wave functions in the form of amplitude versus time. The blue curve represents a zero-phase curve and the red one the same curve with a time shift of $(q - p)$. The green line indicates the peak amplitude of the blue sinusoidal and the distance from the origin to the point $(p, 0)$ is the period of the blue wave function.

at $(0, 0)$ to the point $(0, p)$ represents its wavelength. Other important measures are the angular wavenumber which is the ratio between 2π and the wavelength and is expressed in radians per metre, and the wave speed which is the ratio between the wavelength and the period and is measured in metres per second (m/s). The wave speed depends only on the medium through which the wave is propagated and this will be discussed later on when talking about physical properties of waves.

2.2 Wave Physical Properties

Waves tend to travel in rectilinear trajectories through media with constant properties. We define a medium as being a material that can propagate energy, and the direction and speed with which these waves propagate depend on the intrinsic properties of the medium. In this work, we are only interested in bounded media, i.e. finite in extent, and we will consider both isotropic and anisotropic media. A medium is called isotropic if its physical properties are the same in all directions and anisotropic otherwise. Unless specified, we will use the term anisotropic medium to refer to a medium in which velocity components are different. We will use the terms interface and boundary to refer to the change between two distinct media.

Given the direction that they travel through a medium, mechanical waves can be categorised into longitudinal waves and transverse waves. Longitudinal waves travel in the same direction as the disturbance, while transverse waves travel perpendicular to the direction of the disturbance (see Figures 2.2 and 2.3). Non-electromagnetic longitudinal waves, like acoustic waves, travel through elastic and non-elastic materials while non-electromagnetic transverse waves can only travel through elastic media limiting their propagation to solids and viscoelastic fluids. However, this classification is not binary; waves can present a combination of transverse and longitudinal motions like surface waves also known as Rayleigh or Love waves.

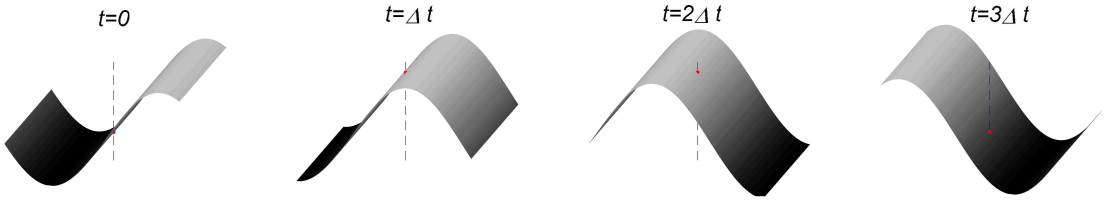

Figure 2.2: Propagation of a transverse planar wave: the oscillations are perpendicular to the direction of propagation of the wave, e.g. the red dot on this surface follows the path represented by the blue line which is perpendicular to this surface at any given time t

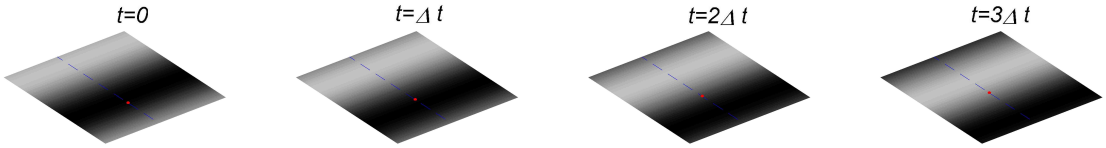

Figure 2.3: Propagation of a longitudinal planar wave: the oscillations are parallel to the direction of propagation of the wave, e.g. the red dot on this surface follows the path represented by the blue line which is parallel to this surface at any given time t

Now we will study the phenomena that describe a wave's behaviour when propagated through a medium and when it reaches a boundary. The simplest phenomenon is absorption, which is the property of a material to transform energy into another

form. For example, magnesium sulphate converts acoustic energy into heat due to ionic relaxation (Fisher and Simmons, 1977) and chloroplasts transform light energy into chemical energy. Jointly with dispersion or scattering, absorption is one of the main causes of attenuation, which is the rate of loss in intensity of a signal through a medium, e.g. sunlight is attenuated by sunglasses. Factors like wave source and frequency might help overcome these problems partially. Take as an example the absorption of sound waves in sea water, where the water itself, boric acid and magnesium sulphate are the main causes of signal loss. Frequencies of up to 10 kHz are absorbed due to the ionic relaxation of boric acid and are highly affected by background noise such as turbulence and micro tremors. Frequencies up to 500 kHz are affected by the ionic relaxation of the magnesium sulphate but frequencies greater than 1 MHz are absorbed by water or damped by thermal noise leaving the range of adequate values fairly restricted depending on the chemical composition of the water.

While these are the main causes of noise and recording problems, the properties that we are mostly interested in analysing are reflections, refractions and diffractions. We will use Huygens' and Fermat's principles to explain their behaviours and derive relevant laws. First we will start with Huygens' principle and later on this Chapter we will state Fermat's principle.

Theorem 2.2.1. *Huygens' principle - Given a wavefront, each point of the wavefront may be regarded as the source of new spherical secondary wavefronts. At a given time, the envelope of these secondary wavefronts determines the future position of the primary wavefront.*

Although Huygens referred to wavefronts of light in his original paper (Huygens, 1690), this principle is applicable to any type of wave motion. Fresnel (1818) and Kirchhoff (1882) are responsible for the mathematical formulation of Huygens' principle and, for an updated discussion in this topic, see Baker and Copson (1950) and Miller (1991).

We know that wavefronts are continuous curves or surfaces and may behave in complicated ways through media and interfaces. To simplify the problem, we assume that wavefronts are sets of infinitely many narrow beams of light, particles

or energy called rays. By discretizing the problem, we can analyse the propagation of waves based on the behaviour of each ray. While ray tracing does not explain phenomena like diffraction and interference, it simplifies the principles of reflection and refraction, giving a good approximation for isotropic conditions. In order to understand the ray tracing approximation, we need to introduce the wave equation and explain a method to solve it. We will come back to reflections and refractions when we establish the viability of this discrete approximation (Cerveny, 2001).

2.3 The Wave Equation

The wave equation is a well-known example of a second-order partial differential equation as it describes mathematically any type of waves. Its most common version is given by

$$\nabla^2\psi = \frac{1}{v(\mathbf{x})^2}\frac{\partial^2\psi}{\partial t^2} \tag{2.1}$$

where ψ is a function of time t and spatial components and v is the propagation velocity of the wave. If we solve Equation 2.1 by assuming that the solution is a plane wave we have that

$$\psi = \psi_0 \exp\left[-i\left(\omega t - \mathbf{k}\cdot\mathbf{x}\right)\right] \tag{2.2}$$

where $\mathbf{k}$ is the wave's directional vector, which determines the direction of propagation of the wave. Substituting 2.2 into Equation 2.1, we have the propagation velocity

$$v^2 = \frac{\omega^2}{|\mathbf{k}|^2}. \tag{2.3}$$

Define the travel time surface $T(\mathbf{x})$ as

$$T(\mathbf{x}) = \frac{1}{\omega}\mathbf{k}\cdot\mathbf{x} \tag{2.4}$$

then

$$\psi = \psi_0 \exp\left[-i\omega\left(t - T(\mathbf{x})\right)\right]. \tag{2.5}$$

Calculating $\nabla^2\psi$

$$\nabla^2\psi = -\psi\omega\left[\omega\left(\nabla T\right)^2 - i\nabla^2 T\right] \tag{2.6}$$

and the second-order partial derivative of ψ in t

$$\frac{\partial^2\psi}{\partial t^2} = -\omega^2\psi, \tag{2.7}$$

we can combine them using Equation 2.1 simplifying to

$$\left(\nabla T\right)^2 = \frac{1}{v^2} \tag{2.8}$$

which is called the eikonal equation which describes the travel time $T(\mathbf{x})$ for a ray passing through a point $\mathbf{x}$ in a medium with velocity $v(\mathbf{x})$.

It can be proven that the eikonal equation is a good approximation to the wave equation in the high-frequency limit. It can be shown that this approximation is valid if variation in the velocity gradient is much smaller than the frequency (Officer, 1958; Wu and Aki, 1988; Lay and Wallace, 1995). This implies that the eikonal equation, and by consequence ray tracing, are not good approximations in cases where the velocity gradient across boundaries is too high compared to the velocities of the layers. However, since we aim to produce macro velocity models, we assume that ray tracing gives an adequate approximation. Taking that into account, we can state Fermat's principle and discuss the laws of reflection and refraction.

2.4 Fermat's Principle

Proposed in a different setting from Huygens' principle, Fermat's principle of least time forms the basis of geometric optics. It explains why rays of light are rectilinear in a homogeneous medium and it can be used to derive the principles of reflection and refraction. Fermat stated his principle of least time as follows.

Theorem 2.4.1. *(Fermat's Principle) A ray, travelling between two points, must traverse an optical path whose length is stationary with respect to variations of the path.*

Like Huygens, Fermat had in mind light rays but, by considering the ray tracing approximation, we can apply this principle to any kind of wave. Fermat's principle is a counter intuitive idea since most of us tend to think that the shortest Euclidian path is always the best one; Fermat says that rays may travel through the critical paths and those do not need to be the shortest ones. This implies that these paths can take the least or the most time or be an inflection. For example, say a source of light is placed in the centre of an ellipsoid mirror. The time the rays take to travel from the centre to the surface and back to the centre is minimum when travelling on the minor axis and it is maximum on the major axis.

It is interesting to note that Fermat's principle can be immediately derived from Huygens' principle if we assume that the rays derived from the eikonal equation are normal to their wavefronts. We now introduce the law of reflection.

Theorem 2.4.2. *(Law of Reflection) In an homogeneous medium, the angle of reflection is the same as the angle of incidence.*

In Figure (2.4), we have an illustration of this law; note that it does not matter if the medium is isotropic or anisotropic for the law of reflections to be valid but it is necessary for it to be homogeneous to guarantee a rectilinear ray path.

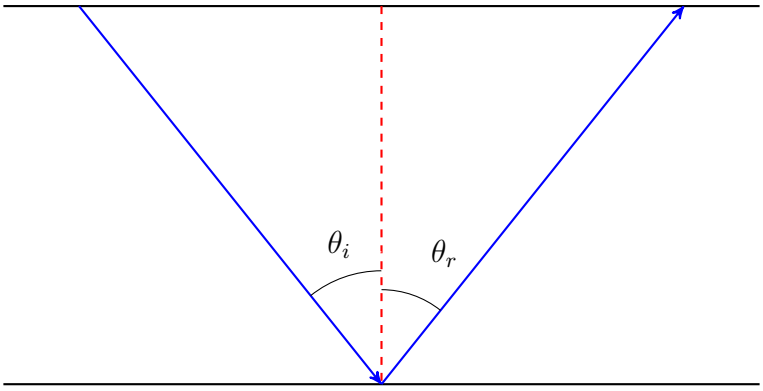

Figure 2.4: Illustration of the Law of Reflection; the angle of incidence θ_i of the blue ray is the same as its angle of reflection θ_r

Next we introduce Snell's Law or the Law of Refraction (see Figure 2.5). Here we require that the two media through which the ray will travel to be isotropic; there are extensions of this law for anisotropic media, but they are more complex.

Theorem 2.4.3. *(Snell's Law) The ratio of the sines of the angles of incidence and refraction equals the ratio of the wave's speed in the two media.*

In Figure 2.5, we have an illustration of Snell's law. If we consider the angle of incidence θ_1 and the angle of refraction θ_2 then

$$\frac{\sin(\theta_1)}{\sin(\theta_2)} = \frac{v_1}{v_2} \tag{2.9}$$

where v_1 is the wave speed in the first medium and v_2, the speed in the second medium.

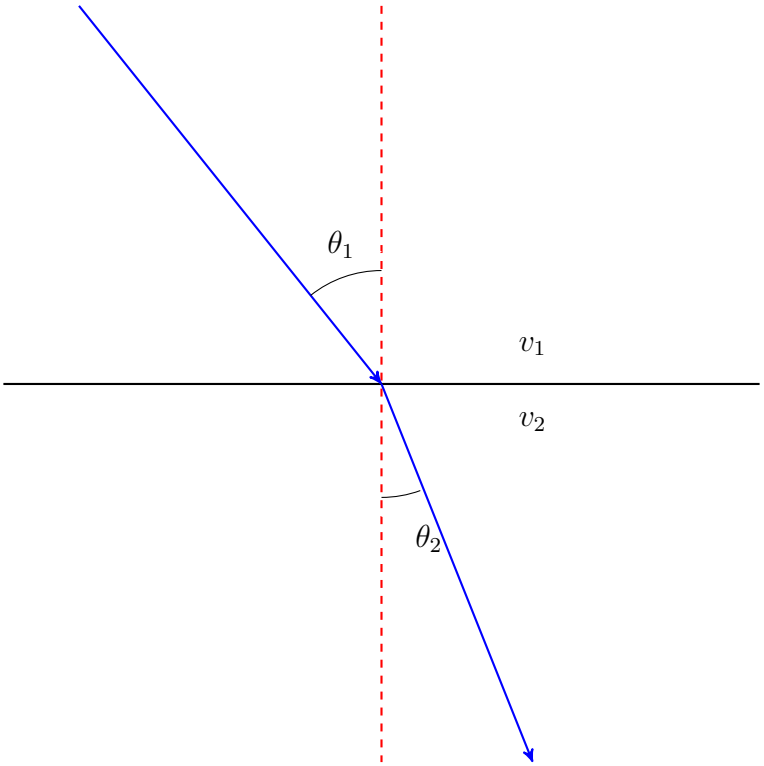

Figure 2.5: Illustration of the Law of Refraction; the ratio of the sines of the angles of incidence θ_1 and refraction θ_2 is the same as the ratio of the speeds before and after refraction, v_1 and v_2.

From Snell's Law, we can derive the idea of total internal reflection and critical angle. Rewriting Equation 2.9 we have

$$\sin(\theta_1) = \frac{v_1}{v_2}\sin(\theta_2) \Rightarrow \theta_1 = \arcsin\left(\frac{v_1}{v_2}\sin(\theta_2)\right). \tag{2.10}$$

When $\theta_2 = \pi/2$, θ_1 reaches its critical angle $\theta_c = \arcsin(v_1/v_2)$. For angles of θ_1 that are greater than θ_c, no ray can propagate across the boundary and we call this phenomenon total internal reflection. For the special case of θ_1 equalling θ_c, the rays are refracted along the boundary and travel horizontally at the velocity of the underlying medium.

The phenomena of reflection and refraction can be easily understood using ray theory but two other phenomena, diffraction and interference, can only be fully understood if the shape and phase of the wavefronts are taken into account.

When two or more waves interact, we have the phenomenon of interference. Under the assumption of linearity, in a homogeneous and additive system, two wavefronts can be linearly combined to create a new wavefront as a consequence of the superposition principle. The two most common examples of interference are illustrated in Figures 2.6 and 2.7. In Figure 2.6, we have two identical wave functions in the same phase. The combined wave function has the same period and phase but higher amplitudes. In Figure 2.7, the same two wave functions are combined but one of them is out of phase by π creating a flat signal. In the first case, we have constructive interference and in the second destructive interference. While linearity does not always hold, waves that do not satisfy the linear wave equation or the superposition principle are rare in the context of seismic reflection. Examples of waves that do not satisfy linearity are shock waves in a sonic boom and electron acoustic waves.

Finally the last phenomenon that we have to discuss is diffraction; it can lead to complex discontinuous waveforms which can be approximated using Fourier transforms. Diffraction refers to the behaviour of an wave when it reaches an obstacle and it appears to behave as if it were bending. The most common example for this phenomenon is the single-slit experiment. Assume we have an unbounded medium and we divide it with a fully reflective boundary in the sense that any energy that reaches it will not be refracted, just reflected. Also assume that this boundary's thickness is negligible and that a small slit is made on this boundary. Now if we create a wavefront that travels parallel to this boundary, when this wavefront hits the boundary, it will be reflected in the opposite direction everywhere but on the

slit where it will be propagated as a series of radial wavefronts through the slit. This is a direct consequence of Huygens' principle combined with the superposition principle. See Keller (1962) for an overview on diffractions.

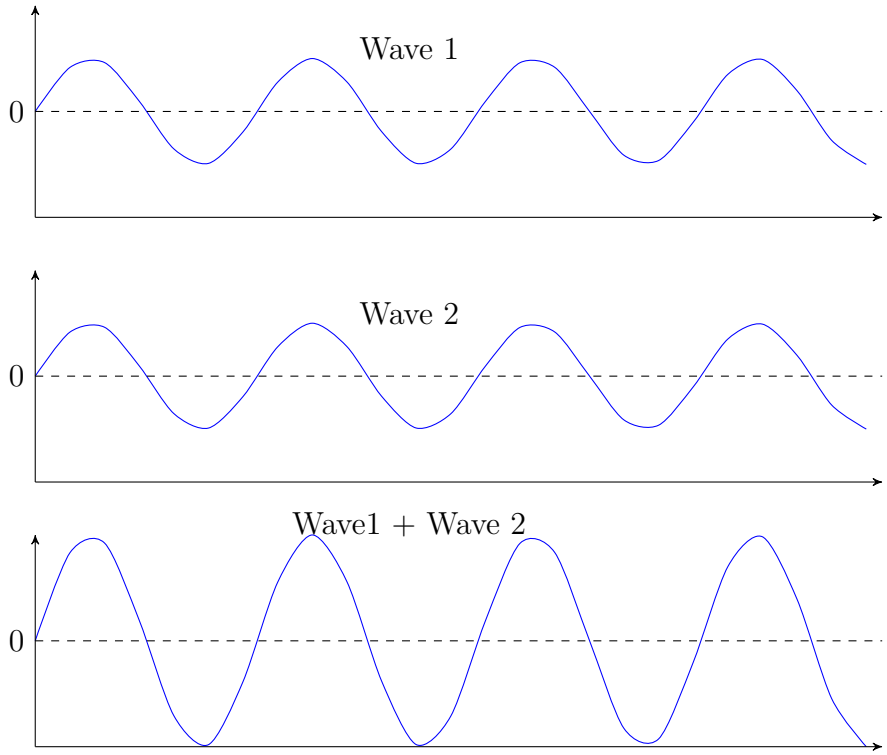

Figure 2.6: Superposition of two identical wave functions, Wave 1 and Wave 2, illustrating constructive interference.

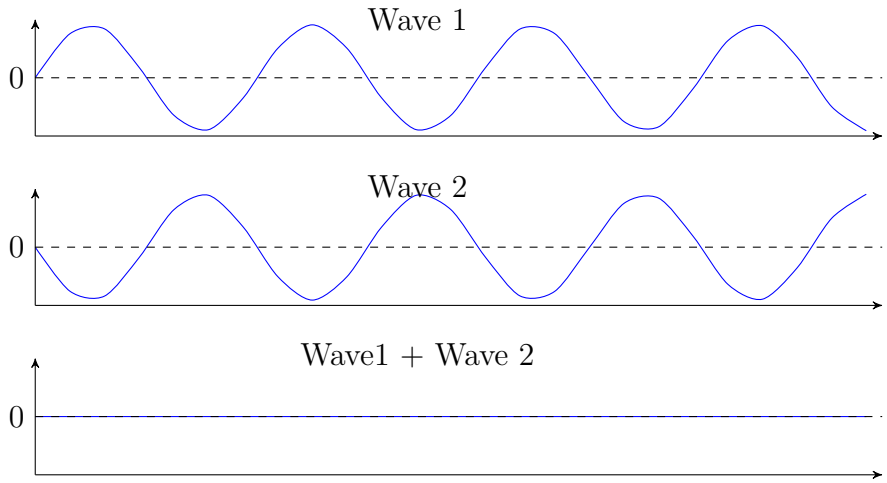

Figure 2.7: Superposition of two wave functions with phase difference π illustrating destructive interference.

2.5 Seismic Waves

All the theory that we have discussed so far applies to any kind of waves but we are mostly interested in the behaviour of seismic waves, how to record and reconstruct their path. Now we explain their mechanisms and describe the necessary tools for reflection seismology and the design of the velocity models studied here. See Sheriff and Geldart (1995), Kearey et al. (2002), Yilmaz (2002) and Jones et al. (2008) for further details.

As the name suggests, seismic waves are waves that travel through the Earth; they are commonly divided into two categories: surface waves and body waves. Surface waves travel along interfaces of different media and, in earthquake seismology, the term refers to waves that travel near the Earth's surface. They are one of the main causes of destruction related to earthquakes and, due to their wide frequency range, they provide a source of error in seismic data collected on land, where they are called ground roll, and that has to be addressed during the design of the experiment, i.e. by setting a source-receiver grid that covers the region to be explored and applying filters to the collected data capable of attenuating the signal produced by surface waves (Halliday et al., 2007). Body waves travel through the interior of the Earth and are the ones studied in reflection seismology. After addressing surface waves, we will focus only on body waves.

There are two types of surface waves: Rayleigh waves and Love waves. Rayleigh waves travel longitudinally and vertically; when travelling on a plane, they form elliptical paths that decrease with depth. Love waves or L-waves are transverse waves that travel in elastic media; unlike Rayleigh waves, they do not travel vertically. In the immediate area of the epicentre of an earthquake, Rayleigh waves are the most destructive; their amplitude decay is inversely proportional to the root of the radius of propagation implying that they can travel for much longer than other types of waves before fading. In terms of velocity of propagation, Rayleigh waves are the slowest and typically travel close to the shear-wave velocity of the surface layers.

Body waves are elastic waves that can be classified as primary or shear waves. Primary waves or P-waves are longitudinal waves while shear waves or S-waves are transverse. Due to their nature, P-waves travel through any material while S-waves,

which are slower than P-waves, only travel through solid materials. P-waves are the first to be recorded and also generate a stronger signal than S-waves making them the most relevant waves in the study of reflection seismology. In the next Chapter we will discuss the geometry of seismic ray paths using the properties and phenomena presented so far and discuss how prior knowledge about basic rock physics can be used to improve the model.

Chapter 3

Velocity Models, Data Recording and BRAINS

In the previous Chapter we discussed wave theory and an approximation for the wave equation and discussed the basic phenomena of reflection and refraction. Under the assumption that the eikonal equation is a suitable approximation for the wave equation, we concluded that we can describe the behaviour of a wavefront by analysing its corresponding rays. We can also assert that, for isotropic media, each ray is normal to such wavefront and consequently, the wavefront is an envelope of the tangent lines perpendicular to these rays at a given time. In this Chapter, we will discuss velocity models and approximations, and the notation used in the rest of this work will be defined here.

To introduce our first velocity model, we will trace the path of a single ray travelling through an isotropic medium bounded by two parallel boundaries as illustrated in Figure 3.1. Let t be the travel time of the blue ray from point S to point R and t_0 the time that a similar ray would take to travel from point M to point C and back. The blue ray travels a distance of $vt/2$, at a constant velocity v, from point S to point C and travels the same distance from point C to point R. Using the Pythagorean theorem, we find that

$$\left(\frac{vt}{2}\right)^2 = z^2 + \left(\frac{x}{2}\right)^2. \tag{3.1}$$

Substituting $z = vt_0/2$ and assuming that $t \geq 0$ then we have the single-layer travel

time equation

$$t = \sqrt{{t_0}^2 + \left(\frac{x}{v}\right)^2} \tag{3.2}$$

where the travel time t_0 is usually called the two-way zero-offset travel time.

In the case where multiple parallel boundaries exist, we can deduce an approximation similar to the single layer result. We start with n isotropic media with varying velocities and thicknesses as represented in Figure 3.2.

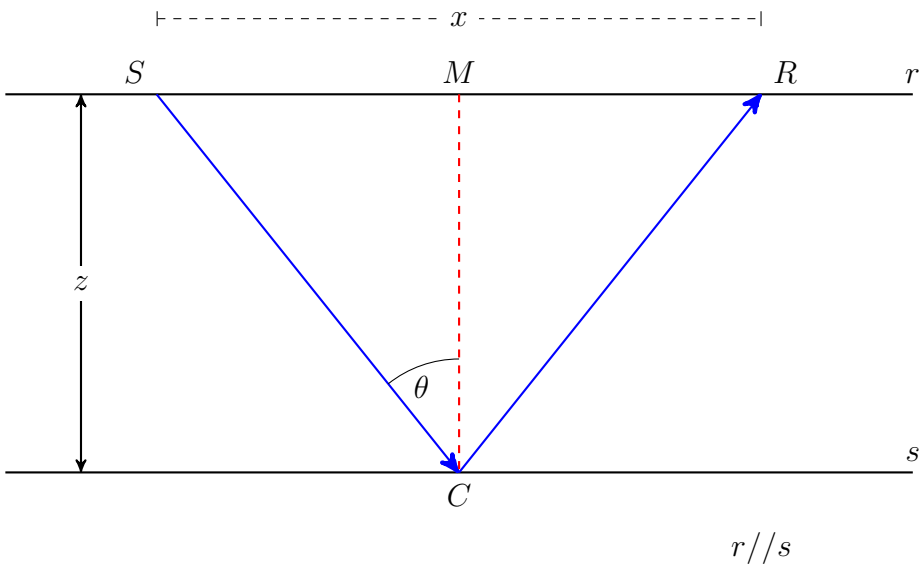

Figure 3.1: Illustration of the travel path of a single ray originated at the point S, reflected from C and recorded at R. The ray travels through the medium bounded by the parallel lines r and s separated by a distance z and with constant velocity v in all directions.

Each layer i, with boundaries $b_{(i-1)}$ and b_i, has a velocity v_i in all directions and thickness Δz_i, therefore the distance between boundaries b_0 and b_i is $z_i = \sum_{k=1}^{i} \Delta z_k$. The two-way zero-offset travel time variation for the i-th layer is Δt_{0_i} meaning that this is the time that a ray takes to travel perpendicularly from boundary $b_{(i-1)}$ to boundary b_i and back to $b_{(i-1)}$, and the total two-way travel time for the i-th layer, $t_{0_i} = \sum_{k=1}^{i} \Delta t_{0_i}$ is the minimum time that a ray takes to travel perpendicularly from boundary b_0 to boundary b_i and back to boundary b_0. Define the ray's travel time in the i-th layer as Δt_i and this value corresponds to the time taken to travel from the

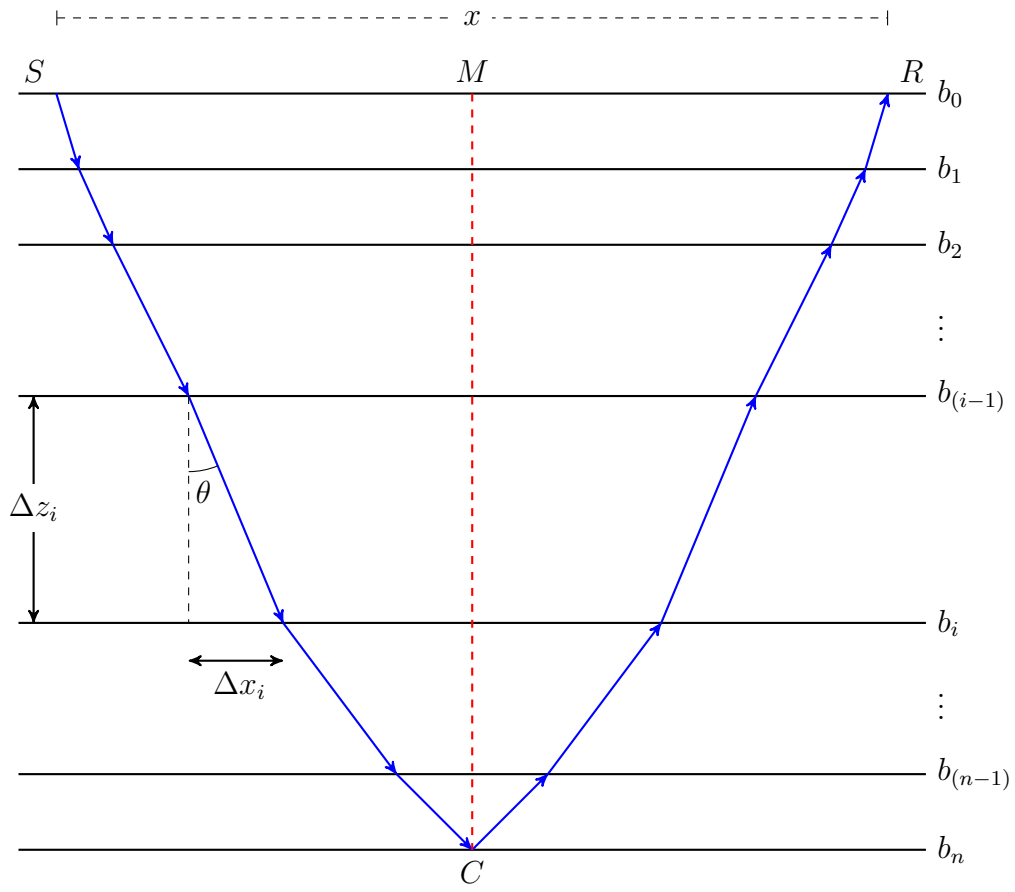

Figure 3.2: Illustration of the travel path of a single ray originated at the point S, refracted through $(n-1)$ boundaries, reflected from C and recorded at R. The ray travels through the media bounded by the parallel lines b_i, $i = 1, \ldots, n$, separated by a distance Δz_i and each with constant velocity v_i in all directions.

point of refraction in boundary $b_{(i-1)}$ to the next point of refraction in b_i. Let Δx_i be the horizontal distance between points of refraction. Now define t_i as the total time taken by a ray to travel from the source in b_0, passing through the refractors b_1 to b_{i-1} and the reflector in b_i and returning to the receiver in b_0 and define x as the distance between the source and receiver in b_0.

If we label each angle of refraction θ_i to indicate the angle formed by the ray with boundary $(i-1)$ after refraction then, from Snell's law, we can deduce that

$$\frac{\sin(\theta_i)}{v_i} = \frac{\sin(\theta_k)}{v_k} \tag{3.3}$$

for any i and k in $\{1, \ldots, n\}$.

From the diagram in Figure 3.2, we can see that

$$\Delta t_i v_i = \sec(\theta_i)\Delta z_i \Rightarrow \Delta t_i = \frac{\Delta t_{0_i}}{2\cos(\theta_i)} \tag{3.4}$$

and

$$\Delta x_i = \Delta z_i \tan(\theta_i) = \frac{1}{2} v_i \Delta t_{0_i} \tan(\theta_i). \tag{3.5}$$

Now define the slowness p, i.e. the horizontal component of velocity, as

$$p = \frac{\sin\theta_i}{v_i},\ \forall i \in \{1, \ldots, n\} \tag{3.6}$$

then

$$t_n = 2\sum_{i=1}^{n} \Delta t_i = \sum_{i=1}^{n} \frac{\Delta t_{0_i}}{\cos\theta_i} = \sum_{i=1}^{n} \frac{\Delta t_{0_i}}{\sqrt{1 - p^2 v_i^2}} \tag{3.7}$$

and

$$x = 2\sum_{i=1}^{n} \Delta x_i = \sum_{i=1}^{n} \frac{\Delta t_{0_i} v_i^2 p}{\sqrt{1 - p^2 v_i^2}}. \tag{3.8}$$

Writing the Taylor expansion for $\sec(\theta_i)$ in p around the point $p = 0$ we have that

$$t_n = \sum_{i=1}^{n} \Delta t_{0_i} \sum_{m=0}^{\infty} \binom{2m}{m} \left(\frac{p v_i}{2}\right)^{2m} \tag{3.9}$$

$$= \sum_{i=1}^{n} \Delta t_{0_i} + \frac{p^2}{2} \sum_{i=1}^{n} \Delta t_{0_i} v_i^2 + O(p^4) \tag{3.10}$$

$$x = \sum_{i=1}^{n} \Delta t_{0_i} v_i^2 p \sum_{m=0}^{\infty} \binom{2m}{m} \left(\frac{p v_i}{2}\right)^{2m} \tag{3.11}$$

$$= p\sum_{i=1}^{n} \Delta t_{0_i} v_i^2 + \frac{p^3}{2} \sum_{i=1}^{n} \Delta t_{0_i} v_i^3 + O(p^5). \tag{3.12}$$

Since $v_i \gg 1$, $\forall i \in \{1, \ldots, n\}$, then $0 < p \ll 1$ and $O(p^4)$ and $O(p^5)$ are small and bound the errors of these approximations. Now, squaring both t_n and x gives us

$$t_n^2 = \left(\sum_{i=1}^{n} \Delta t_{0_i}\right)^2 + p^2 \left(\sum_{i=1}^{n} \Delta t_{0_i}\right)\left(\sum_{i=1}^{n} \Delta t_{0_i} \Delta v_i^2\right) + O(p^4) \tag{3.13}$$

$$x^2 = p^2 \left(\sum_{i=1}^{n} \Delta t_{0_i} v_i^2\right)^2 + O(p^4). \tag{3.14}$$

Define

$$t_{0_n} = \sum_{i=1}^{n} \Delta t_{0_i} \tag{3.15}$$

as the total two-way zero-offset travel time and

$$v_{rms_n} = \sqrt{\frac{1}{t_{0_n}} \sum_{i=1}^{n} \Delta t_{0_i} v_i^2} \tag{3.16}$$

as the n-th root-mean-square (RMS) velocity then

$$t_n^2 = t_{0_n}^2 + p^2 t_{0_n}^2 v_{rms_n}^2 + O(p^4) \tag{3.17}$$

$$x^2 = p^2 t_{0_n}^2 v_{rms_n}^4 + O(p^4) \tag{3.18}$$

and finally

$$t_n^2 \approx t_{0_n}^2 + \frac{x^2}{v_{rms_n}^2}. \tag{3.19}$$

Equation 3.19 is the so-called travel-time approximation and it will be used frequently throughout this work. Since the data that we will work with consists of recorded travel times, amplitudes and offset positions, resolving the number of boundaries and the corresponding interval velocities without this approximation is extremely hard. The RMS velocity provides us with an approximate velocity of what the real velocity would be if every boundary were combined to form a single-layered model. Given that we are able to estimate the RMS velocity for all layers in the model, we can define a relationship between RMS velocities and the velocities v_i which are commonly called interval velocities. Rewriting Equation 3.16, we have that

$$v_{rms_n}^2 = \frac{1}{t_{0_n}} v_n^2 \Delta t_{0_n} + \frac{t_{0_{n-1}}}{t_{0_n}} v_{rms_{n-1}}^2 \tag{3.20}$$

leading to Dix's formula (Dix, 1955, 1952)

$$v_n = \sqrt{\frac{t_{0_n} v_{rms_n}^2 - t_{0_{n-1}} v_{rms_{n-1}}^2}{t_{0_n} - t_{0_{n-1}}}}. \tag{3.21}$$

To produce this approximation, we assumed that the ray approximation was valid, the rays travel in a linear path within boundaries, the boundaries were parallel,

the media were isotropic and that the small-angle approximation was valid when truncating the Taylor series in Equations 3.11 and 3.12, i.e. $\sin(\theta) \approx \theta$ or $\tan(\theta) \approx \theta$. Those are all idealised assumptions but, apart from extreme cases, the model obtained from such approximations is close to reality, e.g. anisotropy is typically small and so is the variation in velocity with propagation direction.

Assume that the previous experiment is replicated m times by symmetrically placing m source-receiver pairs around M and recording the paths of the rays reflected at C. We assume that the sources are fired at different times in order to avoid interference.

Similarly to the problem designed in Figure 3.2, we set the $(n+1)$ parallel boundaries b_i creating n media with velocities v_i, thicknesses Δz_i and two-way zero-offset travel time variation Δt_{0i} for $i \in \{1, \dots, n\}$ and therefore total depth z_i and total travel time t_{0_i}. We label the source-receiver pairs so that the j-th ray is the ray fired from source S_j, reflected from M and recorded by receiver R_j and that distance between S_j and R_j is the offset x_j. Let θ_{ij} be the angle of refraction with boundary $b_{(i-1)}$ for the j-th ray, Δt_{ij} the travel time from the point of refraction in b_i to the next point of refraction in $b_{(i-1)}$ for this same ray and Δx_{ij} the horizontal distance between those two points. In Figure 3.3, we have a description of this setup.

The set of travel times t_{ij} for a fixed i with $j \in \{1, \dots, m\}$ is called a normal move out. Given the assumptions of isotropy and homogeneity of the media, the move out of the time increments is hyperbolic since

$$\Delta t_{ij} = \sqrt{\left(\frac{\Delta x_{ij}}{v_i}\right)^2 + \left(\frac{\Delta t_{0i}}{2}\right)^2},\ i \in \{1, \dots, n\},\ j \in \{1, \dots, m\}. \tag{3.22}$$

Therefore, assuming noise-free data, we can use the move-out to resolve the RMS velocities and two-way zero-offset travel times for each layer and then, in a stepwise process, we can estimate the corresponding interval velocities and thicknesses. We can use the rays reflected from the first boundary to estimate the parameters for the first layer and then, given these estimates, use the recorded travel times for the rays reflected from the second boundary to estimate the next pair and so on.

For the above we assume that all media are approximately isotropic and that all boundaries are locally parallel and, in most cases, these assumptions will be

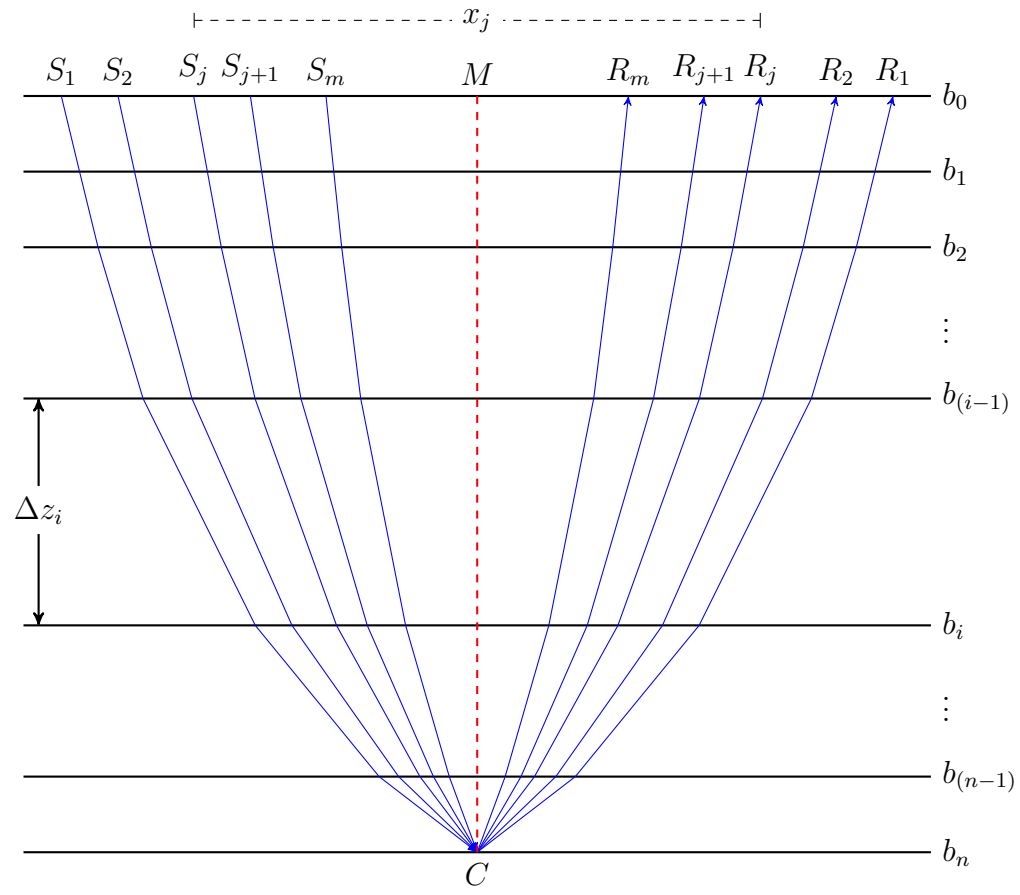

Figure 3.3: Illustration of the travel path of multiple rays travelling from sources S_j, $j = 1, \ldots, m$ to receivers R_j after passing through the media bounded by the parallel lines b_i, $i = 1, \ldots, n$ separated by a distance Δz_i and each with constant velocity v_i in all directions.

sufficient to produce reasonable estimates of the parameters described so far. Later we will analyse examples where these assumptions are not valid and how this affects the errors in our models. Next we introduce the basics of data recording based on the layouts we described in this Section.

3.1 Seismic Data Recording

A seismic survey relies on a configuration of seismic sources and receivers that are usually placed in a rectilinear grid. In the previous Section, we talked about grids for 2D data recording but, by simply adding another axis, we can create a grid

of sources and receivers placed equidistantly from a common midpoint and all the previous calculations still hold. Usually a seismic source, like an air gun (i.e. a pneumatic system pressurized with compressed air fired below water surface) or a vibroseis truck (i.e. a truck designed to produce vibration by shaking its mass), is placed at a predefined position relative to an array of receivers and fired to produce seismic waves. The receivers are set to record a continuous constant analogue signal, this signal is disturbed by the arrival of mechanical waves and the amplitudes of those disturbances are recorded. On land, a geophone detects ground vibrations and in water, a hydrophone detects changes in pressure. The oscillations detected are converted into analogue electrical signals and recorded onto a magnetic tape or a similar medium; these are then converted to a digital data stream by an analogue to digital converter for further analysis.

The amplitude of the signal recorded is a function of the amplitude of the incoming wave and intrinsic properties of both media, i.e. density, P-wave and S-wave velocities, and this relationship is explained by the Zoeppritz equations which are part of a system of linear equations that relate the amplitudes of the reflected and transmitted primary and shear waves to the angles of reflection and refraction of these waves (Yilmaz, 2002). The signal produced by a seismic source is bipolar and locally bandlimited resulting in a wavelet that consists of a short series of amplitude peaks of alternating polarity. The time at which the onset of the first event arrives corresponds to the travel time of the seismic wave from the source, reflected from a certain boundary and to the receiver. Assuming that part of the energy was refracted, when it reaches the next boundary, part of this energy should be reflected creating another event on the recorded signal trace and so on. By the end of the recording, there should be a series of overlapping wavelets with different amplitudes each corresponding to a different boundary from which the energy was reflected.

After the first source is fired and the energy of the seismic waves produced dissipated, we can fire another source in a different position in the same line. If we repeat this process a certain number of times, we will have multiple traces recorded by each receiver from different source positions. The set of traces recorded given a fixed source position is called a common shot gather and the set of traces with a

matching geometrical mid-point between the source and receiver is called a common midpoint gather. For a simple one-dimensional model as used above, the CMP point and the reflection points at each boundary lie on the same line which is normal to all boundaries. However, this is not generally true in the real Earth where lateral velocity changes or geological structure perturb the ray paths. These issues are addressed later through velocity-dependent imaging. Throughout this work, we will use the term CMP gather to refer to a common midpoint gather.

3.2 Analogue to digital conversion, sampling and sources of noise

Now we discuss the recording of analogue signals, the analogue to digital conversion and possible sources of noise. The use of analogue signals is interesting because, at least in theory, the electric signal produced by the receiver is continuous and creates a smooth curve with infinite resolution while digital signals rely on the discretization of a continuous curve and the sampling rate. However, there are issues with analogue signals that cannot be dealt with without digitization. The energy that reaches the receivers also includes environmental or ambient noise that is not part of the required signal and, in addition, the source signal itself may contain unwanted artefacts or ray paths that do not conform to the simple model above, called multiples. This noise degrades the ability to resolve the subsurface structure. The distortion and loss caused by these variations cannot be recovered since the noise and interference are amplified when the signal is amplified. Therefore, the main reason to digitize an analogue signal is that most noise filtering and data processing methods are only available in a digital format.

When converting an analogue signal into a digital format, the data are discretized at a sampling rate that will produce the required resolution after analogue filtering to remove frequencies that would be aliased at the specified sampling rate. According to Nyquist-Shannon sampling theorem, if the signal contains no frequencies higher than α hertz then it is completely determined by a sequence of points $(2\alpha)^{-1}$ apart (Cover and Thomas, 2001). After the digitization, the data goes through digital

filters to improve the signal to noise ratio, smooth the traces and combine them in shot or CMP gathers. Both in land and under water, the designer of a seismic experiment has to take a number of noise sources into account in order to record an usable signal. Typically these comprise of bandpass filters to exclude noise outside of the bandwidth of the seismic source; deconvolution filters that can be used to shape the wavelet to improve the resolution or to remove energy from unwanted ray paths such as multiples (energy that has repeatedly bounced within one or more layers); or more sophisticated filters that transform the data to isolate and suppress unwanted events and noise. Given that a fairly clean signal was recorded, the processing of the data becomes easier and the number of unknowns during modelling is reduced.

Most of the filters applied to seismic data will not affect negatively the analysis that we will carry in the next Chapters as long as the expert applying these filters accounts for possible shifts in time. See Yilmaz (2002) for the most commonly used filters in seismic signal processing.

3.3 BRAINS: Bayesian Regression Analysis in Seismology

In the preceding Sections, we have discussed most of the tools necessary to build functional models for the analysis of seismic reflection problems. While there are many variables to discuss, we aim to estimate velocities and depths given source-receiver positions and recorded travel-times. The observed variables are the position of sources and receivers and the travel times and amplitudes recorded at the receivers.

Let S_j and R_j be a source-receiver pair from a rectilinear array containing m pairs. Assuming that each pair is equidistantly placed around a common midpoint as in Figure 3.3, let X_j be the distance between S_j and R_j. Each receiver R_j records a continuous signal of length T_{tot} seconds which is sampled every T_{sp} seconds generating m_s sampling points for each receiver. Let A_{jk} be the recorded amplitude at receiver j, $j \in \{1, \ldots, m\}$, and at sample $k \in \{1, \ldots, m_s\}$ and let $T_{jk}^{(o)}$ be the recorded time at that same moment. Both matrices $A = [A_{jk}]_{j=1,\ldots,m;k=1,\ldots,m_s}$ and

$T^{(o)} = \left[T^{(o)}_{jk}\right]_{j=1,\ldots,m;k=1,\ldots,m_s}$ are $m \times m_s$ matrices from which we need to extract information for each layer to be modelled. The matrix A is called the common midpoint (CMP) gather; in Figure 3.4, we show a greyscale map of a CMP gather from a real dataset. In the beginning of this Chapter we proved that the recorded travel times reflected from a common reflector were approximately hyperbolic given a number of assumptions and concluded that if these conditions are acceptable locally, like parallelism of boundaries and homogeneity of the media, then we can extend this approximation to real data. In Figure 3.4, we can see these hyperbolic patterns so we need to sort the points in these curves and link them to their corresponding reflector. Note that part of the modeling error will be due to this approximation.

If we assume that we can discretize the number of boundaries through which the signal travels and that there is a finite number of boundaries b_i, $0 \leq i \leq n \leq m_s$, where b_0 is the line that contains the sources and receivers, then we can associate an interval velocity V_i, a thickness ΔZ_i and a two-way zero-offset travel time ΔT_{0_i} to each layer L_i bounded by $b_{(i-1)}$ and b_i. Let Z_i be the depth and T_{0_i} be the two-way zero-offset travel time of boundary b_i, i.e. $Z_i = \sum_{k=1}^{i} \Delta Z_k$ and $T_{0_i} = \sum_{k=1}^{i} \Delta T_{0_k}$.

Let T_{ij} be the real time taken for a ray to travel from source S_j to receiver R_j by refracting through boundaries b_1 to $b_{(i-1)}$, reflecting from b_i and refracting back to receiver R_j. We say that $T = [T_{ij}]_{i=1,\ldots,n;j=1,\ldots,m}$ is the matrix of real travel times for a design with n layers and m source-receiver pairs. Considering the same raypath used to describe T_{ij}, we say that $T^{(r)}_{ij}$ is the recorded travel time and $T^{(r)} = \left[T^{(r)}_{ij}\right]_{i=1,\ldots,n;j=1,\ldots,m}$ is the matrix of recorded travel times associated to T. Finally, let V_{rms_i} be the RMS velocity associated to layer L_i.

We need to estimate the vectors $V = [V_i]_{i=1,\ldots,n}$, $V_{rms} = [V_{rms_i}]_{i=1,\ldots,n}$, $T_0 = [T_{0_i}]_{i=1,\ldots,n}$ and $Z = [Z_i]_{i=1,\ldots,n}$ from observed A, $T^{(o)}$ and $X = [X_j]_{j=1,\ldots,m}$. In order to establish the model, we will assume that the submatrix $T^{(r)}$ of $T^{(o)}$ is known and that X is observed without error and later we will explore ways to select $T^{(r)}$. If the boundaries b_i are parallel and the layers L_i are isotropic, then we can work with the equations deduced in the beginning of this Chapter. Starting with a single-layer

model, i.e. $n = 1$, we have that

$$T_{1j} = \sqrt{T_{01}^2 + \left(\frac{X_j}{V_1}\right)^2} + E_{1j}^{(m)} \tag{3.23}$$

where $E_{1j}^{(m)}$ is the modelling error associated to the isotropic assumption and ray tracing approximation.

Using Equation 3.19, we can extend this to a multi-layered model and

$$T_{ij} = \sqrt{T_{0i}^2 + \left(\frac{X_j}{V_{rms_i}}\right)^2} + E_{ij}^{(m)} \tag{3.24}$$

where V_{rms_i} is given by

$$V_{rms_i}^2 = \frac{1}{T_{0_i}} \sum_{l=1}^{i} \Delta T_{0_l} V_l^2 \tag{3.25}$$

where $E_{ij}^{(m)}$ is the error associated to the isotropic and parallelism assumptions, ray tracing and Dix's approximation.

We have defined $T^{(r)}$ as a submatrix of the matrix of observed travel times $T^{(o)}$ so we can associate a recording error (and a selection error) to each element of T. Say $E^{(r)} = \left[E_{ij}^{(r)}\right]_{i=1,\ldots,n;j=1,\ldots,m}$ is the matrix of recording errors associated to $T^{(r)}$ then

$$T^{(r)} = T + E^{(r)} \tag{3.26}$$

then

$$T_{ij}^{(r)} = \sqrt{T_{0i}^2 + \left(\frac{X_j}{V_{rms_i}}\right)^2} + E_{ij}^{(m)} + E_{ij}^{(r)}. \tag{3.27}$$

We know that the RMS velocities V_{rms_i} are a function of the two-way zero-offset travel times T_{0_i} and interval velocities V_i and the two-way travel times are a function of depths Z_i and interval velocities V_i. We also have that

$$\Delta Z_i = \frac{V_i \Delta T_{0_i}}{2} \tag{3.28}$$

then if we can resolve each two-way zero-offset travel time and velocity pair in a step-wise process, we can estimate the depth of a boundary b_i. For an isotropic medium, we can assume that the layer's thickness and its interval velocity are independent

implying that the two-way zero-offset travel time variation is a fraction of two independent random variables and the RMS velocity is a function of these same two variables. These are the variables necessary to introduce the models we will use in the next Chapters. We start with a model based on RMS velocities and zero-offset two-way travel times for single gathers. All statistical distributions mentioned in this Chapter can be found in Appendix A.

3.3.1 1D Stacking BRAINS

The first model we propose is a nonlinear regression of travel times as a function of RMS velocities and zero-offset travel times on a single gather that imposes strong assumptions. The purpose of this model is to approximate stacking velocities which are the velocities used to correct the travel times of events in each layer by removing their hyperbolic trend. Under the assumptions of homogeneity, isotropy and parallelism, the stacking velocity is the same as the RMS velocity. Stacking velocities are commonly used in intermediate steps during seismic data processing like time migration to improve image quality. Referring to equation 3.27, we assume that $E_{i\cdot}^{(m)}$ and $E_{i\cdot}^{(r)}$ are normally distributed with zero-mean and variance matrices $\Sigma_{i\cdot}^{(m)}$ and $\Sigma_{i\cdot}^{(r)}$, respectively, and, consequently, $T_{i\cdot}^{(r)}$ is also normally distributed with variance matrix $\Sigma_{i\cdot}^{(t)} = \Sigma_{i\cdot}^{(m)} + \Sigma_{i\cdot}^{(r)}$. Each $T_{ij}^{(r)}$ has mean

$$\mu_{ij}^{(t)} = \sqrt{{t_{0i}}^2 + \left(\frac{x_j}{v_{rms_i}}\right)^2},\ i \in \{1,\dots,n\},\ j \in \{1,\dots,m\} \tag{3.29}$$

at the point $(T_{0i}, V_{rms_i}) = (t_{0i}, v_{rms_i})$.

The choices of priors for the zero-offset travel times and RMS velocities depend on the expert's belief; in most cases, it is adequate to define priors over their increments ΔT_{0_i} and ΔV_{rms_i}, $i \in \{1,\dots,n\}$ to reflect weaker assumptions like increasing RMS velocities. When applying 1D Stacking BRAINS (1D S-BRAINS) in this work, we choose to define priors using increments and used Dix equation 3.21 to truncate such priors under the assumption that interval velocities are always nonnegative and sometimes bounded within a range specified by the expert. A simple choice of priors for the increments assume that they are independent within layers and

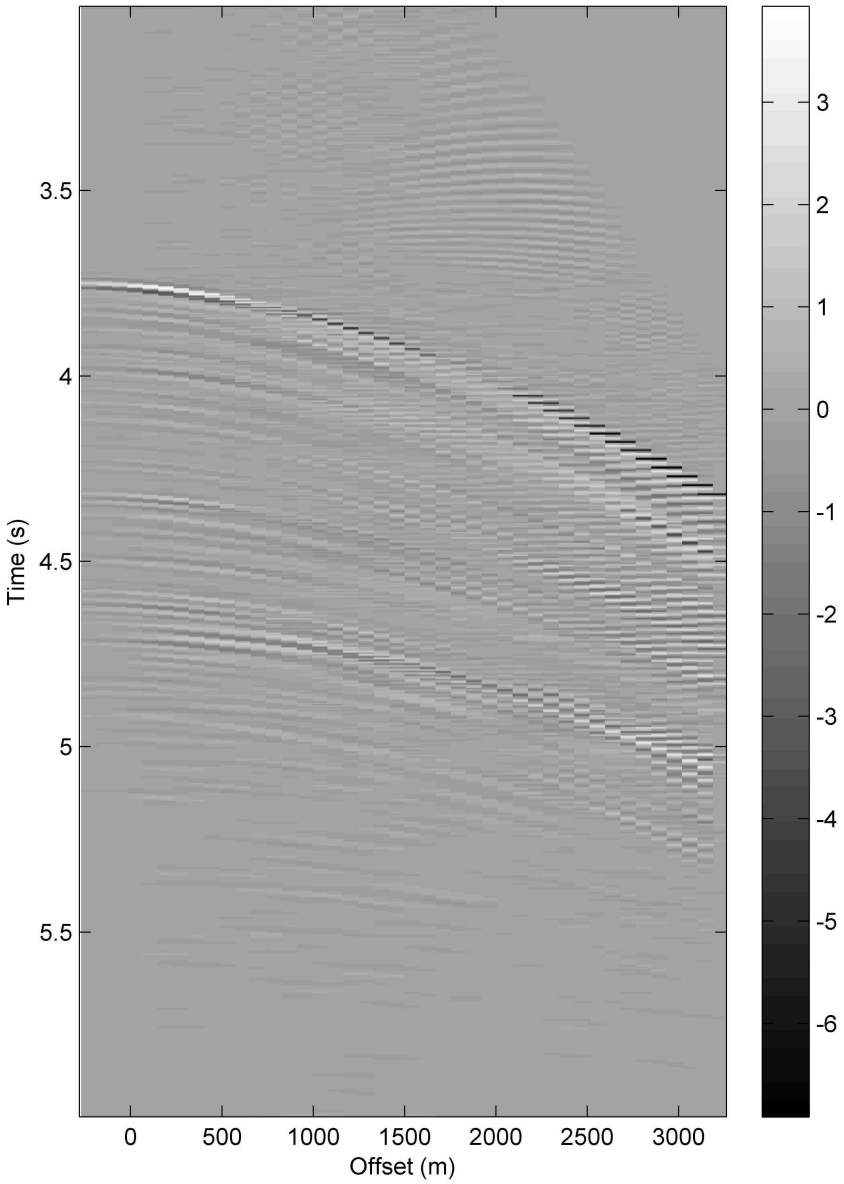

Figure 3.4: Grayscale map of the matrix of recorded amplitudes A from a real CMP gather. The scale bar indicates the intensity of the amplitudes. We are interested in detecting the events with the highest and lowest amplitudes that follow an approximately hyperbolic pattern in order to use them to estimate interval velocities and depths. The hyperbolic behaviour can be approximated by Eq. 3.19.

are normally distributed with means and variances chosen according to the expert's knowledge of the area and truncated according to Dix equation.

Since it is unlikely that prior information for each $\Sigma_{i\cdot}^{(t)}$ would be available, we

have to estimate $\Sigma_{i\cdot}^{(t)}$ so that it can reflect the interdependency of travel times within layers and traces. A crude way to do so is by using a subset of traces in the gather to estimate the sample covariance matrix and the remaining for the inference steps or by using another gather at the neighborhood. Another solution would be to associate a hyperprior to $\Sigma_{i\cdot}^{(t)}$, say a Wishart distribution, and continue the inference using all traces. In this case, the first method is preferred simply because it is much faster than the latter.

In summary,

$$\begin{aligned} E_{i\cdot}^{(t)} &\sim N\left(0, \Sigma_{i\cdot}^{(t)}\right) \\ T_{i\cdot}^{(r)} | \Delta T_{0_i}, \Delta V_{rms_i} &\sim N\left(\mu_{i\cdot}^{(t)}, \Sigma_{i\cdot}^{(t)}\right) \\ \Delta T_{0_i} &\sim N\left(\mu_{t_{0_i}}, \sigma_{t_{0_i}}^2\right),\ i \in \{1, \ldots, n\} \\ \Delta V_{rms_i} &\sim N\left(\mu_{v_i}, \sigma_{v_i}^2\right),\ i \in \{1, \ldots, n\} \end{aligned} \tag{3.30}$$

leading to the posterior

$$\pi\left(v_{rms}, t_0 | t^{(r)}, x\right) = \frac{\pi(t^{(r)}, x | v_{rms}, t_0)\pi(v_{rms}, t_0)}{\pi(t^{(r)}, x)} \tag{3.31}$$

where $\pi(t^{(r)}, x | v_{rms}, t_0)$ is the likelihood function given by

$$\pi(t^{(r)}, x | v_{rms}, t_0) = \prod_{i=1}^{n} (2\pi)^{-\frac{m}{2}} \left|\Sigma_{i\cdot}^{(t)}\right|^{-\frac{1}{2}} \exp\left(-\left(t_{i\cdot}^{(r)} - \mu_{i\cdot}^{(t)}\right)' \left(2\Sigma_{i\cdot}^{(t)}\right)^{-1} \left(t_{i\cdot}^{(r)} - \mu_{i\cdot}^{(t)}\right)\right) \tag{3.32}$$

for $t_{i\cdot}^{(r)} = \left[t_{ij}^{(r)}\right]_{j=1,\ldots,m}$ and $\mu_{i\cdot}^{(t)} = \left[\mu_{ij}^{(t)}\right]_{j=1,\ldots,m}$ for $\mu_{ij}^{(t)}$ as defined in Equation 3.29; $\pi(v_rms, t_0)$ is the prior distribution given by

$$\begin{aligned} \pi(v_{rms}, t_0) = \pi\left(\Delta v_{rms}, \Delta t_0\right) &= \prod_{i=1}^{n} \pi\left(\Delta v_{rms_i}, \Delta t_{0_i}\right) \\ &= \prod_{i=1}^{n} \pi\left(\Delta v_{rms_i}\right) \pi\left(\Delta t_{0_i}\right) \\ &= \prod_{i=1}^{n} \left(2\pi\sigma_{t_{0_i}}^2\right)^{-1/2} \exp\left(-\frac{\left(\Delta t_{0_i} - \mu_{t_{0_i}}\right)^2}{2\sigma_{t_{0_i}}^2}\right) \\ &\times \prod_{i=1}^{n} \left(2\pi\sigma_{v_i}^2\right)^{-1/2} \exp\left(-\frac{\left(\Delta v_{rms_i} - \mu_{v_i}\right)^2}{2\sigma_{v_i}^2}\right) \end{aligned} \tag{3.33}$$

and $\pi(t^{(r)}, x)$ is the normalizing constant which can be computed numerically.

Here we are only interested in producing fast estimates for stacking velocities so we will not separate modeling and recording errors.

3.3.2 1D Semi-Inverse BRAINS

In this model we have a nonlinear regression of travel times as a function of RMS velocities and depths on a single gather which is the isotropic model in (Caiado et al., 2011). We assume that depths and interval velocities are independent within layers; if we write

$$T^{(r)} = T + E^{(r)} \tag{3.34}$$

and take $E_{ij}^{(r)}$, to be normally distributed with zero mean and variance $\sigma_{t_{ij}}$, $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$ then $T_{ij}^{(r)}$ is normally distributed with mean

$$\mu_{t_{ij}} = \frac{1}{v_{rms_i}} \sqrt{4z_i^2 + x_j^2} \tag{3.35}$$

and standard deviation

$$\sigma_{t_{ij}} = q_i \mu_{t_{ij}}. \tag{3.36}$$

In the original model, Q_i is assumed to be Beta-distributed with shape parameters α_i and β_i, which should be set using expert advice, e.g. the expert believes that the standard deviation of travel times is around 1% of their mean with an error of 0.5% then we can estimate α_i and β_i using the mean and standard deviation for a Beta variable, in this example, $\alpha_i = 0.0296$ and $\beta_i = 2.9304$. The RMS velocity of the first layer is assumed to be normally distributed with mean μ_{v_1} and standard deviation σ_{v_1} and we also assume that

$$V_{rms_i} = V_{rms_{(i-1)}} + S_{v_{(i-1)}} D_{v_{(i-1)}},\ i \in \{2, \ldots, n\} \tag{3.37}$$

where $D_{v_{(i-1)}}$, the magnitude of the inter-layer RMS velocity difference, is gamma distributed with shape $k_{d_{(i-1)}}$ and scale $\theta_{d_{(i-1)}}$; and $S_{v_{(i-1)}}$, the sign of this difference, has the following probability distribution

$$P\left(S_{v_{(i-1)}} = s_{(i-1)}\right) = \begin{cases} \gamma_i, & s_{(i-1)} = -1 \\ \delta_i, & s_{(i-1)} = 0 \\ 1 - \gamma_i - \delta_i, & s_{(i-1)} = 1 \end{cases}. \tag{3.38}$$

The RMS velocities are defined recursively so that the trend can be controlled since this trend is expected to be smooth and roughly increasing a priori; most times the expert has strong beliefs about what a reasonable step size for RMS velocities is and the odds of it increasing or decreasing but not much information about the velocities' actual values.

Finally we assume that the thickness of the first layer, Z_1, is normally distributed with mean μ_{z_i} and standard deviation σ_{z_i} and that

$$Z_i = Z_{(i-1)} + D_{z_{(i-1)}},\ i \in \{2, \ldots, n\} \tag{3.39}$$

where $D_{z_{(i-1)}}$ is gamma-distributed with parameters $k_{dz_{(i-1)}}$ and $\theta_{dz_{(i-1)}}$.

Defining the depths recursively guarantees that each layer has a nonnegative thickness and, combined with the recursive definition of RMS velocities, simplifies the detection of spurious events and the update of priors. For a more realistic model, we must also introduce constraints to the RMS velocities based on Dix's equation and assuring that interval velocities derived using Dix's formula are real and nonnegative.

The posterior distribution for this model is given by

$$\pi\left(v_{rms}, z, q | t^{(r)}, x\right) = \frac{\pi(t^{(r)}, x | v_{rms}, z, q)\pi(v_{rms}, z, q)}{\pi(t^{(r)}, x)} \tag{3.40}$$

where $\pi(t^{(r)}, x | v_{rms}, z, q)$ is the likelihood function given by

$$\pi(t^{(r)}, x | v_{rms}, z, q) = \prod_{i=1}^{n}\prod_{j=1}^{m} \frac{v_{rms_i}}{q_i\sqrt{2\pi\left(4z_i^2 + x_j^2\right)}} \exp\left(-\frac{\left(t_{ij}^{(r)} - v_{rms_i}^{-1}\sqrt{4z_i^2 + x_j^2}\right)^2}{2q_i^2 v_{rms_i}^{-2}\left(4z_i^2 + x_j^2\right)}\right) \tag{3.41}$$

for $t_{i\cdot}^{(r)} = \left[t_{ij}^{(r)}\right]_{j=1,\ldots,m}$ and $\pi(v_rms, z, q)$ is the prior distribution given by

$$\begin{aligned}
\pi(v_{rms}, z, q) &= \pi\left(v_{rms_1}, d_{v_{rms}}, s_v, z_1, d_z, q\right) = \pi\left(v_{rms_1}, z_1, q_1\right) \prod_{i=2}^{n} \pi\left(d_{v_i}, d_{z_i}, s_{v_i}, q_i\right) \\
&= \frac{1}{\sigma_{v_1}\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_{v_1^2}}\left(v_1 - \mu_{v_1}\right)^2\right) \frac{1}{\sqrt{2\pi\sigma_{z_1}}} \exp\left(-\frac{1}{2\sigma_{z_1}}\left(z_1 - \mu_{z_1}\right)^2\right) \\
&\times \prod_{i=1}^{n} \frac{q_i^{\alpha_i - 1}\left(1 - q_i\right)^{\beta_i}}{B(\alpha_i, \beta_i)} \prod_{i=2}^{n} P\left(S_{v_{(i-1)}} = s_{v_{(i-1)}}\right) \\
&\times \prod_{i=2}^{n} \frac{1}{\Gamma\left(k_{d_{(i-1)}}\right)\theta_{d_{(i-1)}}^{k_{d_{(i-1)}}}} d_{v_{(i-1)}}^{-1+k_{d_{(i-1)}}} \exp\left(-\frac{d_{v_{(i-1)}}}{\theta_{d_{(i-1)}}}\right) \\
&\times \prod_{i=2}^{n} \frac{1}{\Gamma\left(k_{dz_{(i-1)}}\right)\theta_{dz_{(i-1)}}^{k_{dz_{(i-1)}}}} d_{z_{(i-1)}}^{-1+k_{dz_{(i-1)}}} \exp\left(-\frac{d_{z_{(i-1)}}}{\theta_{dz_{(i-1)}}}\right)
\end{aligned} \tag{3.42}$$

where B is the Beta function and Γ is the Gamma function, and $\pi(t^{(r)}, x)$ is the normalizing constant which can be computed numerically.

We call this model the 1D semi-inverse BRAINS (1D SI-BRAINS) and an alternative for it is the inverse model which uses interval velocities, instead of RMS velocities, and depths in its prior determination.

3.3.3 1D Inverse BRAINS

For the 1D inverse BRAINS (1D I-BRAINS), we assume that the thicknesses of the layers ΔZ_i are independent with prior distribution $\pi_{\left(\Delta Z_i, \boldsymbol{\Theta}_i^{(z)}\right)}(\cdot)$ with parameters $\boldsymbol{\Theta}_i^{(z)}$ and that the interval velocities V_i are also independent with prior distribution $\pi_{\left(V_i, \boldsymbol{\Theta}_i^{(v)}\right)}(\cdot)$ with parameters $\boldsymbol{\Theta}_i^{(v)}$. These assumptions of statistical independency are based in logical independence and complete ignorance about the Earth's subsurface and such assumptions are expected to be reflected on the modelling error. For example, we could assume that anything below surface could be true like a layer of water followed by steel and then sand but a geologist would say that this is unlikely and that the subsurface is more likely formed by a combination of sedimentary rocks. If an expert is able to produce strong priors that contradict such independency e.g. interval velocities are expected to increase with depth, then these should be reformulated accordingly with a joint prior with a correlation structure.

Given the assumption of statistical independency, the two-way zero-offset travel time variations ΔT_{0_i} are independent with distribution

$$\pi_{\left(\Delta T_{0_i},\boldsymbol{\Theta}_i^{(z)},\boldsymbol{\Theta}_i^{(v)}\right)}(\Delta t_{0_i}) = \int_{\mathbb{R}} \left|\frac{v_i}{2}\right| \pi_{\left(V_i,\boldsymbol{\Theta}_i^{(v)}\right)}(v_i)\pi_{\left(\Delta Z_i,\boldsymbol{\Theta}_i^{(z)}\right)}\left(\frac{\Delta t_{0_i} v_i}{2}\right) dv_i. \tag{3.43}$$

Consequently the distribution of T_{0_i} is given by the convolution of $\pi_{\left(\Delta T_{0_l},\boldsymbol{\Theta}_l^{(z)},\boldsymbol{\Theta}_l^{(v)}\right)}(\Delta t_{0_l})$ for $l = 1, \ldots, i$. we can also rewrite the RMS velocity at boundary i, by combining 3.28 and 3.25, as

$$V_{rms_i} = \sqrt{\frac{\sum_{l=1}^{i} V_l \Delta Z_l}{\sum_{l=1}^{i} V_l^{-1} \Delta Z_l}}; \tag{3.44}$$

and while we cannot find an explicit density for this variable, we can sample given the distributions of V_i and ΔZ_i.

Now assume that the recording error $E^{(r)}$ is normally distributed with zero mean and covariance matrix $\Sigma^{(r)}$. If we consider that the source-receiver pairs are independent and that a path of a ray does not affect another one, we can assume that these errors are independent and any residual error is due to a modelling error. So each $E_{ij}^{(r)}$ is normally distributed with mean zero and variance $\sigma^2_{(r)ij}$.

We can assume that each $\sigma_{(r)ij}$ is proportional to its corresponding mean, i.e.

$$\sigma_{(r)ij} = q_i \mu_{(r)ij} \tag{3.45}$$

where Q_i has prior distribution $\pi_{\left(Q_i,\boldsymbol{\Theta}_i^{(q)}\right)}(\cdot)$.

Finally assume that the modelling error $E^{(m)}$ is also normally distributed with mean zero and covariance matrix $\Sigma^{(m)}$ with elements

$$\Sigma_{ij}^{(m)} = \sigma_{ij}^{(r)} \exp\left(\frac{(i-j)^2}{p_i}\right). \tag{3.46}$$

The likelihood function can be written as

$$\pi\left(\mathbf{t^{(r)}}|\mathbf{v},\mathbf{z},\mathbf{x},\mathbf{q},\mathbf{p}\right) \propto \left|\Sigma^{(t)}\right|^{-1/2} \exp\left(-\frac{1}{2}\left(t^{(r)}-\mu^{(t)}\right)'\left(\Sigma^{(t)}\right)^{-1}\left(t^{(r)}-\mu^{(t)}\right)\right) \tag{3.47}$$

and the posterior distribution given $T^{(r)}$ as

$$\begin{aligned}\pi(\mathbf{v},\boldsymbol{\Delta t_{0_i}},\mathbf{q},\mathbf{p}|\mathbf{t^{(r)}},\mathbf{x}) &\propto \pi(\mathbf{t^{(r)}}|\mathbf{v},\boldsymbol{\Delta t_0},\mathbf{x},\mathbf{q},\mathbf{p}) \\ &\times \prod_{i=1}^{n}\left[\pi_{\left(V_i,\boldsymbol{\Theta}_i^{(v)}\right)}(v_i)\pi_{\left(\Delta Z_i,\boldsymbol{\Theta}_i^{(z)}\right)}(z_i)\pi_{\left(Q_i,\boldsymbol{\Theta}_i^{(q)}\right)}(q_i)\pi_{\left(P_i,\boldsymbol{\Theta}_i^{(p)}\right)}(p_i)\right].\end{aligned} \tag{3.48}$$

Next we update the 1D S-BRAINS with a more flexible model that accounts for part of the error related to the hyperbolic assumption.

3.3.4 1D Stacking Smooth BRAINS

The 1D BRAINS models introduced in the previous three subsections are standard non-linear regression models with input values x and observed values $T^{(r)}$ connected by hyperbolic curves; for such models we assumed that the errors are normally distributed with zero-mean and a specific covariance matrix in each case. Because we are restricted to the travel time equations, these models lack flexibility and any additions to the posterior that might improve the posterior credibility interval estimates will come at a computational cost.

A solution to this problem is to work in a function space instead of a weight space. We can assume that the set of travel times related to a certain reflector in a single CMP gather is a sample of a continuous and smooth curve that has a hyperbolic trend. Assume that any finite set of travel times in this curve is a multivariate Gaussian random variable then we can assume that the travel time curve is a Gaussian process over the offset. This will not only account for part of the error created by the discretization of the traces but also the modelling error due to the lack of flexibility of the standard regression methods (Rasmussen and Williams, 2006).

For the 1D Stacking Smooth BRAINS model (1D SS-BRAINS), we have that the recorded travel-time curve for the i-th layer, $\mathcal{T}_i^{(r)}$, is a Gaussian process

$$\mathcal{T}_i^{(r)}(x)|\Delta T_{0_{(1,\dots,i)}}, \Delta V_{rms_{(1,\dots,i)}} \sim \mathcal{GP}\left(m_{t_i}(x), k_i(x,x')\right) \tag{3.49}$$

with mean function

$$m_{t_i}(x) = \left(t_{0_i}^2 + x^2 v_{rms_i}^{-2}\right)^{1/2} \tag{3.50}$$

and covariance function

$$k_i(x,x') = \sigma_{n_i} + \sigma_{s_i} \exp\left(-\frac{(x-x')^2}{d_i}\right) \tag{3.51}$$

where x and x' are two arbitrary points in the offset domain, σ_{n_i} is a noise parameter, σ_{s_i} is a scale parameter and d_i is a length parameter.

We choose the joint prior for ΔT_{0_i} and ΔV_{rms_i} as

$$\begin{bmatrix} \Delta T_{0_i} \\ \Delta V_{rms_i} \end{bmatrix} \sim N\left(\begin{pmatrix} \mu_{t_{0_i}} \\ \mu_{v_i} \end{pmatrix}, \Sigma_{(t_0, v_{rms})_i} \right). \tag{3.52}$$

and we treat σ_{n_i}, σ_{s_i} and d_i as constants that need to be set either manually or by an optimization process. Here the posterior distribution is given by

$$\pi\left(v_{rms}, t_0 | t^{(r)}\right) = \pi(v_{rms}, t_0) \prod_{i=1}^{n} \int_X \frac{\pi(t_i^{(r)}(x) | \Delta t_{0_{(1,\ldots,i)}} \Delta v_{rms_{(1,\ldots,i)}})}{\pi(t^{(r)}(x))} dx \tag{3.53}$$

where

$$\pi\left(t_i^{(r)}(x) | v_{rms_i}, t_{0_i}\right) = \pi\left(t_i^{(r)}(x) | \Delta t_{0_{(1,\ldots,i)}}, \Delta v_{rms_{(1,\ldots,i)}}\right) \tag{3.54}$$

is the likelihood of the Gaussian process in Equation 3.49 and $\pi(v_{rms}, t_0)$ is the prior distribution given by

$$\pi\left(v_{rms}, t_0\right) = \prod_{i=1}^{n} \pi\left(\Delta t_{0_i}, \Delta v_{rms_i}\right) \tag{3.55}$$

where $\pi\left(\Delta t_{0_i}, \Delta v_{rms_i}\right)$ is the density of the Normal distribution in Equation 3.52.

This concludes the single-gathered BRAINS models, next we introduce their multi-gathered analogues.

3.4 Multi-Gathered BRAINS

In the previous Section, we proposed the 1D stacking BRAINS (1D S-BRAINS) which estimates zero-offset travel times and RMS velocities, the 1D semi-inverse BRAINS (1D SI-BRAINS) which estimates RMS velocities and depths and the 1D inverse BRAINS (1D I-BRAINS) which estimates interval velocities and depths all for a single-gathered dataset. We also proposed the 1D smooth stacking BRAINS (1D SS-BRAINS) which generalizes the 1D S-BRAINS model and accounts for part of the discretization error. Next we generalize the 1D SS-BRAINS and the 1D I-BRAINS to multi-gathered problems.

3.4.1 Stacking Smooth BRAINS: Generalizing the 1D SS-BRAINS

When dealing with synthetic and real examples, we will see that the previous models work well when we need to optimize the stacking transformation for a single gather but, in multiple-gathered setups, they do not reflect the inter-gather correlation structure. Based on the assumption of homogeneity, we could say that the RMS velocity across gathers for the same interface is constant; however, this is a strong assumption and we should assume that it is valid locally but not globally, therefore, instead of assuming a constant velocity, we can assume that the RMS velocity as a function of the CMP positions is continuous and smooth. If we take the RMS velocity across CMPs to be a Gaussian process, we can generalize the 1D SS-BRAINS model for the multi-gathered setup accounting for the inter-gather variability.

We define $T_i^{(r)}(x, x_c)$ as the recorded two-way travel time of a ray reflected from the i-th layer corresponding to a source-receiver pair separated by an offset x and with a common midpoint position x_c. If we are analysing a sequence of CMP gathers in a line, x_c is a scalar and, if our CMP gathers are derived from a grid of sources and receivers, x_c is a vector of coordinates. Here we will focus on CMP gathers in a line but the method naturally extends to gridded profiles. Say $T_{0_i}(x_c)$ is the two-way zero-offset travel time for the i-th layer in the gather with CMP position x_c and $\Delta T_{0_i}(x_c)$, the two-way zero-offset travel time increment for this same layer. Similarly we define $V_{rms_i}(x_c)$ as the RMS velocity for the i-th layer in the gather with CMP position x_c and $\Delta V_{rms_i}(x_c)$ as the RMS velocity increment for this same layer. Then we assume that

$$\mathcal{T}_i^{(r)}(x, x_c)|\Delta T_{0_{(1,\dots,i)}}(x_c), \Delta V_{rms_{(1,\dots,i)}}(x_c) \sim \mathcal{GP}\left(m_{t_i}(x, x_c), k_i(x, x', x_c)\right) \tag{3.56}$$

with mean function

$$m_{t_i}(x, x_c) = \left(t_{0_i}(x_c)^2 + x^2 v_{rms_i}(x_c)^{-2}\right)^{-1/2} \tag{3.57}$$

and covariance function

$$k_i(x, x', x_c) = \sigma_{n_i}(x_c) + \sigma_{s_i}(x_c) \exp\left(-\frac{(x - x')^2}{d_i(x_c)}\right) \tag{3.58}$$

where x and x' are two arbitrary points in the offset domain, $\sigma_{n_i}(x_c)$ is a noise parameter, $\sigma_{s_i}(x_c)$ is a scale parameter and $d_i(x_c)$ is a length parameter. For an ideal dataset recorded in an area where an expert believes that the subsurfaces are approximately homogeneous and isotropic, the parameters σ_{n_i}, σ_{s_i} and d_i can be treated as constants across gathers or small sets of gathers eliminating the dependency on x_c.

Now if we assume that the RMS velocity increment as a function of x_c follows a Gaussian process, we can choose its prior to be

$$\Delta V_{rms_i}(x_c) \sim \mathcal{GP}\left(m_v(x_c), \sigma_{nv_i} + \sigma_{sv_i} \exp\left(-\frac{(x_c - x_c')^2}{d_{vi}}\right)\right) \tag{3.59}$$

where $m_v(x_c)$ is a polynomial mean function, x_c and x_c' are two points in the CMP line, σ_{nv_i} is a noise parameter, σ_{sv_i} is a scale parameter and d_{vi} is a length parameter. The mean function shape should be selected according to the level of information about the RMS velocity variation in a specific layer across gathers; a constant or low-order polynomial mean should suffice in most cases where the only assumption is smoothness.

Similarly, we can expect the zero-offset travel time increments to vary smoothly across gathers so we choose an equivalent prior

$$\Delta T_{0_i}(x_c) \sim \mathcal{GP}\left(m_{t0}(x_c), \sigma_{nt_i} + \sigma_{st_i} \exp\left(-\frac{(x_c - x_c')^2}{d_{ti}}\right)\right) \tag{3.60}$$

where $m_{t0}(x_c)$ is a polynomial mean function, x_c and x_c' are two points in the CMP line, σ_{st_i} is a scale parameter and d_{ti} is a length parameter. Usually the expert is able to produce sequences of zero-offset travel time picks which are obtained by visual inspection and manual picking using a stack of gathers; if that is the case, we can use this information to update our prior.

In this model, we choose to assume that the RMS velocity and zero-offset travel time increments for a given layer are independent a priori; there was no noticeable gain when using a joint prior since the likelihood tends to account for most of the information in the posterior. As this model can cope with lateral variation, we drop the 1D prefix and call this model Stacking Smooth BRAINS or SS-BRAINS. Next we present our last model, the Inverse Smooth BRAINS.

3.4.2 The Inverse Smooth BRAINS method: inverse velocity modeling

Estimating stacking velocities is part of the standard seismic data analysis process and is used as an intermediary process before a time or depth migration with the objective of improving the lateral resolution and ultimately predicting depths. Here we propose a model that estimates depths and interval velocities for a sequence or grid of gathers within our modeling constraints. First we need to redefine our variables and update their notation in order to account for the CMP position of each gather; let $\Delta Z_{0_i}(x_c)$ be the thickness of the i-th layer for a gather with CMP at x_c and let $V_{int_i}(x_c)$ be the interval velocity in this same layer. As in the SS-BRAINS model, we start by assuming that the recorded travel times of the i-th layer in a gather with CMP at x_c is a Gaussian process described as follows

$$\mathcal{T}_i^{(r)}(x, x_c)|\Delta Z_{(1,\dots,i)}(x_c), V_{int_i}(x_c) \sim \mathcal{GP}\left(m_{z_i}(x, x_c), k_{z_i}(x, x', x_c)\right) \tag{3.61}$$

with mean function

$$m_{z_i}(x, x_c) = \frac{1}{v_{rms_i}(x_c)}\left(4z_i(x_c)^2 + x^2\right)^{1/2} \tag{3.62}$$

with v_{rms_i} as in Equation 3.44 and covariance function

$$k_{z_i}(x, x', x_c) = \sigma_{n_i}(x_c) + \sigma_{s_i}(x_c)\exp\left(-\frac{(x - x')^2}{d_i(x_c)}\right) \tag{3.63}$$

where x and x' are two arbitrary points in the offset domain, $\sigma_{n_i}(x_c)$ is a noise parameter, $\sigma_{s_i}(x_c)$ is a scale parameter and $d_i(x_c)$ is a length parameter.

Assuming we are tracking a single lithological unit, the I interval velocity, as a function of x_c, is expected to be continuous and smooth; thickness is expected to follow the same type of behaviour but could present discontinuities. We can assume that, a priori, the interval velocities and the thicknesses follow Gaussian processes with polynomial mean functions

$$\Delta Z_i(x_c) \sim \mathcal{GP}\left(m_z(x_c), \sigma_{nz_i} + \sigma_{sz_i}\exp\left(-\frac{(x_c - x_c')^2}{d_{zi}}\right)\right) \tag{3.64}$$

and

$$V_{int_i}(x_c) \sim \mathcal{GP}\left(m_{v_{int}}(x_c), \sigma_{nv_int_i} + \sigma_{sv_int_i} \exp\left(-\frac{(x_c - x'_c)^2}{d_{v_int_i}}\right)\right) \tag{3.65}$$

where x_c and x'_c are two points in the CMP line, σ_{nz_i} and $\sigma_{nv_int_i}$ are noise parameters, σ_{sz_i} and $\sigma_{sv_int_i}$ are a scale parameters and d_{zi} and $d_{v_int_i}$ are length parameters. The covariance and mean functions for the process that describes $\Delta Z_i(x_c)$ should be modified to reflect the expert's beliefs about the presence of discontinuities either by assigning independent priors for each gather or by using a piecewise mean function. We call this model the Inverse Smooth BRAINS (IS-BRAINS).

These six models form the current BRAINS toolbox and each model is aimed at a different stage of seismic data analysis. Other alternatives that include information about anisotropy or dipping reflectors can be explored by modifying the mean functions of the multi-gathered models. Now we discuss a few issues in prior selection.

3.5 Interpreting prior information

Regardless of the parametrisation adopted or the data being analysed, there are a number of weak prior specifications that always hold due to physical and mathematical constraints and limitations of the Earth's geology. The velocity of a P-wave in solids depends on the material's density and bulk and shear moduli while the velocity of S-waves depends only on the shear modulus and density. In liquids, the P-wave velocity is a function of density and bulk modulus.

The bulk modulus is defined as

$$K = -V\frac{\delta P}{\delta V} \tag{3.66}$$

where V is volume and P is pressure and the shear modulus is defined as

$$G = \frac{F/S}{d/h} \tag{3.67}$$

where F is a force acting on a planar area S orthogonally to a vector of length h and d is the transverse displacement of S. The ratio F/S is called shear stress and the ratio d/h, shear strain.

In solids, the speed of P-waves is given by

$$s_{p,s} = \sqrt{\frac{K + (4/3)G}{\rho}} \tag{3.68}$$

where ρ is the density of the solid and the speed of S-waves is given by

$$s_{s,s} = \sqrt{\frac{G}{\rho}}. \tag{3.69}$$

In fluids, the speed of P-waves is given by

$$s_{p,f} = \sqrt{\frac{K}{\rho}} \tag{3.70}$$

and S-waves cannot propagate in fluids since the shear stress is null. In all three cases, we have that speed is inversely proportional to the density of a material. The density of the most common types of rock in the lithosphere range from 1 to 4 grams per cubic centimetre and the density of water is around 1 gram per cubic centimetre. Given just this piece of information, we would be led to believe that the speed of a seismic wave decreases when it reaches the ocean's seabed. This is not the case since the shear and bulk moduli dominates the speed of a wave.

The speed of seismic waves in seawater usually varies between 1450 and 1500 metres per second; temperature and depth are the main factors responsible for this variation. It is common to see higher velocities in deep waters than in shallow waters but more variation in shallow waters due to changes in the atmosphere above. The highest speeds occur in lower crustal rocks and salt while the lowest occur in clays; the P-wave velocity usually has an upper boundary of around 8000 metres per second and most media in the Earth's subsurface have speeds greater than the water layer.

Anomalies in speed are usually caused by bodies of salt, porous rocks saturated with gas or fluids, or basalt. In the first two cases, the scatter and attenuation of seismic waves are higher because of the geometry and composition of these materials. In the case of salt and basalts, it becomes harder to record velocities below these bodies since most of the energy is reflected or lost when passing through these media. For saturated porous rocks, there is usually a decrease in speed that is proportional to the level of saturation.

Apart from these cases, the interval and RMS velocities of seismic waves tend to increase with depth. The RMS velocities usually vary smoothly and the interval velocities in small steps. This is the first piece of prior information that needs to be quantified either as a boundary condition, by assigning prior distributions to the hyperparameters of the distributions of the interval velocities or by redefining the model in terms of velocity variation. We will discuss these options in the next Chapter.

Other priors can be defined before the data collection. When proposing a seismic survey, the experts have a certain depth target for exploration that is then converted to a maximum recording time. For example, assume that we want to determine the depth and velocity of a boundary about 5 kilometres deep and the experts think that the RMS velocity at that boundary does not exceed 2000 metres per second. Then it is necessary to record more than 2.5 seconds of data to account for the hyperbolic moveout. In one of our examples with real data we have just under 13 seconds of recorded data which sets our maximum two-way zero-offset time at that time. We can also infer a maximum depth according to the experts' prior beliefs about the geology of that Section.

More specific priors and information about their shapes can be obtained according to the experts' knowledge of the area. We will comment on these priors and the robustness of the model when we start analysing synthetic and real datasets.

Chapter 4

Picking BRAINS: preparing the data for analysis

In the previous Chapter, we proposed a family of models used to estimate RMS velocities, zero-offset travel times, interval velocities and depths in one or more gathers. All models assumed that the events corresponding to a certain reflector were identified, sorted and stored in a matrix of recorded travel times. Here we propose a method to execute this task and introduce tools and algorithms that we will use in the following Chapters to analyse synthetic and real datasets.

4.1 Trace Analysis

Given a set of picked travel times, the estimation of the variables defined in the previous chapter depends mainly on the complexity of the posterior distribution. However, selecting this required set of travel times from a gather manually is time consuming and adds a different type of uncertainty associated to the expert's accuracy.

During the manual picking process, the expert looks for similarities on each trace in order to match these points. In Figure 4.1, we have a sequence of 10 traces extracted from a synthetic CMP gather. The sequence of disturbances on these traces at around 2 seconds comes from a horizontal boundary of an isotropic interface with constant speed of 1480m/s and 1480m thick. In a non-filtered real

dataset, the first sign of disturbance of the trace would indicate the real arrival time of an event. Since the amplitude of the signal at the instant of arrival is really close to noise level, instead of trying to pick that sequence of points, it is common to apply a compensation filter to convert the wavelet to zero phase, centred on the real arrival time, since it is much easier to track extremes rather than the onset. Applying such filter is common practice and this filter should account for possible time shifts derived; since the error that might result from this filter is usually within the sampling frequency, it can be ignored or assumed to be included in the modelling error. Because we are not trying to model the amplitude of the peaks but their positions, the impact of the zero-phasing filter on the magnitude of the signal is not relevant for the regression analysis. Now we present a procedure to pick the local maxima and minima of a given dataset that correspond to geologically plausible boundaries.

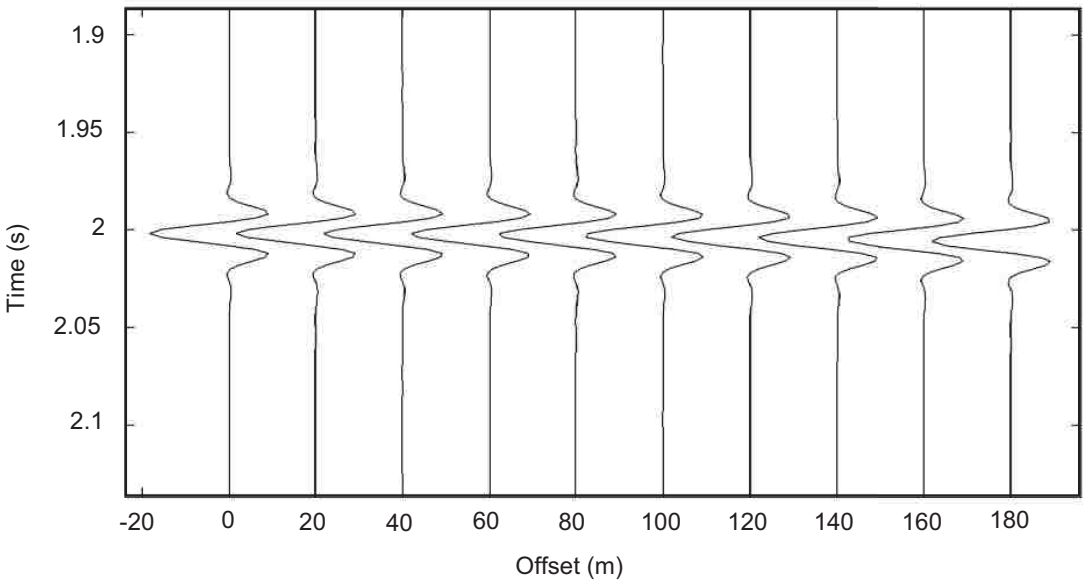

Figure 4.1: Sequence of traces from a synthetic CMP gather showing one recorded event at around 2 seconds.

If we look at the first trace in Figure 4.1 and zoom at the recorded disturbance (see Figure 4.2), we can see four local maxima and three local minima from which we can easily pick the absolute minimum marked in green and two local maxima with similar amplitude around the minimum. In all other traces of Figure 4.1,

this pattern is repeated and the local extremes have fairly similar amplitudes. Our objective is to pick the sequence of extremes with the maximum absolute amplitude; in this case, a sequence of absolute minima. In Figure 4.3, we have a grayscale map of this same CMP gather with the absolute minima (in green) and the local maxima highlighted in Figure 4.2 (in red) picked in each trace. In this case, the sequence we are looking for is the one formed by the green dots and they correspond to the reflection time of that boundary for this zero-phase wavelet. In Figure 4.4, we have the recorded amplitudes of the points picked in each sequence highlighted in Figure 4.3; note that the three sequences are fairly smooth and, combined with the idea that the green series is approximately hyperbolic in the time domain (see Eq. 3.19), we can develop a tracking system to identify events and store their corresponding recorded amplitudes and travel times for further analysis, but first we introduce the semblance, a measure commonly used in reflection seismology to identify events.

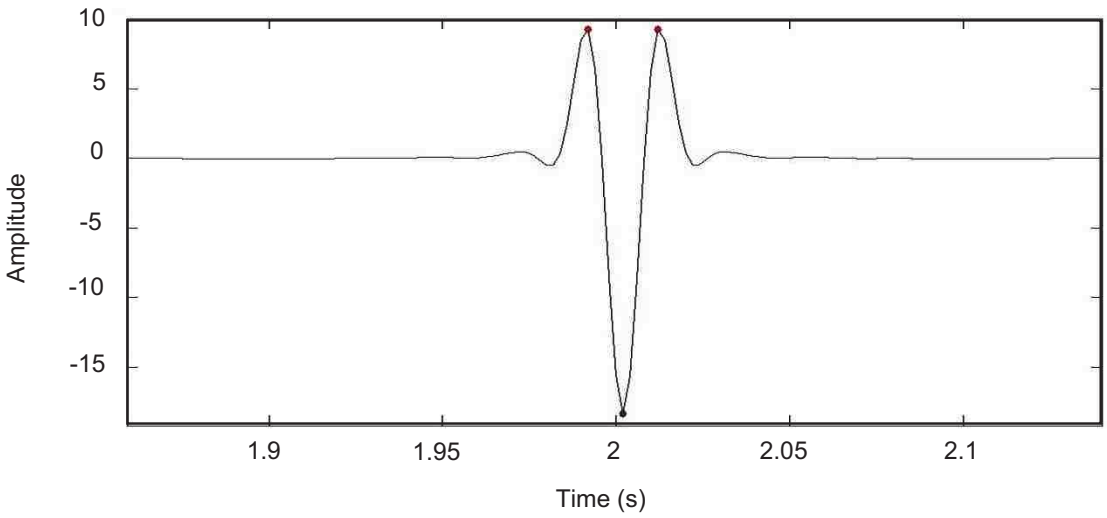

Figure 4.2: First trace of the synthetic CMP gather in Figure 4.1. The green dot marks the absolute minimum and the two red dots mark two local maxima symmetric around the absolute minimum.

4.2 Semblance Analysis

In seismic data processing, semblance is a measure used to assess the level of coherence between sequences of events on sequence of traces in a CMP gather. Given

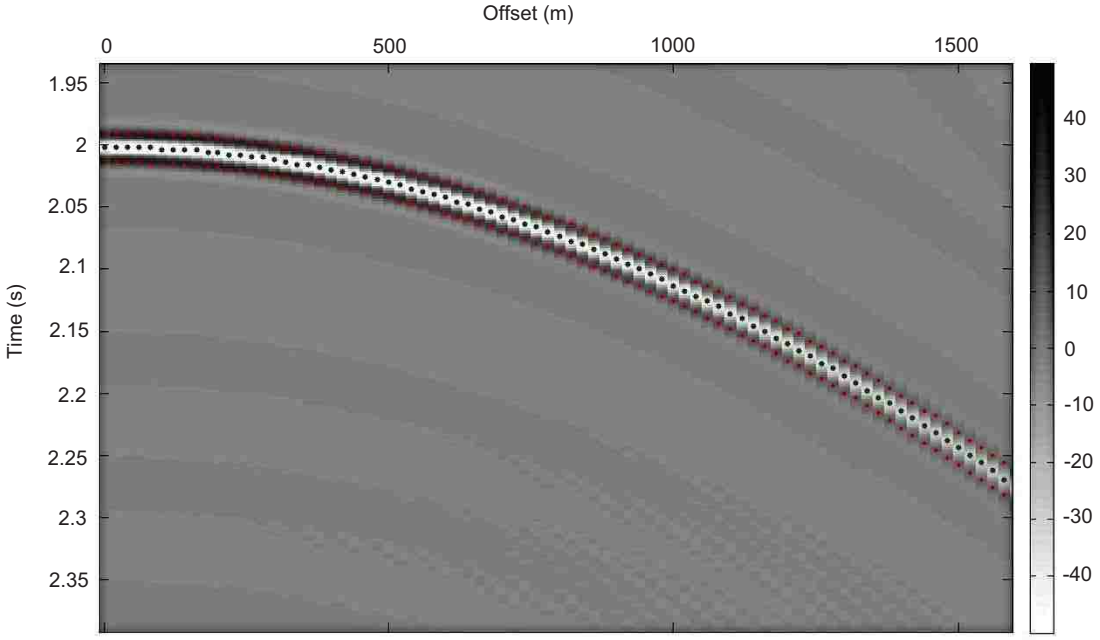

Figure 4.3: Grayscale map of the synthetic CMP gather used in Figure 4.1 with picked extremes highlighted. The scale bar on the right indicates the amplitude of the recorded signal at each point. The green dots correspond to the absolute minima and the red dots correspond to the local maxima highlighted in Figure 4.2.

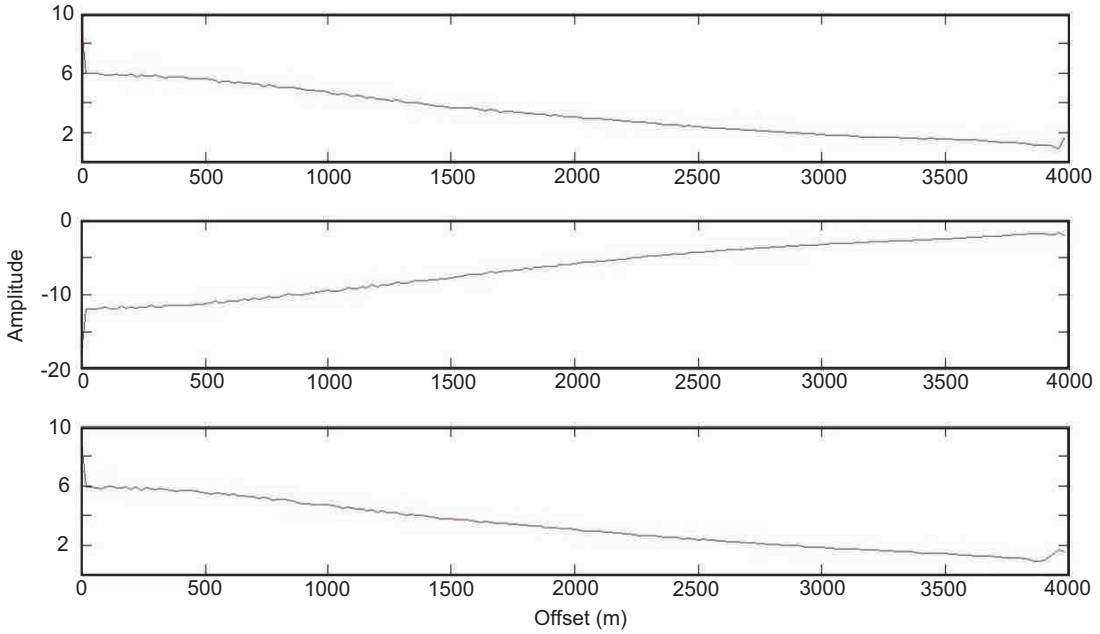

Figure 4.4: Amplitude series for the three sequence of points picked in Figure 4.3.

a set of recorded amplitudes of a sequence of picks (say the sequence of absolute minima in Figure 4.1), the semblance is defined as the quotient of the square of the average of the amplitudes and the average of their squares. Say we have n traces and at each trace i, $i \in \{1, \ldots, n\}$, we pick a point with amplitude a_i then the semblance S_a for this sequence is given by

$$S_a = \frac{\left(\sum_{i=1}^{n} a_i\right)^2}{n \sum_{i=1}^{n} a_i^2}, \tag{4.1}$$

which is a value in the interval $[0, 1]$. From a statistical point of view, maximizing the semblance value is equivalent to minimizing the variance of the amplitudes using an estimator based on the method of moments which uses sample moments to estimate population moments. Apart from the bias resulted from this estimation method, a problem with this measure is that it does not account for the order or position of the picks, i.e. the offset or the recording times so, if both red sequences in 4.3 were identical in the amplitude domain, any combination of points of these two sequences would return the same semblance value rendering this measure ambiguous in this case.

Using this measure, semblance maps can be generated for a given gather. Since the travel times are not used in the calculation of the semblance, it is necessary to fix a trend for the picks to be selected, e.g. the hyperbolic moveout (see 3.19). For the case of the hyperbolic moveout, our unknowns are the zero-offset travel time and the RMS velocity so, to generate a semblance map, we start with a regular grid of zero-offset travel times and RMS velocities. For the CMP gather represented in Figure 4.3, we produced a regular grid with zero-offset travel times varying from 1.9 to 2.1 seconds with steps of 2 milliseconds and with RMS velocities from 1450 to 1500 metres per seconds with steps of 0.1 metres per seconds. For each zero-offset travel time and RMS velocity pair in this regular grid, we pick the samples and corresponding amplitudes in the CMP gather which are closest to the trend (in this case the hyperbolic moveout) by using splines to interpolate in the amplitude domain and then we calculate the semblance for this sequence. If we repeat this calculation for each pair in the regular grid, we obtain the leftmost map in Figure

4.5 which is a raw semblance map without any local weighting or smoothing. The sequence of red areas illustrate the situation mentioned previously where multiple sequences have similar semblance values.

In Figure 4.5, the white lines indicate the position of the actual time and velocity of the reflector used to create this model, i.e. zero-offset travel time at 2 seconds and RMS velocity of 1480 metres per second. As we would expect, the semblance for the sequences of local maxima and minima is high and it is hard to identify the region that contains the absolute extreme. In fact, the sequence containing the absolute extreme does not hold the maximum semblance value in the map. In the centre, we display a map of the mean amplitudes of each sequence used to create the semblance plot and we can see that the sequence we are looking for is formed by a sequence of absolute minima (blue region). If we multiply these two maps, we can generate a weighted semblance map (right) and easily identify the region that contains the sequence corresponding to the real parameters of the reflector (dark red).

Weighted semblance maps are commonly used by experts during manual picking to identify possible reflectors and, while this tool lacks precision, it provides a method to produce estimates for time and velocity. In our case, we can use this strategy to identify search regions for the picking method described in the next Section. In the case of multi-layered models, the final map should be generated using local weighting to enhance weak reflectors and areas of low signal to noise ratio. While weighting and smoothing might solve this problem partially, a semblance map still relies on a fixed trend to be chosen beforehand like the hyperbolic moveout; there is no room for modelling errors and it strongly relies on its grid. For these reasons, we need a method capable of tracking points in both time and amplitude domains. Therefore, we introduce a more robust strategy in the next Sections.

4.3 Searching in calm waters

In the following Sections, we propose a search method that identifies sequences of local extrema that may correspond to a real reflector. We start with a strategy for noise-free or low noise data and, in the following Section, discuss how to approach

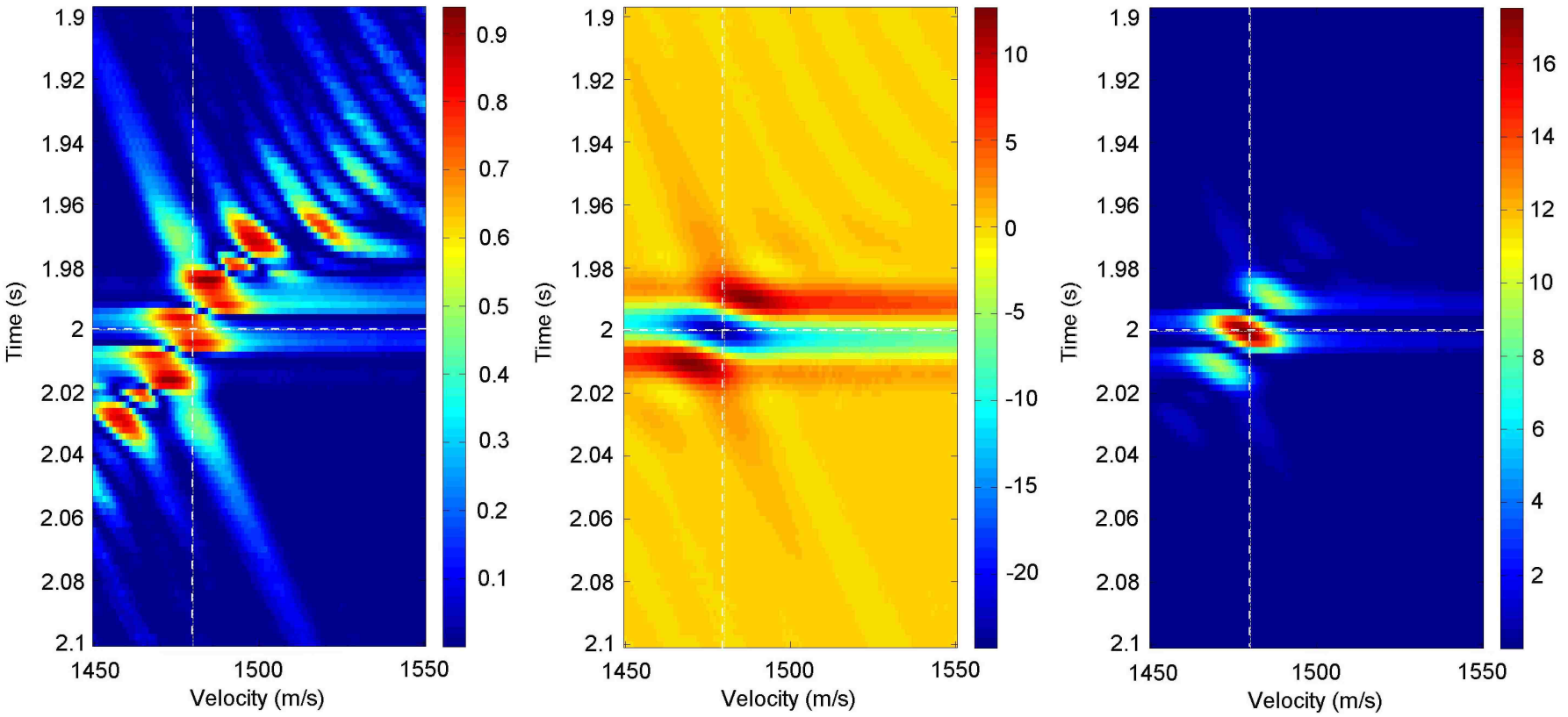

Figure 4.5: On the left, raw semblance map generated for the synthetic gather in Figure 4.3, here the colour bar represents semblance. In the centre, map of the mean amplitudes (represented by the coulour bar) of the sequences used to generate the semblance map on the left. On the right, semblance plot reweighted using the map in the centre, the colour bar indicates the product of the semblance by the absolute mean amplitude.

noisy and real data. As shown in the previous Section, given a CMP gather processed using a zero-phase filter, we need to identify a sequence of local extrema for each possible reflector. We introduced the concept of semblance maps and showed how a weighted map can help identify regions that contain a candidate. Depending on the expert confidence on the quality of the semblance maps and knowledge of the area, a minimum semblance value can be set for the whole gather or for specific blocks. The introduction of semblance maps is not necessary to aid the picking of events but it can speed the process up by reweighting unlikely regions. Note that this is a pre-prior setup and, while the data is used to produce the semblance map, the maps are only used as a guide for the search algorithm. When expressing confidence about the semblance map, one must be careful not to impose a search area that limits the functionality of the uncertainty analysis. Areas that contain a reflector normally produce higher semblances and conversely areas of low semblance are unlikely to contain an event but that does not imply that one should derive point estimators (e.g. a zero-offset travel time and RMS velocity pair) directly from these maps or deduce priors; instead we can use them to reduce the posterior space.

We make two simple assumptions in order to build a picking algorithm. The first is that, in the time domain, a set of points corresponding to a certain event lie around an approximately hyperbolic curve as proposed in Equation 3.19. Second we can assume that the amplitudes follow a smooth trend. From Equation 3.19, we have that for a small offset-velocity quotient, the two-way zero-offset term dominates the approximation and, for a small offset-velocity quotient, the slope of the tangent line is close to zero. Moreover, if we look at the square of two-way travel times as a function of the squared offset, we have an approximately linear fit.

In Figure 4.6, we have a three-layered isotropic synthetic CMP gather with parallel boundaries. Here we have a 2048×200 matrix of recorded amplitudes, A, formed by 200 traces recorded at a sampling rate of $2ms$ for 4.096 seconds. The source-receiver pairs are symmetrically positioned around a common midpoint each $20 \cdot (i-1)$ metres apart, $i \in \{1, \ldots, 200\}$. On the left, we have a grayscale map in the usual time versus offset scale where we can see three distinct curves that resemble hyperbolas and, on the right, the same dataset transformed to the t^2 versus x^2

domain where these curves appear to have been linearized.

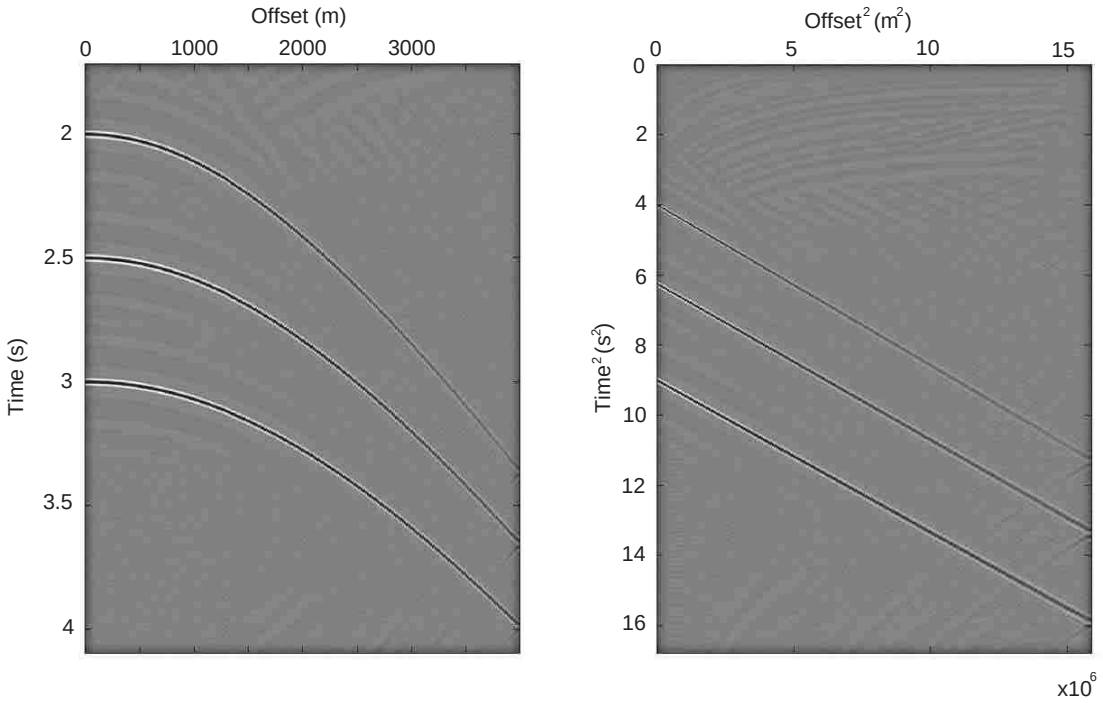

Figure 4.6: Left: three-layered isotropic synthetic CMP gather. Right: the same CMP gather as the one on the left in the $t^2 - x^2$ plot.

If we assume that the number of layers is unknown and no prior information about the layers' location is given, we start a search in A for possible events. First we select all local maxima and minima for each trace independently and set all other points on the trace to zero creating matrix A^* of local extremes. This process will help reduce the number of points to filter and facilitate the matching process. Another option is to use the semblance maps to find the regions in the gather that contain the points with local extremes. In Figure 4.7, we have the semblance map, the average amplitude map and the weighted semblance map for this three-layered model. Note that the regions of maximum semblance contain the points with the reflectors' real travel times and velocities. At this stage, a minimum absolute peak height and minimum peak distance can be set but that is not necessary. In Figure 4.8, we have a plot of the same CMP gather with a scatter of extremes picked from each trace individually; a median filter was used to remove background noise, i.e. say A_w is a $w \times w$ submatrix of A for an odd w then replace the centre point of A_w

with the median of the values in A_w.

Now we return to the extremes picked in the first trace; we want to match each of them to a point of the set of extremes from the second trace. For a given point in the first trace, we need to search for the point with the closest amplitude in the second trace that is in the neighbourhood of such point. For points in near offsets, the travel time difference between corresponding peaks will be extremely small and, as we move further from the CMP, the curvature of the hyperbolic curve will become more evident. From Equation 3.19, we have

$$t_{ij} \approx \sqrt{t_{0_j}^2 + \frac{x_i^2}{v_{rms_j}^2}} \tag{4.2}$$

for traces $i \in \{1, \ldots, n\}$ and layers $j \in \{1, \ldots, m\}$. Here $n = 200$ and m should be 3. If we apply a two-term Taylor expansion in x_i around 0 then we have

$$t_{ij} \approx t_{0_j} + \frac{x_i^2}{2t_{0_j} v_{rms_j}^2}. \tag{4.3}$$

We can use this approximation to create a search window for near offsets when v_{rms_j} is large. So given an extreme in the first trace, assume that its travel time is a good approximation for the zero-offset two way travel time at that layer candidate. Also assume that we can set a global minimum RMS velocity, v_{min}, then using Equation 4.3 we have

$$\Delta t_{ij} = t_{ij} - t_{(i-1)j} \approx \frac{1}{2t_{0_j} v_{rms_j}^2} \left(x_i^2 - x_{(i-1)}^2\right) \leq \frac{t_{0_j}}{v_{min}^2} \left(x_i^2 - x_{(i-1)}^2\right). \tag{4.4}$$

Since the travel time sequence for a given layer is increasing with offset, a search window with width $2w_{ij}$, where

$$w_{ij} = \frac{1}{2t_{0_j} v_{min}^2} \left(x_i^2 - x_{(i-1)}^2\right), \tag{4.5}$$

centred at $t_{(i-1)j}$ should be wide enough to locate the best candidate for t_{ij} in the following trace for near offsets under the assumption that $t_{0_j} \approx t_{1j}$. Note that if prior information is available about RMS velocities lower and upper boundaries for a specific layer, this window should be modified to speed up the search.

In our example, we assume that v_{min} is 1000m/s, and if we use the semblance maps as a support, we can increase v_{min} to 1450m/s since the semblance for lower

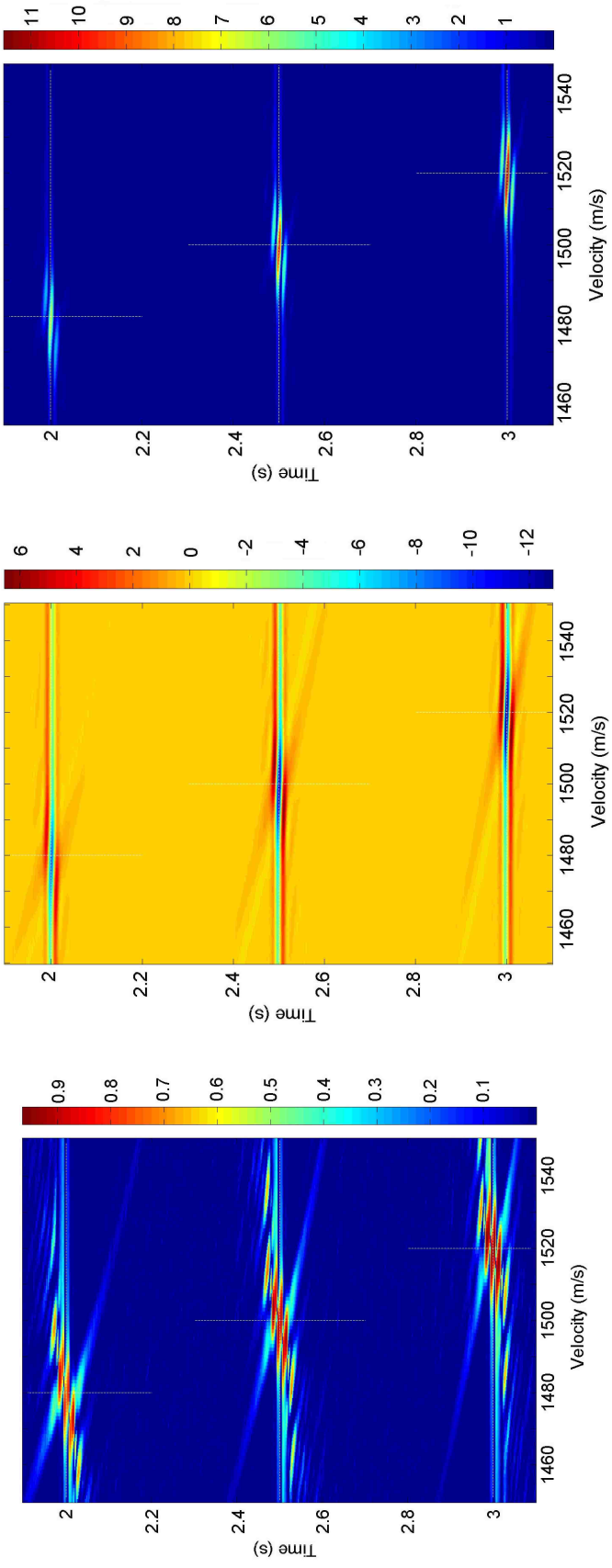

Figure 4.7: On the left, raw semblance map generated for the synthetic gather in Figure 4.6, the colour bar indicates the raw semblance value. In the centre, map of the mean amplitudes of the sequences used to generate the semblance map on the left, the colour bar represents the mean amplitude value. On the right, semblance plot reweighted using the map in the centre, the colour bar represents the product of the semblance by the absolute mean amplitude value.

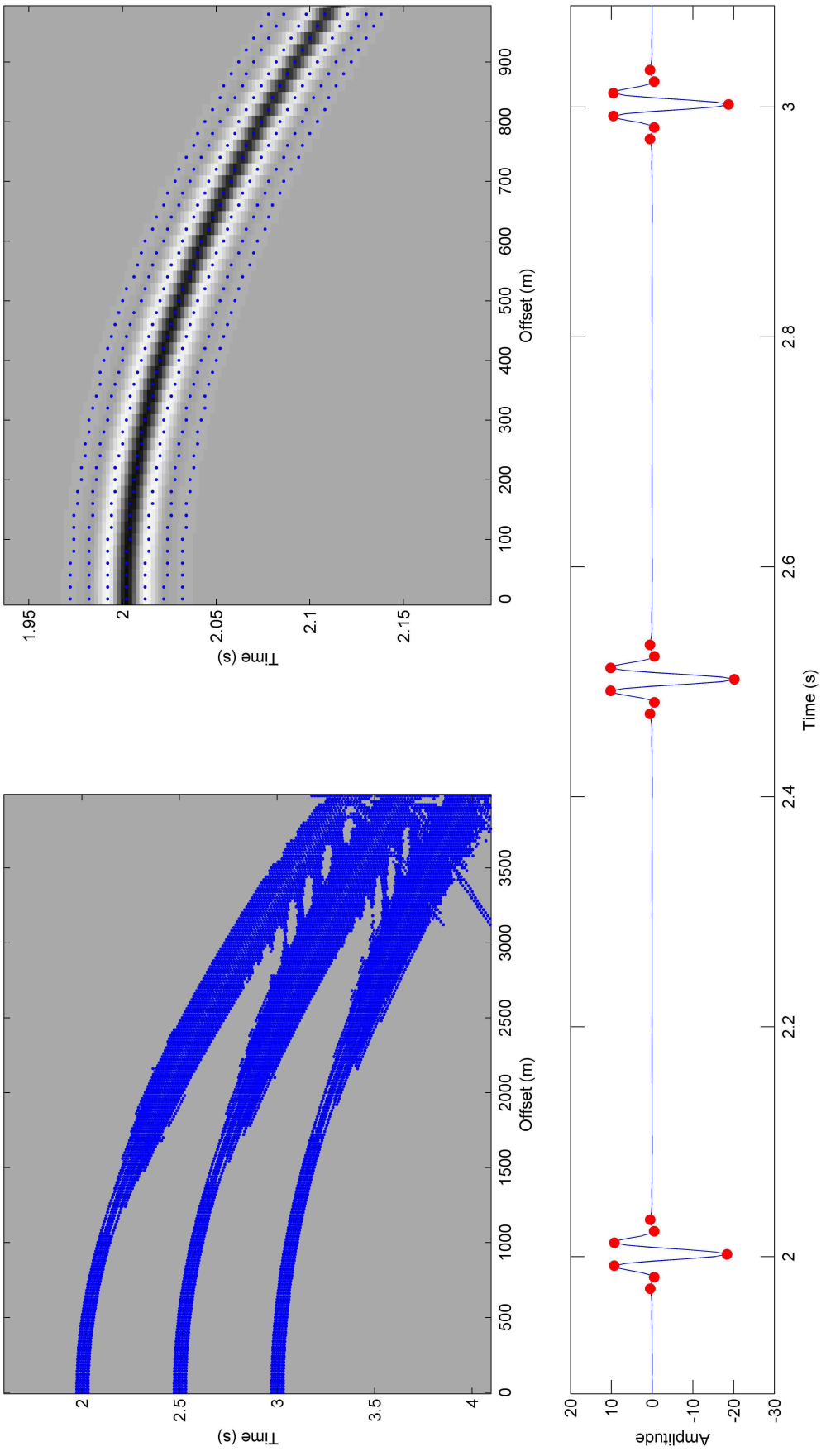

Figure 4.8: CMP gather from Figure 4.6 with scatter of picked extremes on the left and a zoomed in section of the same on the right. The bottom plot shows the picked extremes in the first trace.

velocities is negligible. We use the search window for the first 50 traces then, given a travel time in a trace i, we set a search window around that time and pick the point with the closest amplitude. By doing so, we obtain the curves displayed in Figure 4.9.

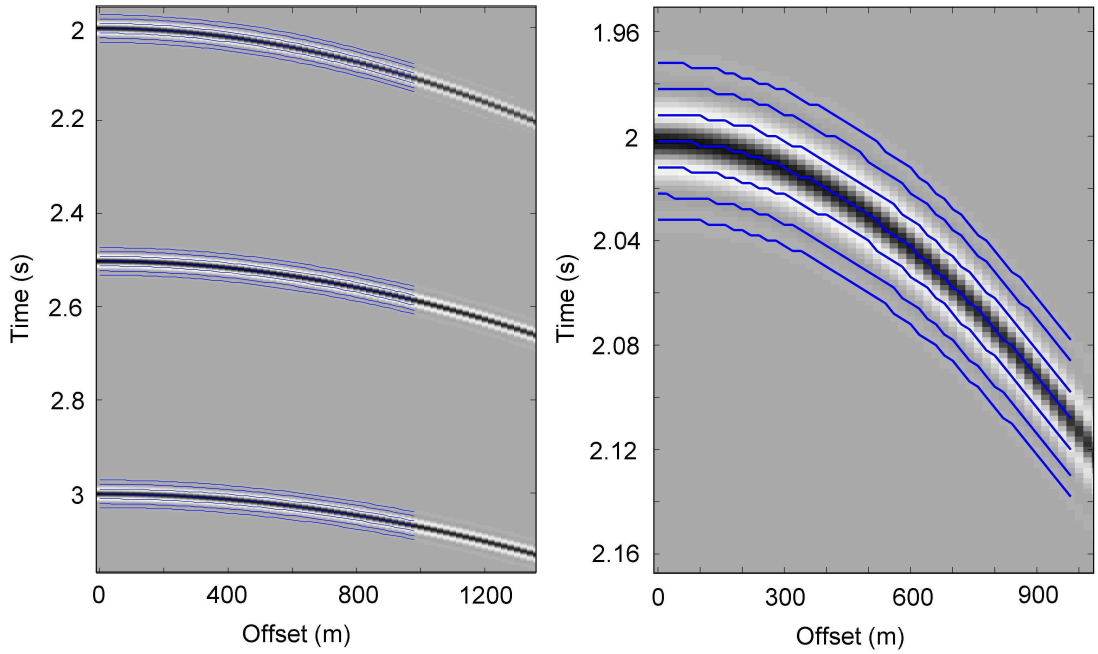

Figure 4.9: Three-layered isotropic synthetic CMP gather with layer candidates obtained using the first phase of the picking algorithm.

We started with 2048 candidates which is the same as the the number of samples in the first trace. After searching for the local maxima and minima in the first trace and using the search window mentioned above, we are left with 21 layer candidates that seem to be clustered in groups of seven and we want to single out one layer in each cluster. From each cluster, we want the sequence with the highest absolute amplitude. In Figure 4.10, we have a scatter of the extremes picked in the first trace against the absolute accumulated amplitude. It is fairly simple to pick the three local maxima.

After spotting the relevant sequences, we can either proceed to pick points in the sequence using the t^2-x^2 transformation, use the information in the semblance maps to eliminate low semblance sequences, or fit the model presented in the previous chapter to the points obtained so far to estimate the velocity and zero-offset travel

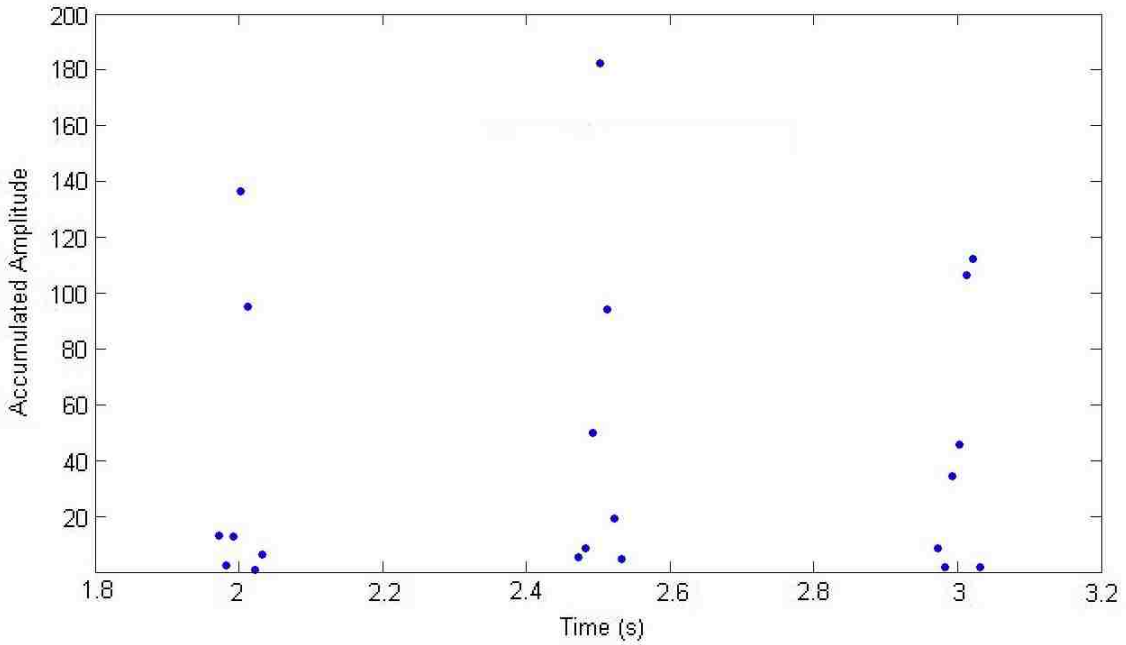

Figure 4.10: Scatter of picked extremes in the first trace against corresponding absolute accumulated amplitude.

time ranges for each layer and use that information to update the search window in 4.5 for the remaining points. In the next chapter, we combine this algorithm with the model introduced in the previous chapter to explore this synthetic gather and others. In this case, the three highest points in Figure 4.10, correspond to the first points in the reflectors' sequence of extremes and the sequences that contain these points to the points of highest semblance on the weighted semblance map. In the case of noise-free synthetic gathers with fairly sparse reflectors, this will always be true and the points of maximum semblance will be close to the real travel times and velocities apart from situations with high anisotropy.

Therefore, if used with caution and under similar circumstances as the ones described in the previous paragraph, weighted semblance maps may be used as a guide to reduce the search region. In cases where reflectors are too close to each other, semblance maps may only be used to reduce the velocity search range, otherwise information about one or more reflectors might be lost, i.e. they might appear to be part of the same semblance region.

4.4 Adapting to the real world

In an ideal world, the Earth is flat, all of its subsurfaces are constant and parallel to the surface everywhere, all layers are isotropic and homogeneous and all the data recorded is noise-free. However, we can settle for a locally flat Earth with locally parallel subsurfaces which are piecewise continuous and, if the anisotropy or inhomogeneity in its layers are not too strong, we can estimate their parameters fairly well, but we are still left with the noise issue.

The types of noise that interfere with the picking method proposed in the previous section are the ones that affect the amplitude of the signal recorded in each trace. More specifically noise that affects regions the signal with low amplitudes. In noise-free data, a low amplitude peak is still a local extreme and regardless of the magnitude of the amplitudes of the peaks associated to other events. However, in reality, all measurements are disturbed by noise, coherent and incoherent. The issue of coherent noise will be addressed in the next section; here we are interested in background noise that might difficult the detection of events with small magnitude.

The term "incoherent noise" is used to refer to noise that appear to be random, like white or brown noise, or that are not coherent throughout a number of traces; a real event is expected to be recorded (not necessarily detectable) in all traces. Incoherent noise does not follow a trend or at least not a trend that resembles a real event trend and, if a real event is just above the incoherent noise level, it will be detected as a local extreme. There are many algorithms available capable of improving the signal to noise ratio by attenuating background noise that can be used before the velocity modeling takes place. For example, the use of bandpass filters can attenuate background noise by removing frequencies outside a given range, and the use of a Hilbert transform filter envelopes the sinusoidal peaks related to one event into a single peak. Any such filters must be applied by an expert since their incorrect use will result in data loss.

Given that we have a filtered gather with attenuated background noise, we can use the method described in the previous section to identify possible events. Now we discuss sources of coherent noise.

4.5 Multiples: the ghosts that did not dissipate

Multiples are unwanted reflections of the energy that remained after a primary reflection. After energy leaves a source, it is expected to travel until it reaches a boundary where it is reflected or refracted. Then the refracted energy keeps on travelling until it finds another boundary where it is reflected or refracted and so on until all the energy has dissipated. However, we forgot to account for everything that happens to the energy that was reflected. Ideally all the energy would be absorbed by the receiver or the interfaces and this energy would follow the same path it travelled downwards upwards but that is not the case. The same way energy can be reflected, refracted and absorbed when travelling downwards, the same can happen when it is travelling upwards.

Say we have one source and one receiver placed on a line and there is only one subsurface beneath and this subsurface is parallel to the surface that contains the line. When we fire the source, what we would like to see is the energy travelling downwards, maybe partially dissipating on its path, until it reaches the subsurface boundary where it is totally reflected upwards. The energy reflected travels back in the same manner and reaches the surface where it is recorded by the receiver and any remaining energy disappears. The problem is that this energy does not disappear, it can be absorbed but it can also be reflected by the surface following a similar path as the one it followed when it first left the source. In marine exploration, the "surface" that contains the sources and receivers is water and the coefficient of reflection from water to air is close to -1 meaning that most of the energy reaching the water-air boundary is reflected.

The number of boundaries in the problem is irrelevant, any wave that was reflected once may be partially reflected again from any boundary on its way upwards generating multiples on the seismic data. Multiples can be classified as long-path, short-path and peg-leg multiples. Long-path and short-path multiples behave the same way as a primary reflection while the wave in a peg-leg multiple might reverberate between interfaces before making its way back to the surface; peg-leg multiples are defined in various different ways in the literature, here we choose to define a peg-leg multiple as a multiple with an asymmetric path. Short-path multiples tend

to be reverberations from a shallow subsurface like the seabed and tend to arrive near to their primary reflections. In semblance maps, they tend to form a distinct pattern of events that seem to have similar velocities but placed at different times; normally this pattern can be separated from the real velocity trend since the RMS velocity trend would be nearly constant and the interval velocities derived would be almost the same as the RMS velocities. Unlike short-path multiples, long-path multiples do not appear close to their primary events but tend to appear as an event outside of the RMS velocity trend since it presents itself with the same RMS velocity as its primary but at a later time where events are expected to have much higher velocities. Finally, peg-leg multiples can be harder to detect in comparison to long-path and short-path multiples; in the case of short-path peg-leg multiples, there can be a noticeable interference with primary reflections. That is likely to happen when the period of the peg-leg is small; however, peg-leg multiples with longer paths or with high-period and peg-legs usually do not fit the RMS trend.

Many filters have been developed to remove multiples and they are usually applied to the data before the velocity analysis takes place (Verschuur, 2006). If that is not the case, expert knowledge of the interval and RMS velocity trends can also eliminate part of this source of noise.

4.6 Measuring Flatness

Multiples, diffractions, noise and other forms of coherent noise introduce ambiguity in the layer selection process and are usually detected as a possible real event. Even though part of these artefacts can be eliminated with data processing, there are still cases where we do not have enough information to eliminate or keep an event in the model even with expert advice. Since seismic data is subject to interpretation, the information given by experts depend on their judgement of what interfaces are important for the model. When analysing real data, it is unlikely that we would be able to resolve all layers in the subsurface because that implies a complete discretization of the layers which is rare; however, experts are interested in identifying layers that carry relevant information for their analysis and not every possible layer.

Depending on the expert's background, their interest might be in finding the most geologically coherent model according to their knowledge in geological eras or they might be interested in detecting the interfaces that better resolve the seismic image given coherence across gathers regardless of its geological interpretation. When no explicit prior information about the geology of the subsurface is available, or when other forms of ambiguity are present, we choose the model that better stacks the data.

If we are looking for the velocity trend that best stacks the data, we have to find a way to compare different models according to the amount of information provided by them. During semblance analysis, the expert usually has a semblance map created using a guide curve, e.g. the travel time approximation in Equation 3.19 and the CMP gather displayed side by side as in Figure 4.11. The expert then picks the trend that best removes the moveout effect of the events in the gather that is geologically accurate according to their prior knowledge. If we transform this gather using the actual values used to create this example, we obtain the results in Figure 4.12; the events in the transformed gather on the right appear to be flat and that indicates that the data was transformed with a suitable velocity trend. If one or more values were incorrect, one or more events would appear curved instead of linear. Say the velocity of the second layer were wrong. If it were lower than its true value then the migrated event would appear to be curved upwards and, if higher, downwards as illustrated in Figures 4.13 and 4.14; when the velocity is higher than its actual value than the event is said to be overcorrected and undercorrected when lower.

In this case, it is clear that undercorrected and the overcorrected models are not good since the velocity picks for the second layer are clearly not optimal and we can easily "move" the pick to the red zone. However, in real cases, the layers are usually not as well defined or clearly isolated as this one and choosing an appropriate number of layers or a reasonable set of layers to migrate the data is a complex task. For this reason, we need a measure that can compare the level of "flatness" in the migrated gathers and a simple way to do so is by using correlations. If the migrated gather were perfectly flat, all traces would be proportional to each other

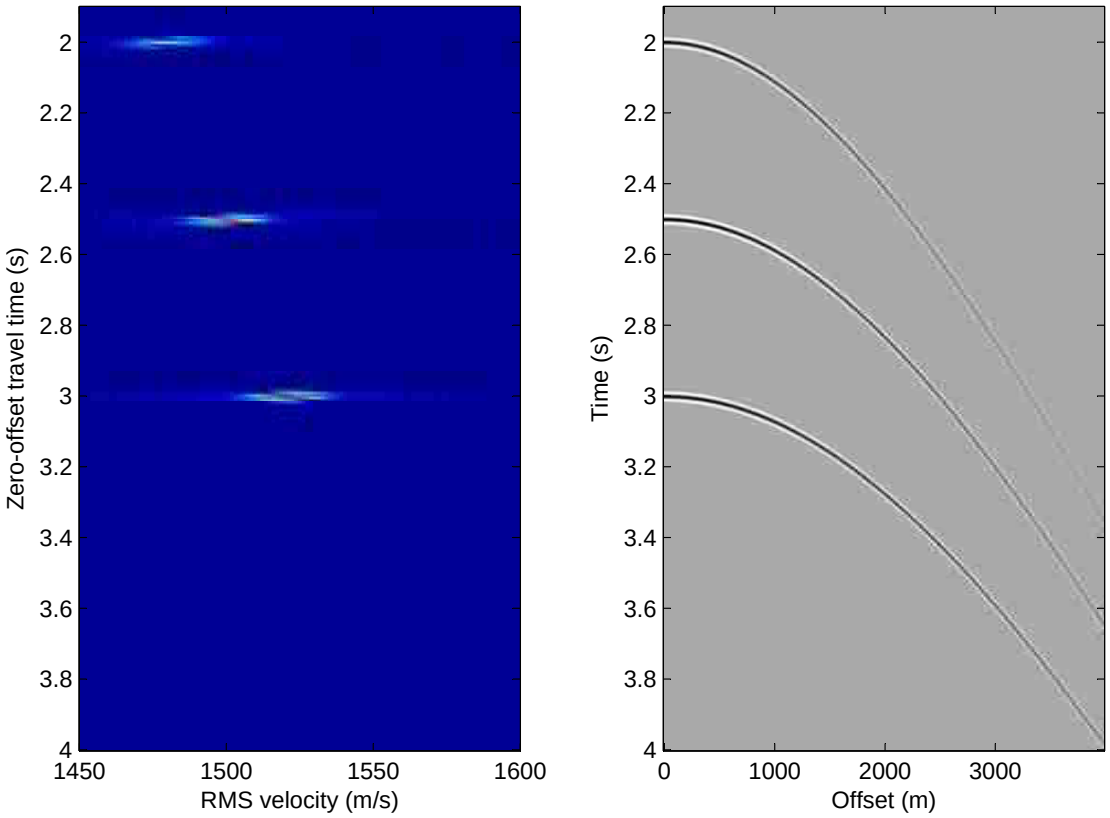

Figure 4.11: A weighted semblance map generated using the travel-time equation in 3.19 on the left and its corresponding CMP gather on the right.

in amplitude at least locally, i.e. the trace-by-trace correlation would be maximal if the amplitudes in one trace were proportional to the ones in the other traces. Given a gather formed by m traces with m_s samples each, if we calculate the sample correlation coefficient of the amplitudes for every pair of traces, we would obtain a $m \times m$ matrix of correlation coefficients. If all traces were perfectly correlated then all the elements in this matrix would be one, otherwise these values would be anywhere in the interval $[-1, 1]$. If we compute such matrices for the original CMP gather in Figure 4.11 and the migrated CMP gather in Figure 4.12, we obtain the matrices represented as a colour map in Figure 4.15. After visual inspection, it is clear that the traces of the transformed gather are better correlated and therefore "flatter" than the ones in the original gather; however, it would be more efficient to calculate a measure that represents this level of flatness.

For this reason, we propose an index of flatness (IF) which we define as the

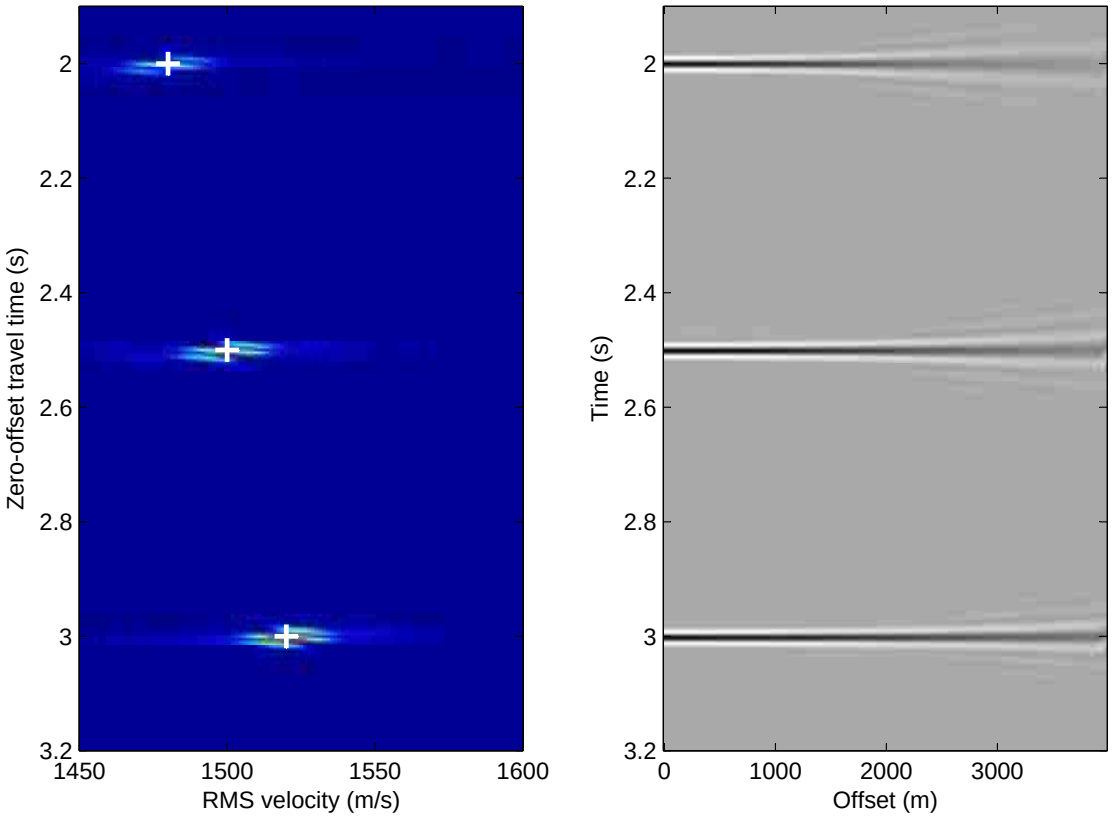

Figure 4.12: Left: A weighted semblance map generated using the travel-time Equation in 3.19 with the actual values used to create the original gather marked in white. Right: migrated CMP gather on the right using this same equation and the actual values used to generate the original CMP gather.

mean of the entries of the upper triangular submatrix excluding the diagonal of this correlation matrix. Say we have a gather α with m traces then the correlation matrix $CM^{(\alpha)}$ is $m \times m$ and IF_α is given by

$$IF_\alpha = \frac{2}{m(m-1)} \sum_{i=1}^{m} \sum_{j=i+1}^{m} CM_{ij}^{(\alpha)} \tag{4.6}$$

and $IF_\alpha \in [-1, 1]$; so the closer IF_α is to 1, the flatter is the transformed gather.

In the example in Figure 4.11, the IF coefficient for the original gather is $IF_{orig} = -0.0074$ and for the transformed gather using the actual values we have $IF_{trans} = 0.8047$. If we were to calculate this matrix for the under and overcorrected gathers in Figures 4.14 and 4.13, their IF coefficients would be $IF_{under} = 0.5457$ and $IF_{over} = 0.5521$, respectively. As we would expect, the gather transformed with the actual

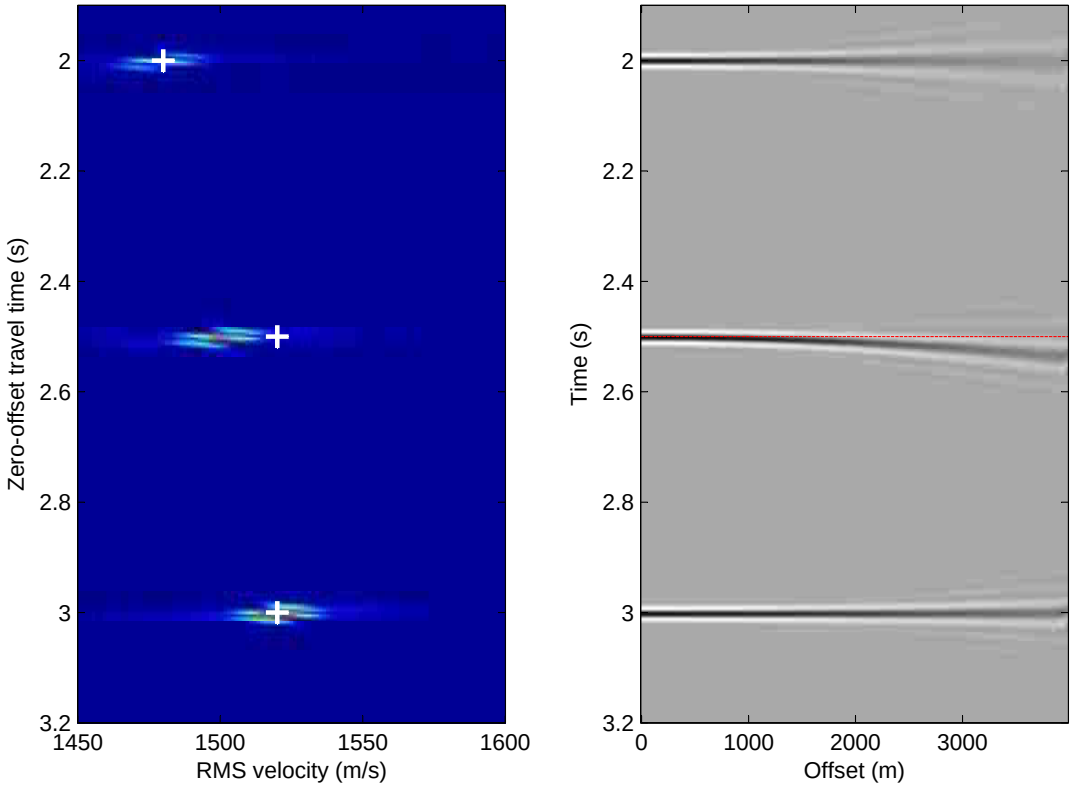

Figure 4.13: Weighted semblance map from Figure 4.11 and gather migrated with the true values for layers 1 and 3 and a RMS velocity for the second layer which is lower than its actual value. The red lines indicated where the curve would be expected to be centred if the velocities were correct.

values has a higher coefficient than the unmigrated and the under and overcorrected gathers. We can also use this measure to verify if the introduction or removal of a layer in a model improves it. Return to the example in Figure 4.11 and suppose that we had selected all three layers independently but we were not able to verify if all three are relevant for the model so we calculate the IF for the transformed gather for each possible layer combination. The results are displayed on table 4.6 and the three-layered model is in fact the flattest one.

4.7 Seismic Event Tracking Algorithm

Here we can finally summarize the previous Sections in one tool, the seismic event tracking algorithm (SET algorithm). A similar algorithm was introduced in Ca-

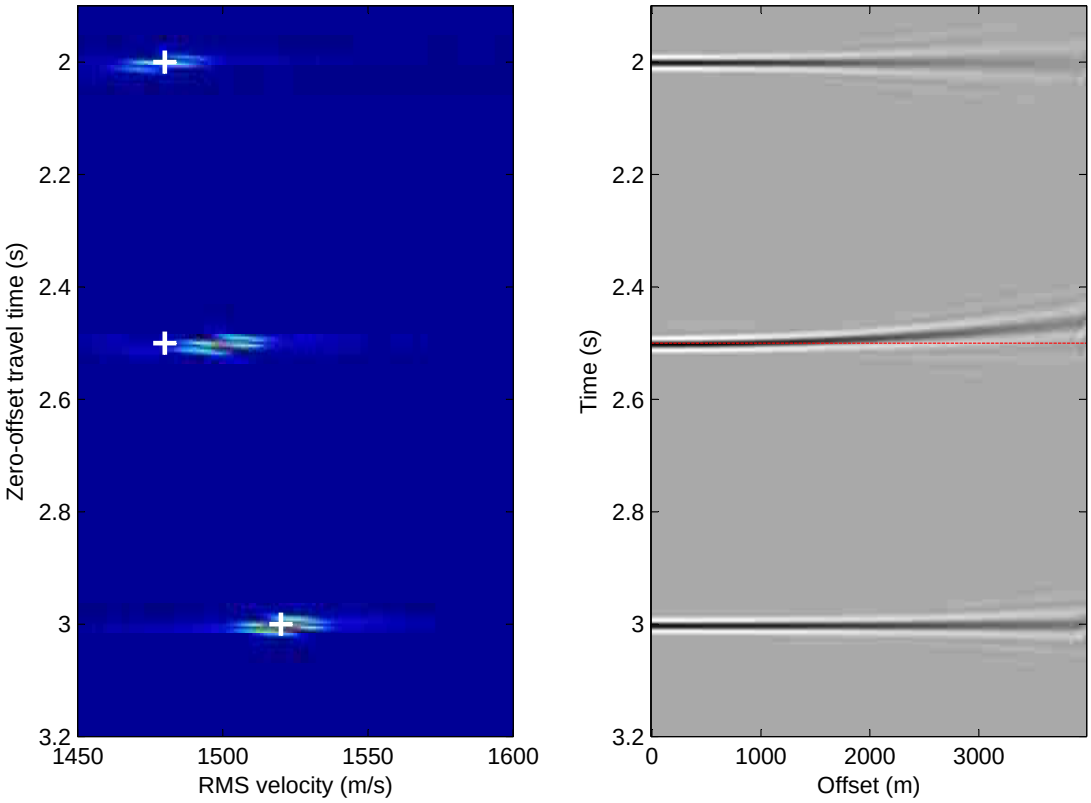

Figure 4.14: Weighted semblance map from Figure 4.14 and gather migrated with the true values for layers 1 and 3 and a RMS velocity for the second layer which is higher than its actual value. The red lines indicated where the curve would be expected to be centred if the velocities were correct.

iado et al. (2011) as a method to select the data points to be used during velocity modeling; the steps of the picking algorithm used here are described below:

1. Given the m_s-by-m matrix of recorded amplitudes A with m traces each with m_s samples, pick all m_p extremes in the first trace $A_{.1}$. In noisy gathers, m_p might be too large since every disturbance in the signal would produce one or more extremes. To avoid this problem, we can set a minimum peak height, minimum peak distance and threshold to reduce the number of noisy picks. For example, the minimum peak height could be set as the receiver's factory standard recording error of the mean of the absolute height of the last 5% points in a trace if the expert believes that no relevant events were in that area.

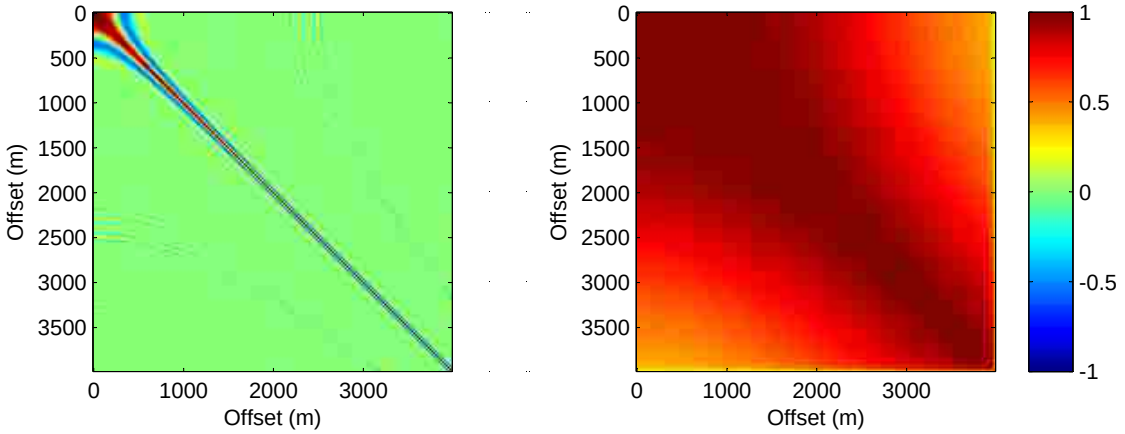

Figure 4.15: Correlation matrices for the CMP gather in Figure 4.11 (left) and the transformed CMP gather in Figure 4.12 (right). The entries of these matrices are the correlation coefficients calculated for each pair of traces in these gathers and the flatter the gather appears to be, the closer to 1 these coefficients are.

2. We need to loop through all m traces matching the m_p extremes in the first traces to extremes in the remaining $m-1$ traces. Start the counter $i_p = 1, \ldots, m_p$ and the counter $j = 1, \ldots, m$.

3. Given the recorded time at i_p in trace j, assume that the zero-offset travel time can be approximated using the recorded time i_p in trace 1 and use the search window in 4.5 to restrict the search area in trace $j+1$. Inside this search window in trace $j+1$, pick the sample with the minimum absolute distance to the sample picked in trace j in the amplitude domain. We are trying to create the matrix $T^{(o)}$ but we cannot refer to it as such until we assure that all the layer candidates follow the approximately hyperbolic trend.

4. After step 3, there might be too many layers picked and some of them might not follow the hyperbolic pattern that we would expect from a seismic event. The quickest way to exclude spurious events is by fitting a hyperbolic least-squares model to each layer candidate using Equation 3.19. We can eliminate any candidates with a low R-squared since that indicates a probable misfit.

5. Fit the 1D S-BRAINS model to each of these candidates with vague priors

Layer(s)	IF
1	0.1955
2	0.3605
3	0.4195
(1,2)	0.6945
(1,3)	0.7822
(2,3)	0.6865
(1,2,3)	0.8047

Table 4.1: Calculated index of flatness (IF) for transformed gathers using all possible combinations of the three layers in the synthetic model.

as if each candidate were part of a single-gather model. Identify the RMS velocity trend produced by the posterior mean estimates in each candidate. If prior information about the trend is given by the expert, use it to flag outliers, e.g. the trend is expected to increase smoothly with time then fit a spline to the trend and produce credibility intervals to identify outliers. This will be explained further when analysing synthetic and real datasets.

6. Use the index of flatness (IF) to identify the best possible combinations of candidates and compare models.

After eliminating all unlikely candidates, we have that $T^{(r)}$ is the matrix of recorded times for a given set of candidates and further analysis can be carried using the BRAINS models in Chapter 3. Next we introduce a set of algorithms used to approximate the posteriors in BRAINS.

4.8 Markov Chain Monte Carlo Algorithms

Monte Carlo methods is the name given for a class of algorithms based on stochatic sampling; Enrico Fermi was one of the first scientists to use MC methods in the 1930s, but without computers his version was limited. This class of methods were formally introduced by Ulam in 1946 when he attempted to calculate the probability

of winning a game of solitaire; the idea was popularized by von Neumann who applied it to his experiments in thermonuclear and fission problems. The first Markov Chain Monte Carlo method was published in 1949 by Metropolis (Metropolis and Ulam, 1949) who combined Monte Carlo methods and random walks (i.e. Markov chains) by including an acceptance rate to each new sample. Hastings (1970) generalized the Metropolis algorithm focusing on the dimensionality of the problem and the properties of the chains used creating the Metropolis-Hastings algorithm. MCMC methods became popular among statisticians in the late 1980s (Besag, 1989; Geman and Geman, 1984) and the early 1990s (Smith and Roberts, 1993; Gelfand and Dey, 1994) when computers were considerably faster and widely available. At the same time, important advances in the theory of MCMC algorithms were made in the works of Tierney (1994) and Roberts and Rosenthal (2007) where the assumptions necessary to analyse these chains were formalized. These were followed by works that discussed rates of convergence (Mengersen and Tweedie, 1996), acceptance rates (Roberts et al., 1997) and number of iterations (Roberts and Rosenthal, 1999). Two notable methods in MCMC theory are introduced by Green (1995) where he defines the reversible-jump Markov Chain Monte Carlo method which allows sampling in spaces of varying dimensions, and Skilling (2006) who introduced nested sampling.

The Metropolis-Hastings algorithm is used to generate sequences of samples of a multivariate probability distribution with unknown normalizing constant, say $\pi(\mathbf{x})$, in order to approximate such distribution or evaluate an integral. The algorithm is started by generating a candidate, $\mathbf{x_0}$ in the target space and starting the counter at $t = 1$. For a given t, a new sample $\mathbf{x}'$ is drawn from a proposal distribution $p\left(\mathbf{x}'|\mathbf{x_{(t-1)}}\right)$. The new sample is accepted with probability

$$\alpha = \min\left(1, \frac{\pi\left(\mathbf{x}'\right) p\left(\mathbf{x_{(t-1)}}|\mathbf{x}'\right)}{\pi\left(\mathbf{x_{(t-1)}}\right) p\left(\mathbf{x}'|\mathbf{x_{(t-1)}}\right)}\right) \tag{4.7}$$

and $\mathbf{x_{(t)}} = \mathbf{x}'$. Otherwise, reject the proposal and $\mathbf{x_{(t)}} = \mathbf{x_{(t-1)}}$.

As $t \to \infty$, $p(\mathbf{x}) \to \pi(\mathbf{x})$.

When the proposal distribution is symmetric, i.e., $p\left(\mathbf{x}'|\mathbf{x_{(t-1)}}\right) = p\left(\mathbf{x_{(t-1)}}|\mathbf{x}'\right)$, the algorithm is simplified to the Metropolis case and it is a good option when no information about π is available. Another commonly used proposal distribution is

the independent proposal where $p\left(\mathbf{x}'|\mathbf{x}_{(\mathbf{t-1})}\right) = p\left(\mathbf{x}'\right)$ which works adequately for multimodal π given that p is close to π since it assumes that the current sample does not depend on the previous one. The use of local proposals which are highly dependent on the previous state should be discouraged unless one is certain that π is unimodal.

For high-dimensional $\mathbf{x}$ (say d dimensional), we might be unable to derive a proposal density for π so, instead of updating the whole vector $\mathbf{x}$ in one step, we can update each of its variables x_i, $i = 1, \ldots, d$ with a proposal density p_i. At a given time t, simulate x_i' from p_i so that $x' = (x_1, \ldots, x_i', \ldots, x_d)$ and $x_{(t-1)} = (x_1, \ldots, x_d)$. Compute

$$\alpha = \min\left(1, \frac{\pi\left(\mathbf{x}'\right) p\left(\mathbf{x}', x_i\right)}{\pi\left(\mathbf{x_{(t-1)}}\right) p\left(\mathbf{x_{(t-1)}}, x_i'\right)}\right), \tag{4.8}$$

make $\mathbf{x_t} = \mathbf{x}'$ with probability α and $\mathbf{x_t} = \mathbf{x_{(t-1)}}$ with probability $1-\alpha$. This process is called the variable-at-a-time Metropolis-Hastings update and tends to cover more of the space in the same number of runs as the classic Metropolis-Hastings algorithm.

Here we choose our variant of the Metropolis-Hastings according to the dimensionality of the problem and level of prior information. The original Metropolis algorithm is used in small problems involving only RMS velocities and zero-offset travel times and the variable-at-a-time update is used with more accurate models where we need to estimate depths and interval velocities.

As an example, we take a random sample of size 20 from the sine function in the interval $[0, 10]$ and add gaussian noise with standard deviation 0.05 as illustrated in Figure 4.16. We then assume that we have no information about the underlying function and let these points be a sample of a Gaussian process with a constant mean and covariance function as in 3.58. We assume that a priori the mean parameter and the covariance function parameters are all normally distributed with zero mean and standard deviation 5; these distributions are arbitrary and reflect lack of knowledge of the underlying function. We draw 10^4 samples using the M-H algorithm and discard the first 10^3 samples in a process called burn-in. In principle, the length of the chain and the length of burn-in are arbitrary; the burn-in phase is meant to be

as long as it is necessary to make the chain "forget" about its starting point and the actual chain should run for as long as it takes for it to "converge". In the proposal of the M-H algorithm, there is no underlying requirement for a burn-in phase and there is no evidence that this is actually necessary; in theory, if the chain ran for long enough, the impact of the burn-in samples should be negligible. However, the ideal length for such a chain is infinite and it is unlikely that we would be able to simulate it for that long so we settle for the next best thing, we run the chain until it seems to have converged.

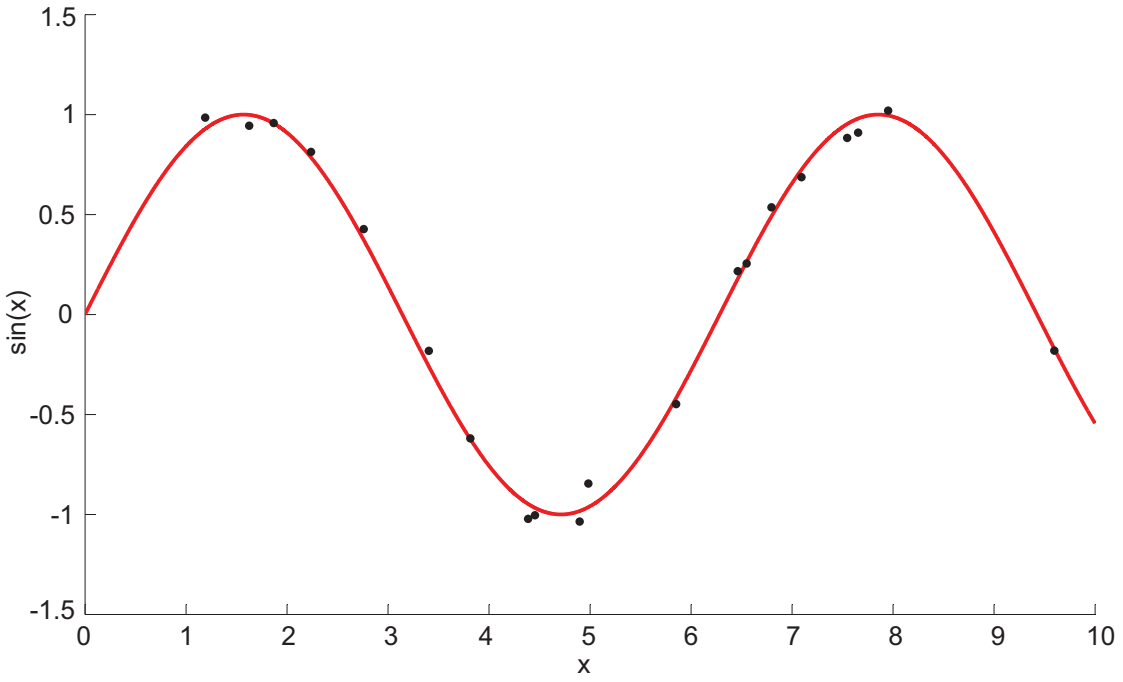

Figure 4.16: Sine function in red and random sample of size 20 with gaussian noise ($s = 0.05$) in black.

If we draw the 10^4 samples and throw away the first 10^3, we obtain the samples illustrated by the gray lines in the top plot of Figure 4.17. The mean curve is represented by the black line, the 2.5% and the 97.5% quantile curves are plotted in blue and the real function is the dashed red line. If we choose to keep the burn-in samples, we obtain the results in the bottom plot in the same Figure. Despite having a plot that looks messier when all samples are displayed, the posterior means and quantiles are virtually the same indicating that the impact of the burn-in samples were irrelevant. That will not always be the case since this behaviour depends on the

chain's rate of convergence which is directly linked to the shape of the distribution we are sampling from.

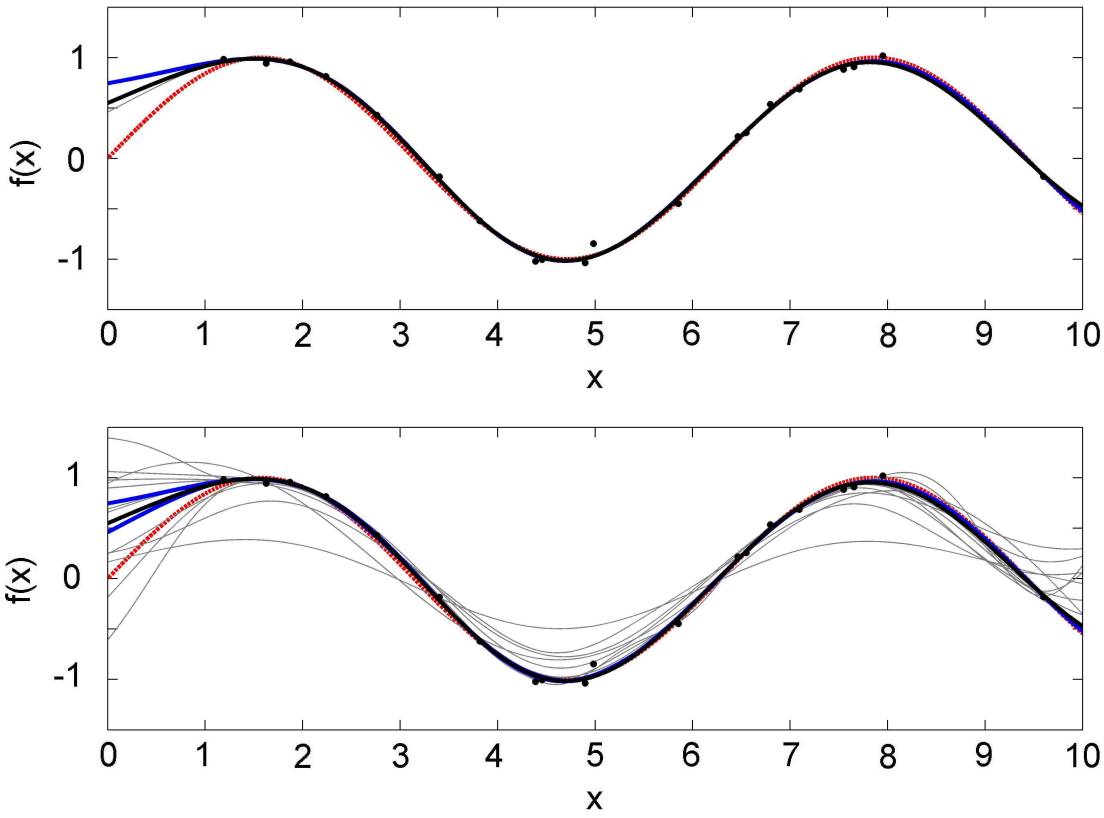

Figure 4.17: Curves sampled obtained using the M-H algorithm including (top) and excluding (bottom) burn-in samples. The gray lines are the M-H samples, the black line is the posterior mean curve, the blue lines are the posterior quantiles (2.5% and 97.5%) and the red dashed line is the real underlying curve.

From these results, it is clear that the mean of the chain "converged" to a curve close to the real one; to assess convergence, we calculate the slope of the running means and the running quantiles for each parameter and assume that the chain converged if the slope is close to zero for all sequences. This method is not fail proof; the chain could have got stuck in that specific region of the search space or our choice of priors could have been too narrow. For this specific example, that is not the case but we need to incorporate automatic or semi-automatic convergence diagnostics in our methods later on. In the next Chapters, we analyse single-gathered and multi-gathered synthetic examples and a real example.

Chapter 5

Stacking BRAINS in Single-Gathered Synthetic Examples

In this Chapter, we analyse a number of synthetic datasets using the models proposed in Chapter 3 and the algorithms described in Chapter 4. The main purpose of this work is to obtain a suitable velocity model for a given dataset with point and interval estimates for each physical parameter and, in order to achieve that, we need to verify the precision and limitations of the proposed method by analysing synthetic data. Since processing speed and automation are two important factors, we aim to optimize the workflow to minimize the need of expert interaction. First we describe the method used to create these synthetic examples.

5.1 Seismic Forward Modeling

Finite difference, finite element and spectral analysis are the most used techniques to simulate seismic reflection data by numerically solving the wave equation. The finite element method approximates solutions of partial differential equations (PDE) by first approximating the domain of the PDE with an union of geometrical objects to create a mesh, then approximating the solution of the PDE using a piecewise linear function on the nodes of the mesh and finally interpolating the nodal solutions on

the faces of the mesh (Zienkiewicz and Taylor, 2000). The finite difference method also starts with a mesh or grid of the domain of the PDEs to be solved and then the derivatives in the PDEs are replaced with finite differences resulting in a system of linear or non-linear equations to be solved (Mitchell and Griffiths, 1980). The spectral methods are fairly similar to the finite element method with the difference that the spectral methods approximate the solution of the PDEs as a linear combination of continuous functions, e.g. Chebyschev polynomials and Fourier series (Gottlieb and Orszag, 1993).

The phase-screen method is a spectral method used to model the transmission of seismic waves in heterogeneous media (Stoffa et al., 1995; Wu, 1994); it is vastly used to model reflection seismic data and, while it is computationally efficient and simulate diffractions, it is restricted to small angles and does not simulate multiples. This method built on diffraction theory and it consists of approximating the differential equations on screens which are orthogonal to the direction of propagation using fast Fourier transforms (Brigham, 2002). The method used to create our synthetic gathers is an extension of the phase-screen method for isotropic and anisotropic media which simulates nonnormal raypaths; first the domain is scanned to assess the complexity of each screen, then the propagating wavefields are modeled as a function of the horizontal wavenumber and each subset is transformed separately (White and Hobbs, 2007).

5.2 Single-Layered Isotropic Synthetic Examples

We start with two single-layered isotropic zero-offset synthetics; model (1.a) (SM-1a) has one boundary at $1480m$ with interval velocity of $1480m/s$ and model (1.b) (SM-1b) has one boundary at $3000m$ with interval velocity of $3000m/s$; their CMP gathers are displayed in Figure 5.1. Given a CMP gather, an expert would try to pick RMS velocity and zero-offset travel time pairs with the aid of the CMP gather and a semblance map (see Section 4.1) with the NMO moveout approximation in Equation 3.19 or a Taylor expansion of this same equation. Here we use Equation 3.19 and select each set of samples used to calculate the semblance using a linear

extrapolation, we also calculate the mean amplitudes in each sequence and create a weighted semblance map using both maps. Figures 5.2 and 5.3 show the maps for SM-1a and SM-1b respectively. In all maps the horizontal grid spacing is $0.1m/s$ and the vertical grid spacing is $0.002s$.

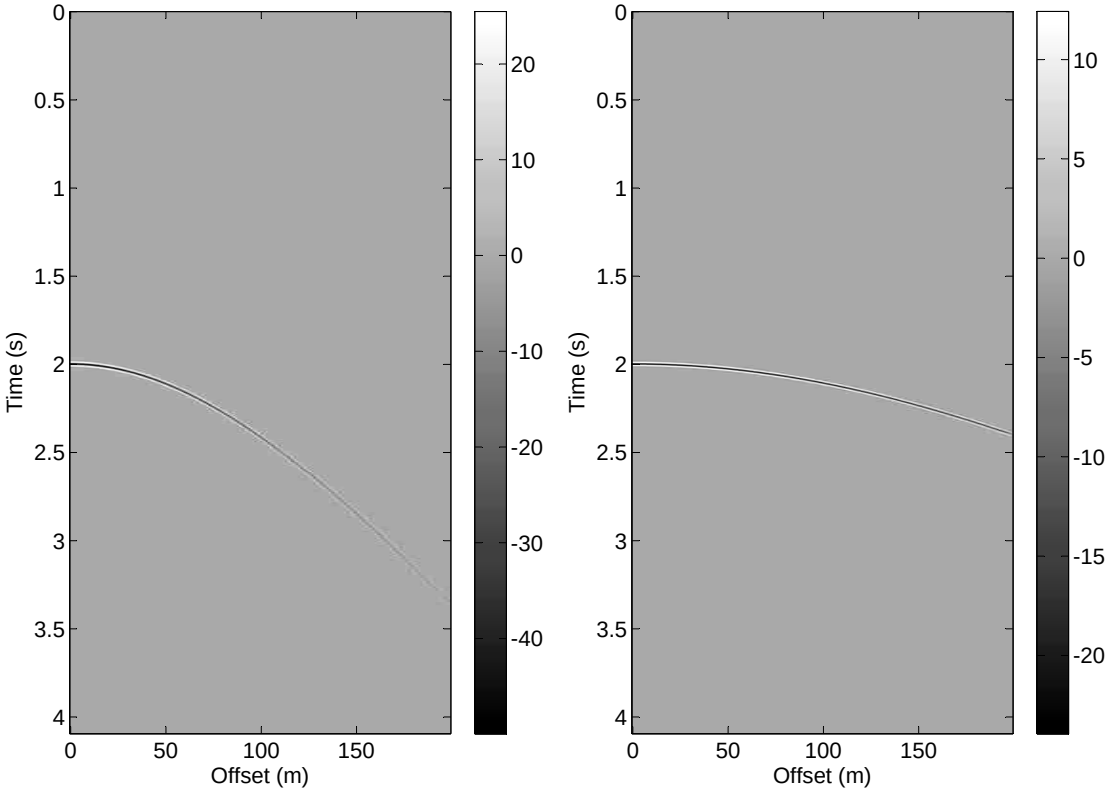

Figure 5.1: On the left, CMP gather for the synthetic model SM-1a and, on the right, CMP gather for the synthetic model SM-1b.

Note that in both cases the values used to create these synthetics correspond to points of high semblance and high absolute mean as expected. However, many other velocity and zero-offset travel time pairs in the neighbourhood of the actual values produce a similar result in both domains making it harder to select just one appropriate point without optimization tools or prior information. The point of maximum semblance in SM-1a correspond to a RMS velocity of $1485.6m/s$ and zero-offset travel time of $1.986s$ while in SM-1b, the maximum semblance is at the point $(3051.3m/s, 1.952s)$. Neither correspond to the values used to create the synthetics and the main reasons for that are high levels of coherent noise since the semblance only measures coherence and not magnitude, and the fact that the Equation 3.19

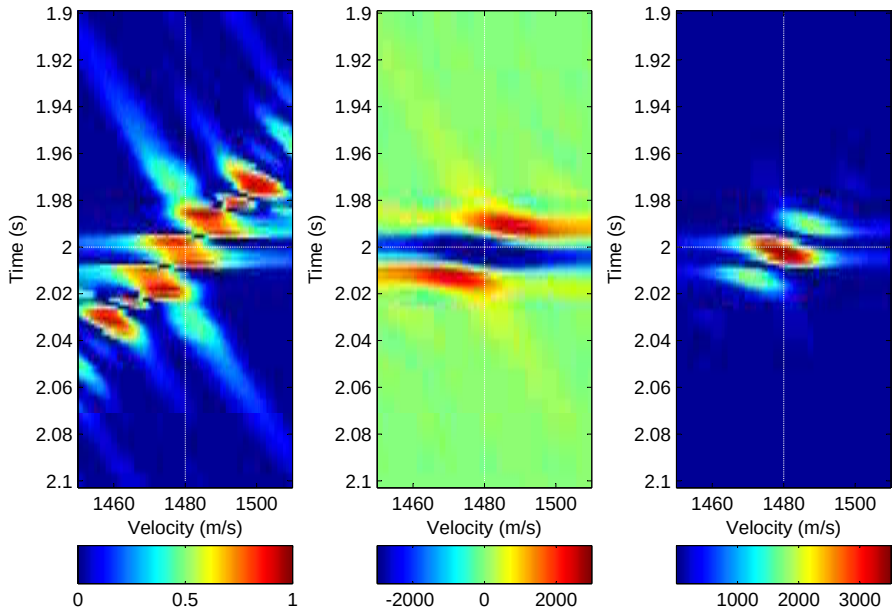

Figure 5.2: On the left, raw semblance map generated for SM-1a. In the centre, map of the mean amplitudes of the sequences used to generate the semblance map on the left. On the right, semblance plot weighted using the map in the centre.

used to create the maps might not be a good fit.

If we look at the maps of mean amplitudes, the point of extreme mean amplitude in SM-1a corresponds to a RMS velocity of $1479.3m/s$ and zero-offset travel time $2.002s$ and in SM-1b, the extreme corresponds to a RMS velocity of $2998.7m/s$ and zero-offset travel time $2.002s$. These results are closer to the actual values than the values obtained by picking the points of maximum semblance. Since we have noise-free zero-phase wavelet synthetics, the samples of highest absolute amplitude in the signal should correspond to the events produced by the reflector and that is why these results are closer to the actual values. Finally, if we combine semblance and mean amplitude maps to create the weighted semblance maps, we obtain the same results as the ones given by the mean amplitude map but now we also have the information that the chosen points were high semblance points.

The remaining error in these picks could be attributed to a modelling error de-

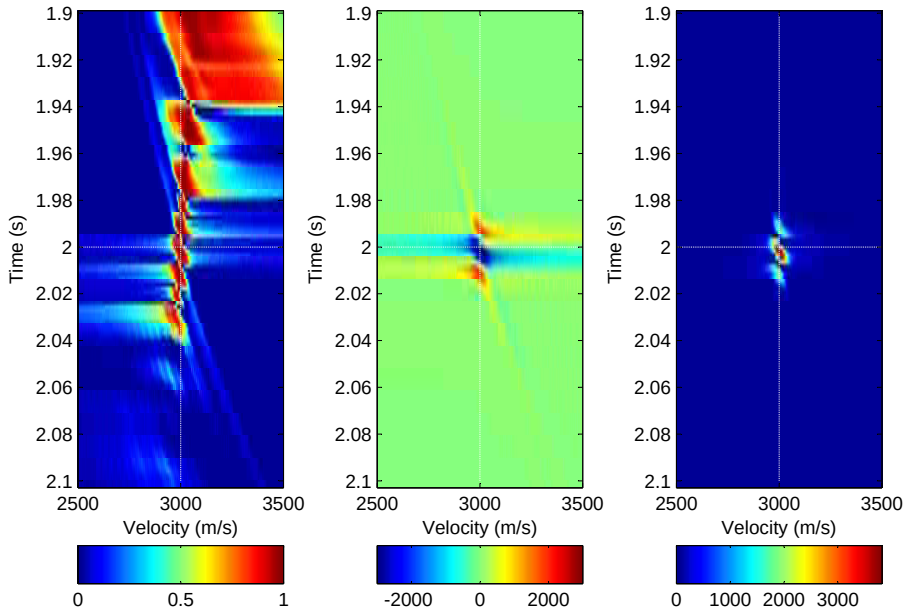

Figure 5.3: On the left, raw semblance map generated for SM-1b. In the centre, map of the mean amplitudes of the sequences used to generate the semblance map on the left. On the right, semblance plot weighted using the map in the centre.

rived from the use of Equation 3.19 or the discretization of the signal given that there is a $0.002s$ difference between the actual zero-offset travel time and the picked ones which is the same as the data sampling period. In both cases, the semblance analysis alone produced unreliable results while the weighted method, which is similar to the one used by experts, returned results close to the real values. In addition, experts use the actual CMP gather to help them choose the best pick and after selecting zero-offset travel time and velocity pairs, the experts apply a transformation based on equation 3.19 to the original gather and check how "flat" the events appear after the move-out correction has been applied. In Figure 5.4, we have the gather used in SM-1a on the top left and its transformation using equation 3.19 based on the true zero-offset travel time and velocity values on the top right. The bottom left map shows the correlation matrix for the original gather $CM^{(1a_o)}$ and the bottom right, the correlation matrix for the transformed gather $CM^{(1a_t)}$(see Section 4.6).

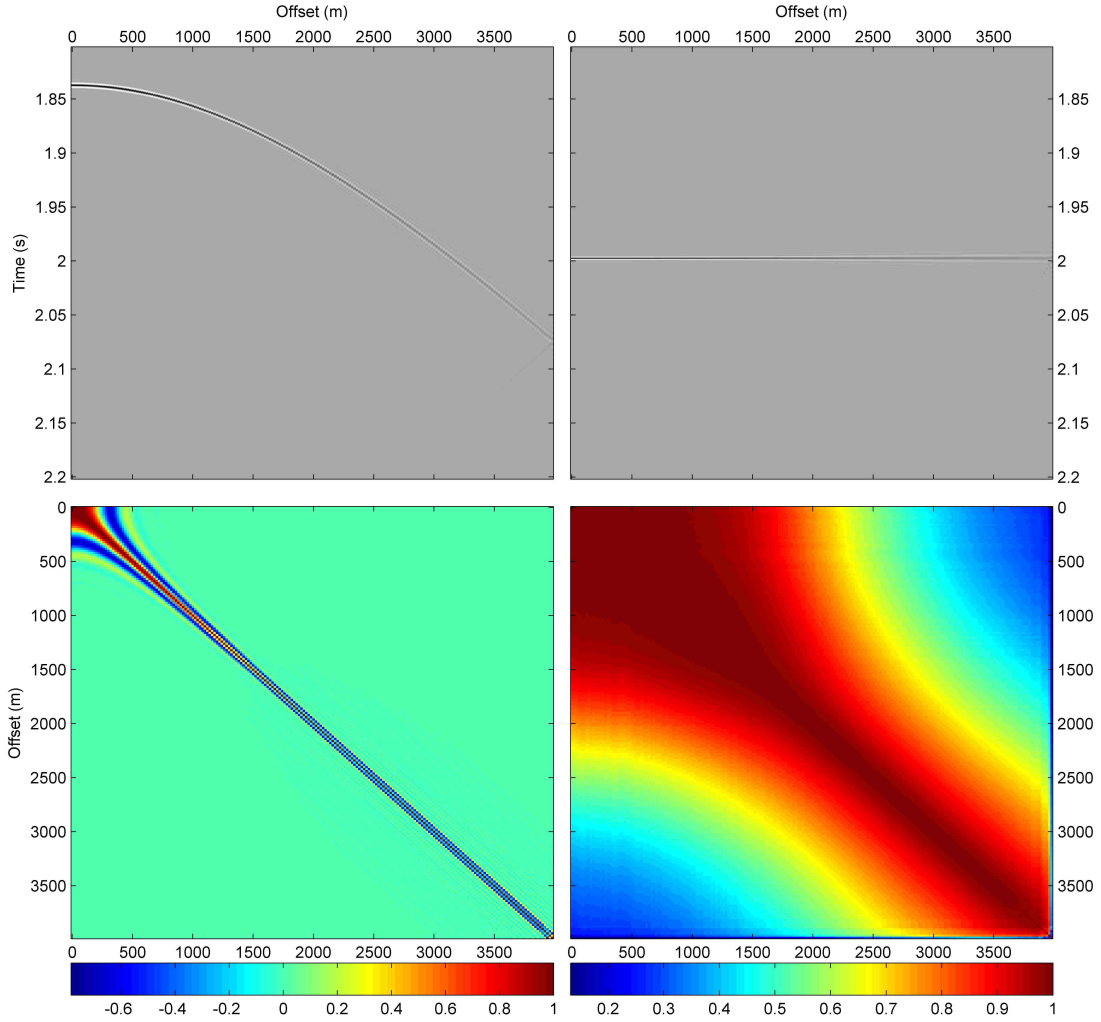

Figure 5.4: On the top left, we have the CMP gather for SM-1a and, on the bottom left, the correlation matrix associated to it. On the top right, we have the same CMP gather transformed using equation 3.19 based on the true zero-offset travel time and velocity values for this synthetic and, on the bottom right, the correlation matrix corresponding to it.

In order to compare these two correlation matrices, we calculate the indexes of flatness for both and obtain $IF_{(1a_o)} = -0.0029$ and $IF_{(1a_t)} = 0.7262$ for the original gather in SM-1a and the transformed gather with the actual values, respectively. For SM-1b, the index of flatness for the original gather is $IF_{(1b_o)} = 0.0015$ and for the migrated gather with the actual values is $IF_{(1b_t)} = 0.9758$. If we repeat the same transformation with the best semblance pick using the semblance map weighted

with the absolute means, we get $IF_{(1a_w)} = 0.7040$ for SM-1a and $IF_{(1b_w)} = 0.97372$ for SM-1b. If we use an optimization algorithm like the quasi-Newton method to maximize the index of flatness, we will find that the true values are in fact the maxima in both cases. See Deuflhard (2004) for more information on Quasi-Newton methods.

Even though optimizing the IF gives us point estimates that equal the actual values, this process does not return any information about the error in those estimates. If we apply the SET algorithm to SM-1a and fit the 1D S-BRAINS from Subsection 3.3.1 with an uniform hyperprior for the zero-offset prior mean $\mu_{t_{0_1}}$ in the interval $(1.5, 2.5)$ and uniform hyperprior for the RMS velocity prior mean μ_{v_i} in $(1450, 1550)$, and uniform hyperpriors for the prior standard deviations $\sigma_{t_{0_1}} \in (0, 0.5)$ and $\sigma_{v_1} \in (0, 50)$; these values are chosen arbitrarily to reflect vague prior information about this model. We also choose an uniform hyperprior for $\Sigma^{(t)}$ in the interval $(0, .5)$ and approximate the posterior distribution using the Metropolis-Hastings algorithm.

We use an independent M-H algorithm to draw 10^5 samples and discard the first 10^4, estimating the posterior mean using the remaining samples, we get $2.0001s$ and $1480.0m/s$. If use these same samples to estimate the 95% posterior credibility intervals, we obtain $(1.996s, 2.0052s)$ for the zero-offset travel time and $(1472.6m/s, 1489.2m/s)$ for the RMS velocity; we repeated the simulation with a random-walk M-H algorithm and obtained similar results. In all cases the acceptance rates were reasonable at around 35%, the chains showed good mixing and the running means and quantiles were stable.

For SM-1b, we used the same hyperpriors for the zero-offset parameters and used uniform hyperpriors for μ_{v_1} in $(2500, 3500)$ and σ_{v_1} in $(0, 100)$. Using the independent M-H to draw 10^5 samples and discarding 10^4 samples, the estimated posterior means are $2.0000s$ and $3001.4m/s$ for the zero-offset travel time and RMS velocity respectively and the 95% posterior credibility intervals are $(1.998s, 2.002s)$ for the zero-offset travel time and $(2984.2m/s, 3020.5m/s)$ for the RMS velocity.

In both examples, the posterior credibility intervals are fairly wide for a noise-free dataset analysis and varying our prior choices does not produce intervals much

narrower than these; while these results could reflect the uncertainty in the model and the data, this is more likely to be due to the fact that the model used is too rigid and does not account for the variability within traces properly. In order to verify this statement, we use the 1D SS-BRAINS defined in Section 3.3.4 in both datasets. In both cases, we use the even-numbered samples to estimate the covariance function parameters in 3.58 and the odd-numbered samples to estimate the RMS-velocity and the zero-offset travel time. Since we are not directly interested in the estimates for the covariance function parameters, we use a gradient search to find a local maximum in the likelihood and retrieve starting points for these parameters in an area of high probability. Afterwards, we use these estimates in the 1D SS-BRAINS model to estimate the RMS velocities and zero-offset travel times. The gradient search also returns optimal values for the zero-offset travel time and the RMS velocity and we use these values as the means of our priors. Another result obtained from a gradient search is the hessian matrix which approximates the inverse of the covariance matrix. We use the submatrix of the inverse hessian that corresponds to the covariances of the zero-offset travel time and the RMS velocity as an approximate for $\Sigma_{(t_0,v_{rms})_1}$.

In Figure 5.5, we have a plot of a section of the gather in SM-1a with the points selected by the picking algorithm in blue and the 95% posterior credibility interval for the predicted travel times created using the steps described above. The posterior mean is at the actual values used to create the synthetic, $1480m/s$ and $2s$ with 3 decimal places of precision; the joint posterior distribution of the RMS velocity and the zero-offset travel time and the 95% credibility contour is illustrated in Figure 5.6. In comparison with the previous model applied, the 95% posterior credibility intervals are $(1.996s, 2.004s)$ and $(1474.2m/s, 1483.6m/s)$; the zero-offset travel time interval does not show any improvement but the RMS velocity interval is now smaller. While these intervals still seem to be wide, they were produced without any expert prior information so they reflect the information in the data given our modelling assumptions.

If we use the 1D SS-BRAINS method in SM-1b, we get the joint posterior distribution illustrated in Figure 5.7; the 95% credibility contour is marked in black and the posterior means by the white dashed lines. The posterior means are the same

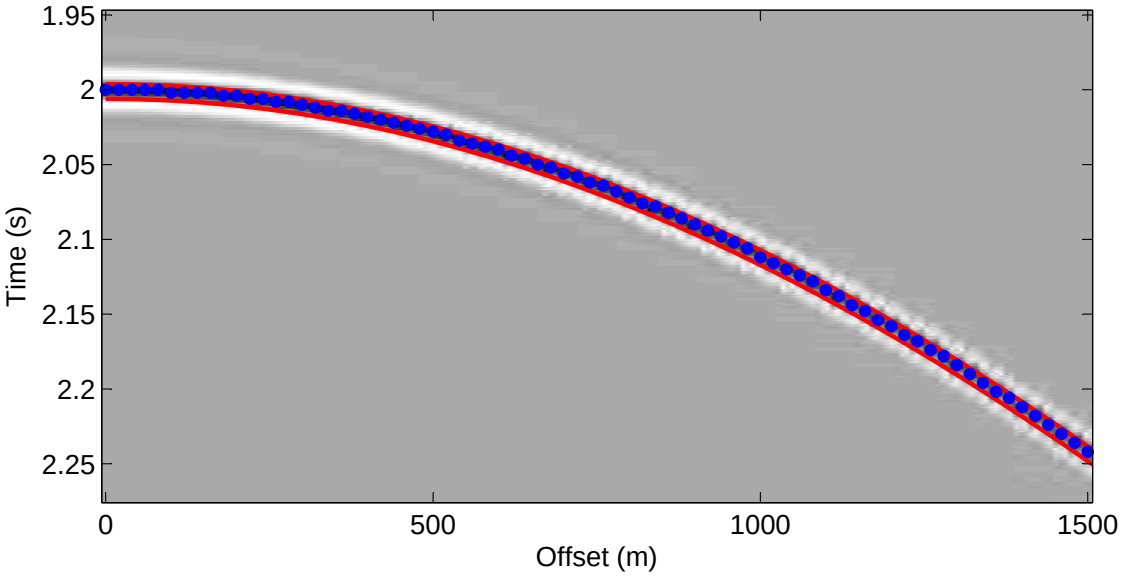

Figure 5.5: Section of the CMP gather in SM-1a with points selected by the picking algorithm in blue and the 95% credibility interval in red.

as the actual values and, unlike SM-1a, the 95% credibility intervals are extremely narrow reflecting the steepness of the marginal likelihood which suggests that the hyperbolic approximation is a better fit for deeper reflectors.

Here we can conclude that, despite being crude, the semblance analysis method picked by an expert works fairly well in a noise-free environment but relies on grid size and does not give any information about the uncertainty of the selected points or the chosen curve fit. However, when applying the Bayesian method combined with the picking algorithm, we have an accurate result with uncertainties. In the next section, we analyse a multi-layered synthetic model.

5.3 Multi-Layered Isotropic Synthetic Example

In this section we work with a multi-layered isotropic synthetic example. Model (2.a) (SM-2a) is formed by three layers and its CMP gather and weighted semblance map are displayed in Figure 5.8. In table 5.1, we have the true zero-offset travel times and RMS velocities used to create this model.

If we were to model just RMS velocities and zero-offset travel times and our aim

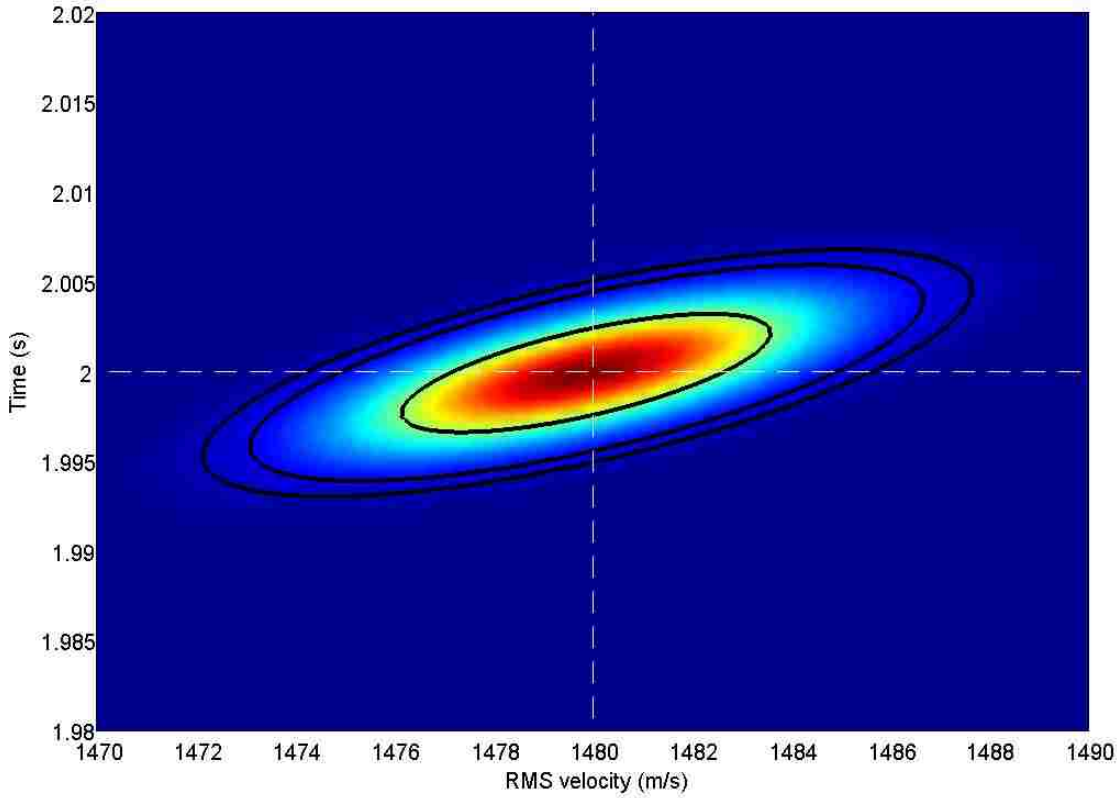

Figure 5.6: Illustration of the joint posterior distribution of the RMS velocity and the zero-offset travel time with the 95%, 90% and 50% credibility contours in black (outer to inner curves) and posterior mean in the intersection of the white dashed lines.

Layer	Zero-offset travel time	RMS velocity
1	2.0	1480
2	2.5	1500
3	3.0	1520

Table 5.1: Zero-offset travel time t_{0_i} (in seconds) and RMS velocity v_i (in metres per second) pairs, $i = 1, 2, 3$ for SM-2a.

was to produce a transformation that maximizes the linear correlation in this case ignoring the physical plausibility of the model, we find that the actual values are not maximal. In Figure 5.9, we have three correlation matrices, the first is the correlation matrix of the untransformed gather $CM^{(orig)}$, the second is the matrix relative to

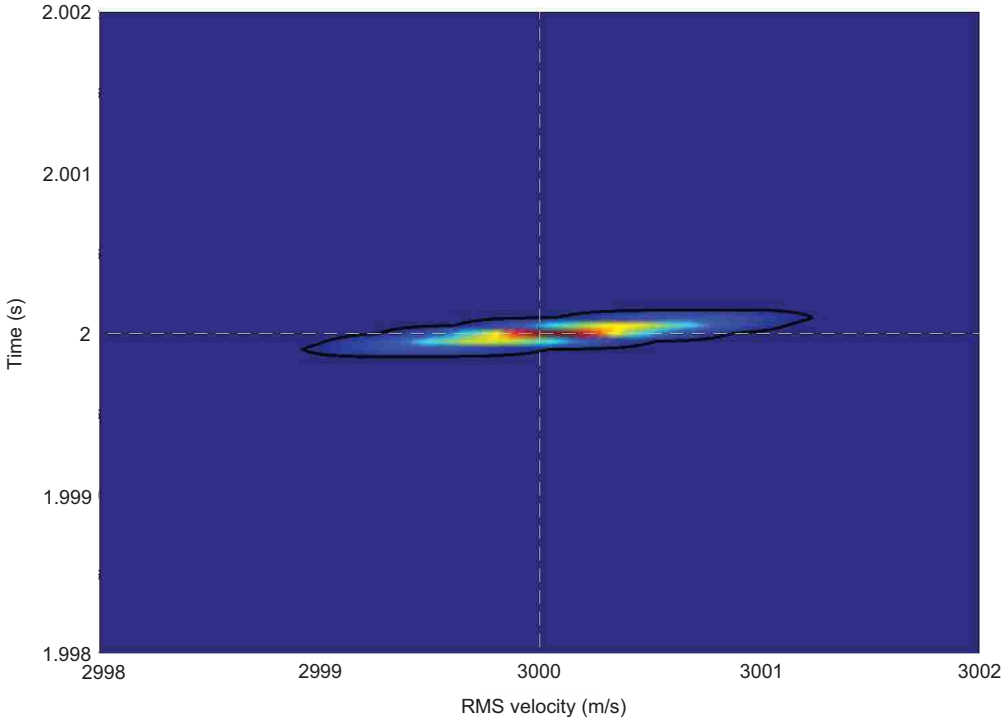

Figure 5.7: Illustration of the joint posterior distribution of the RMS velocity and the zero-offset travel time with the 95% credibility contour in black and posterior mean in the intersection of the white dashed lines.

the transformation using the actual values $CM^{(real)}$ and the last one shows the correlation matrix of the transformation that uses all local maxima in the weighted semblance map $CM^{(max)}$; the weighted semblance map and the local maxima are illustrated in Figure 5.10. Calculating the indexes of flatness for these matrices, we have that $IF_{orig} = -0.0074$ for the untransformed gather, $IF_{real} = 0.8047$ for the gather transformed with the actual values and $IF_{max} = 0.8794$ for the gather transformed with the local maxima.

Even though the local maxima "linearizes" the gather better than the actual values, the local maxima model is implausible; while the actual values produce an increasing piecewise linear interval velocity trend, the local maxima model returns complex-valued interval velocities. So we actually need to find the transformation with the highest index of flatness within acceptable physical conditions, i.e. a piecewise smooth RMS trend which accounts for the fact that the interval velocity trend

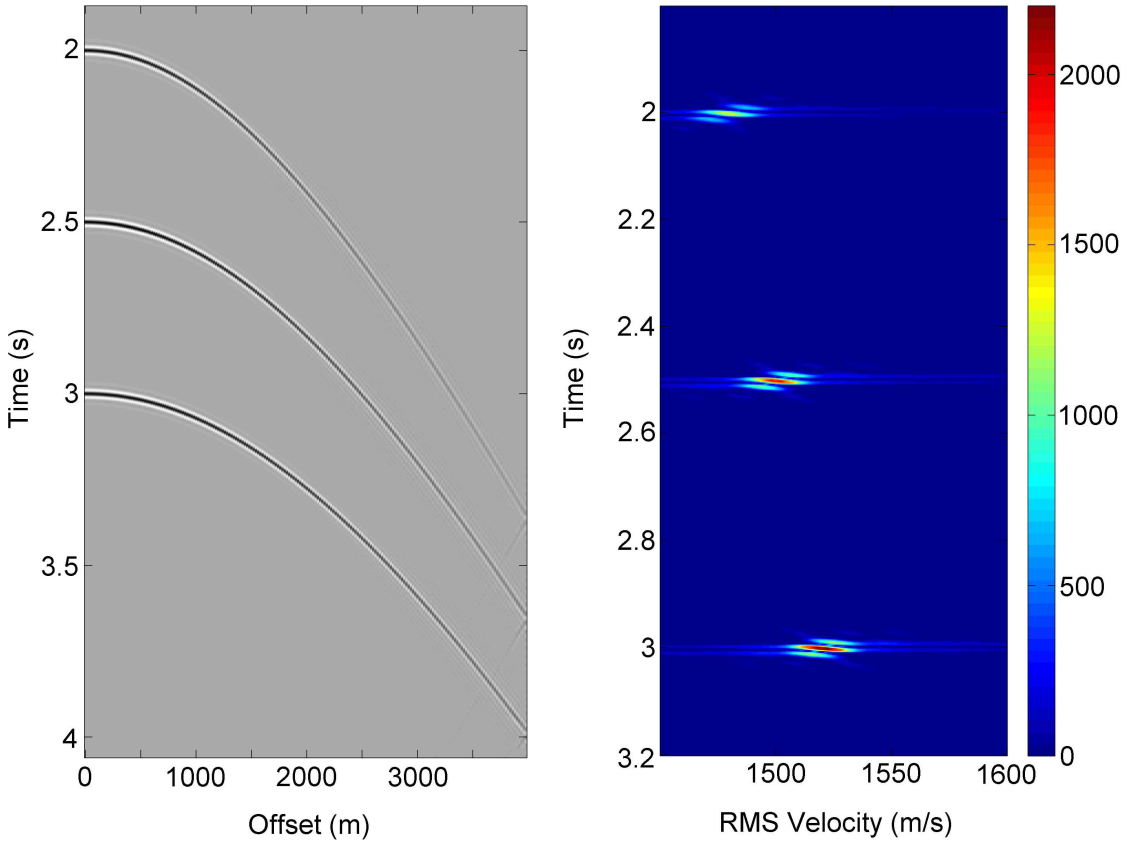

Figure 5.8: On the left, CMP gather for SM-2a and, on the right, its corresponding weighted semblance map.

is real-valued and piecewise constant. Moreover, boundaries and velocity picks must be associated to sequences of local extremes in the CMP gathers. Therefore we are looking for the transformation that maximizes the linear correlation conditional on the plausibility of the velocity trend.

Now we use the picking algorithm and the 1D SS-BRAINS method to estimate the zero-offset travel time and RMS velocity pairs for SM-2a. The points corresponding to the three layers in this model are selected automatically using the picking algorithm under the assumption that the RMS velocities are in the interval $(1400m/s, 1600m/s)$; the local extreme sequences for each layer are clear so there is no ambiguity in their determination.

The results obtained applying the 1D SS-BRAINS are displayed in Table 5.2; the values are consistent with the actual values used to create the synthetic gather and the small deviation of the mean RMS velocity from the actual values in the second

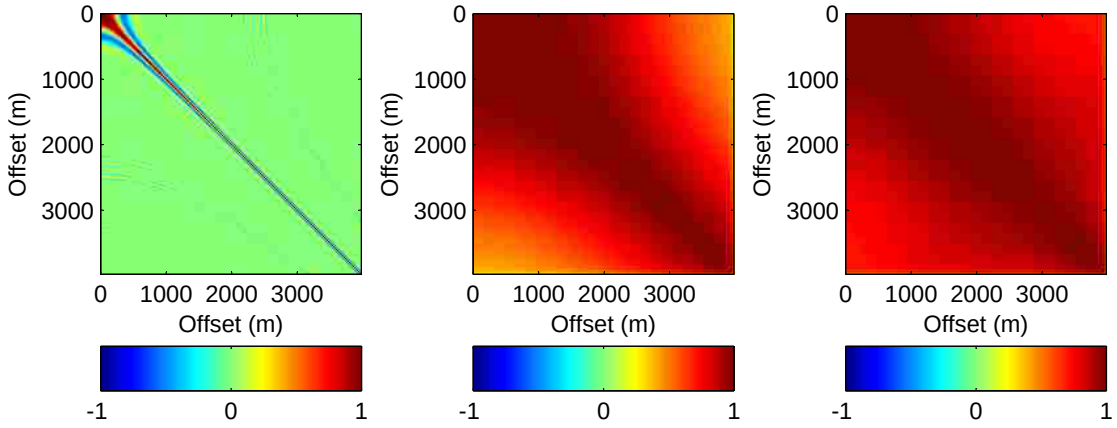

Figure 5.9: Correlation matrices for SM-2a. On the left, correlation matrix of the original CMP gather in SM-2a $CM^{(orig)}$; in the centre, correlation matrix of the transformed gather using the true values in the model $CM^{(real)}$ and; on the right, correlation matrix of the transformed gather using the local maxima in the weighted semblance map $CM^{(max)}$.

and third layers can be attributed to one or more of the following: the hyperbolic approximation, the precision of the M-H algorithm or the precision of the algorithm used to create the synthetic.

Layer	RMS velocity (m/s)				Zero-offset travel time (s)			
	True	Mean	$Q_{0.025}$	$Q_{0.975}$	True	Mean	$Q_{0.025}$	$Q_{0.975}$
1	1480	1480.0	1479.6	1480.2	2.0	2.0	2.0	2.0
2	1500	1500.6	1499.8	1500.4	2.5	2.5	2.5	2.5
3	1520	1520.6	1519.2	1521.1	3.0	3.0	3.0	3.0

Table 5.2: Posterior means and credibility intervals for SM-2a using the 1D SS-BRAINS; the means and credibility intervals of the zero-offset travel times for each layer are the same as the actual values up to the sixth decimal case

In order to verify the impact of the use of the M-H algorithm in these deviations, we also used a fine grid to approximate the posterior distribution and estimated the posterior means and credibility intervals; the results were the same up to the first decimal case for the RMS velocities and up to the sixth decimal case for the zero-offset travel times indicating that these results are consistent and that the

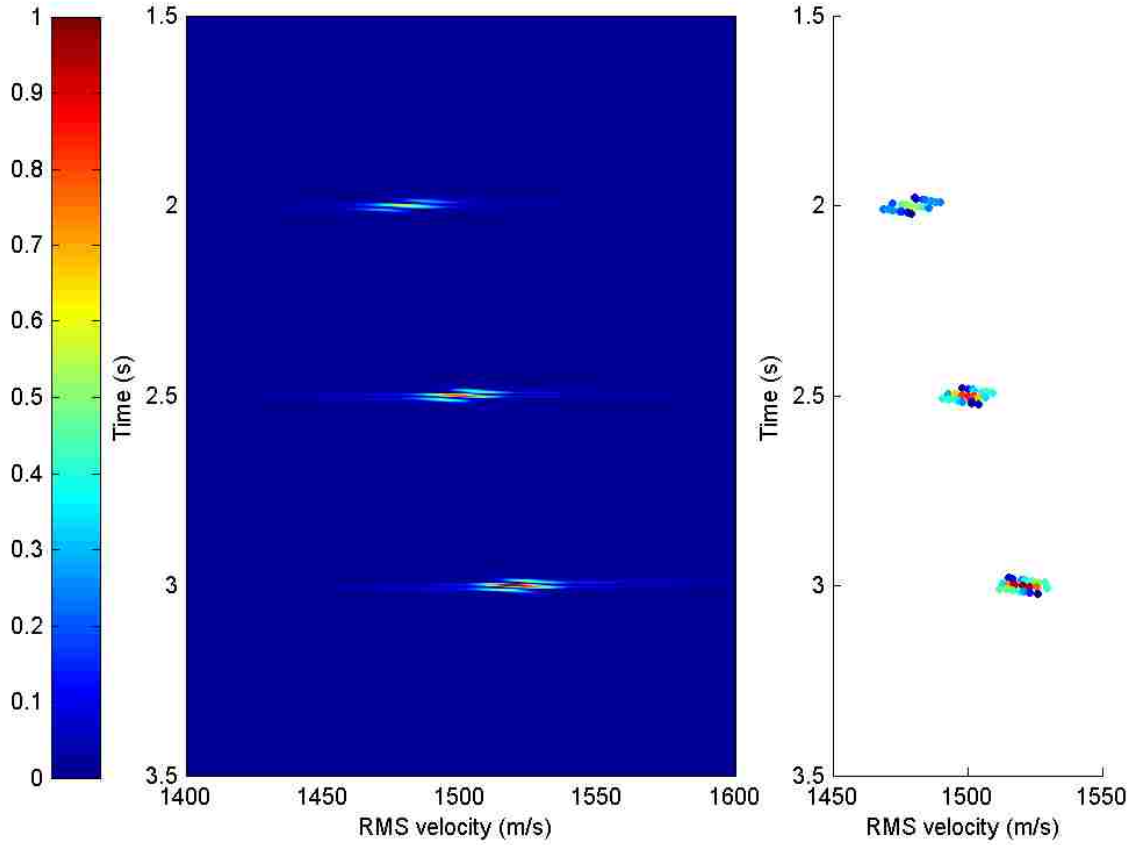

Figure 5.10: Left: weighted semblance map for SM-2a. Right: scatter of the local maxima for the weighted semblance map.

M-H algorithm is a suitable sampler. In this case, we used an independent M-H algorithm with a variable-at-a-time update to draw 10^5 samples and set the burn-in at 10^4 samples; the acceptance rate was approximately 21% and the running means and quantiles for each variable were stable. The precision of the algorithm used to create the synthetic depends on several factors including run time and coarseness of the grid used for the FFT deconvolution but a similar effect could be produced by the discretization of the signal or the application of a filter in real data; therefore, this error is likely to be intrinsic of the data as a type of coherent noise.

Here we used the 1D S-BRAINS and the 1D SS-BRAINS and showed how both can perform to produce point and interval estimates for zero-offset travel times and stacking velocities. While the first returns accurate point estimates for the posterior means and is easier to simulate than the latter, it does not account for the variability in the model correctly so the posterior interval estimates are wider than we would

expect. The 1D SS-BRAINS resolves part of these issues by adapting a flexible fit to the picked travel times which accounts for the trace-by-trace variability. In the next Chapter, we discuss multi-gathered synthetic examples and discuss the error structure within gathers.

Chapter 6

Breaking the BRAINS rules

In the previous Chapters, we made a number of assumptions to model an ideal Earth and we worked with synthetic examples containing a single CMP gather and analysed them using the 1D S-BRAINS and the 1D SS-BRAINS models looking for the best stacking velocity. Analysing a single CMP gather provides information about the local subsurface around its common midpoint; however, we are usually interested in analysing a profile composed by multiple CMP gathers obtained from multiple sources and receivers organized in a line or grid since features that contradict our assumptions usually can only be identified across multiple gathers.

In the first two Sections, we analyse sequences of synthetic gathers and test the flexibility of the BRAINS models when the the assumptions of continuity and parallelism are not valid. In the last Section, we investigate issues related to anisotropy. The synthetics in this Chapter were generated using the phase-screen method in Section 5.1 as described in White and Hobbs (2007).

6.1 A simple discontinuity: breaking the space time continuum

Our first example, synthetic model (3.a) (SM-3a), is a subsurface with a step discontinuity as represented in Figure 6.1. The subsurface is formed by two parallel semi-planes connected by another semi-plane which is perpendicular to both, the first semi-plane is located at 1480m below surface, the second at 1554m and the dis-

continuity at the CMP position of 4000m; the subsurface is assumed to be isotropic and homogeneous with a constant interval velocity of 1480m/s. The array of sources and receivers is linearly placed in a line represented by the black dots in Figure 6.1; the synthetic was modelled in two dimensions with a cylindrical source/receiver spreading function, i.e. we assume that the waves that leave the sources travel on the plane orthogonal to the surface. Given the linear trajectory of seismic rays, we can predict that the signal before and after the discontinuity will be continuous and, close to the discontinuity, unstable due to the interference caused by diffracting rays.

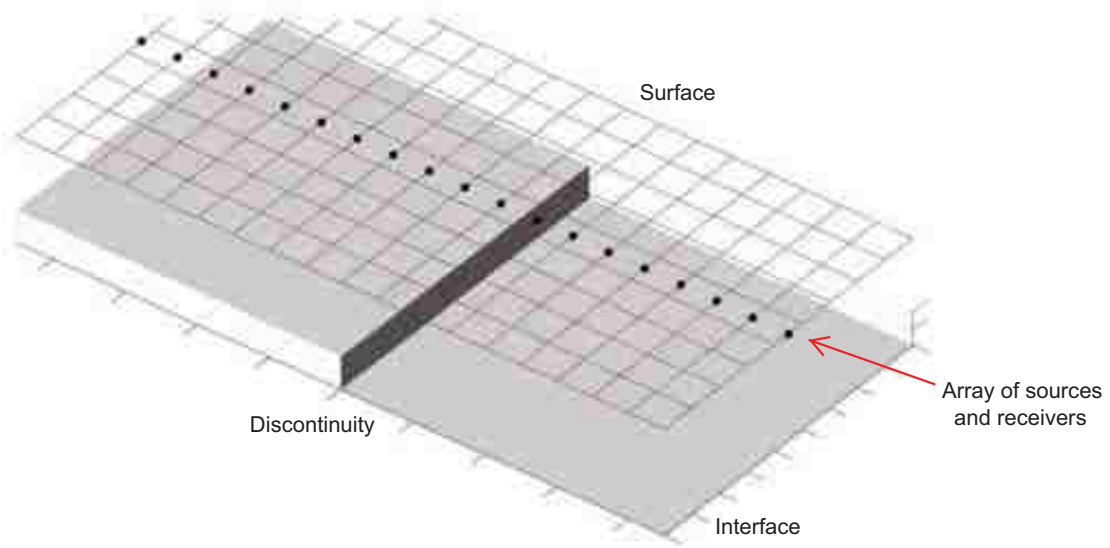

Figure 6.1: Representation of the subsurface in SM-3a which contains one step discontinuity and is assumed to be isotropic and homogeneous with constant interval velocity of 1480m/s. Before the discontinuity the subsurface is 1480m below the surface and, after the discontinuity, 1554m. The surface is represented by the dark red grid and the subsurface by the gray planes; the possible positions for the array of sources and receivers are represented by the black dots in surface grid.

For this model, we have a sequence of 301 CMP gathers equally spaced 10 metres apart. Each gather consists of 88 traces with offset from 0 to 3480 metres for odd gathers and 20 to 3500 metres for even gathers all with a 40 metres spacing and each trace was recorded with a sampling period of 2 milliseconds for 4.094 seconds; therefore, each CMP gather is represented by a 2048×88 matrix of recorded amplitudes with 88 equally spaced traces each containing 2048 samples. Ten of the

odd numbered CMP gathers located, 300 metres apart each, are shown in Figure 6.2. Note that the discontinuity is noticeable when comparing the first and last gathers but presents itself as a possible two-layer or even three-layer model where it is possible to discern diffractions from both the top and bottom of the discontinuity throughout the transition. When approaching the discontinuity, we have two apparent reflections; one originates from the real reflector and the other is an artifact derived from diffractions. Running the picking algorithm in each of these gathers individually, we get the surfaces displayed in Figure 6.3. The surface on the left corresponds to the first set of arrivals and the one on the right is a patch that corresponds to the second set of events. The colour map on the surfaces represents the amplitude of each pick; we can see that the magnitude of the peak amplitude gets smaller for higher offsets and close to the discontinuity. The first is mainly explained by the fact that energy loss increases with the distance travelled by the wave but also includes the AVO effect and, at the largest offsets, an amplitude suppression by the forward modelling used to create the synthetic gathers; the second is explained by the interference caused by diffracting events. Other features produced by the pattern of the diffracting rays are the apparent ridges on the lower and top steps which correspond to the points where the diffracting rays and the reflected rays cross paths and constructively and destructively interfere on the same phase altering the signal.

Now we model each gather separately using the 1D SS-BRAINS method first considering that the surface of first arrivals represents the only true event then, for the gathers where two events were detected, we assume that the second set of arrivals represent the true event and finally, we assume that all events are part of the model. In Figure 6.4, we have the results for the first two cases; in blue, we have the posterior means for the surface of first arrivals and, in dashed red, the posterior means for the surface of second arrivals; the posterior means for the case with both surfaces are the same as the individual surfaces. Both the red and blue trends indicate an instability in the model near the discontinuity but neither seem to be correct on their own. Recalling the index of flatness presented in Section 4.6. We will use the notation IF_S to indicate the value of this measure calculated for the

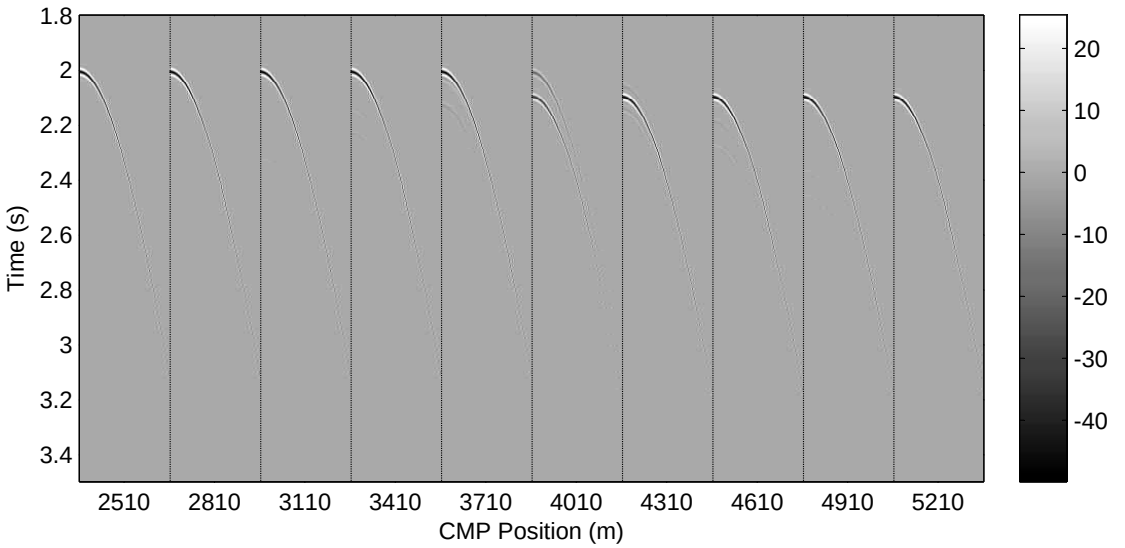

Figure 6.2: Sequence of 10 equally spaced CMP gathers from SM-3a located 300 metres apart. It is implied that each gather has an offset in metres on its x-axis and in this case they all have the same offset range from 0 to 3480 metres and their traces are 40 metres apart.

migrated gather using the NMO correction and the posterior means of the layers in S. Say we use only the first layer to migrate the gather then we calculate $IF_{(1)}$ and if we use both layers, we calculate $IF_{(1,2)}$; to indicate the sum for the untransformed gather, we use the notation $IF_{(orig)}$.

For SM-3a, we calculate $IF_{(1)}$, $IF_{(2)}$ and $IF_{(1,2)}$ for each gather and obtain the results in Figure 6.5. Overall $IF_{(1)}$ seems to be higher than $IF_{(2)}$ almost everywhere apart from the area around the discontinuity. In Figure 6.6, we have the surfaces that represent the best picks according to the index of flatness; on the left plot we have all picks corresponding to the first layer, in the centre we have a subset of the picks corresponding to the second layer, and last we have a plot of the best picks selected using the index of flatness which is a combination of the two surfaces in Figure 6.3. If we model each gather independently, we obtain the posterior means plotted in Figure 6.7; note that the trends are consistent almost everywhere apart from the area around the discontinuity where the velocity curves intercept each other. This model might produce flatter gathers after migration but the velocities are not smooth across gathers.

Figure 6.3: Surfaces of the picks obtained by applying the picking algorithm in each CMP gather of SM-3a, the left surface represents the first set of arrivals and the second surface, the second set of arrivals. The colour map displayed on the surfaces represents the amplitude of the selected pick at each point.

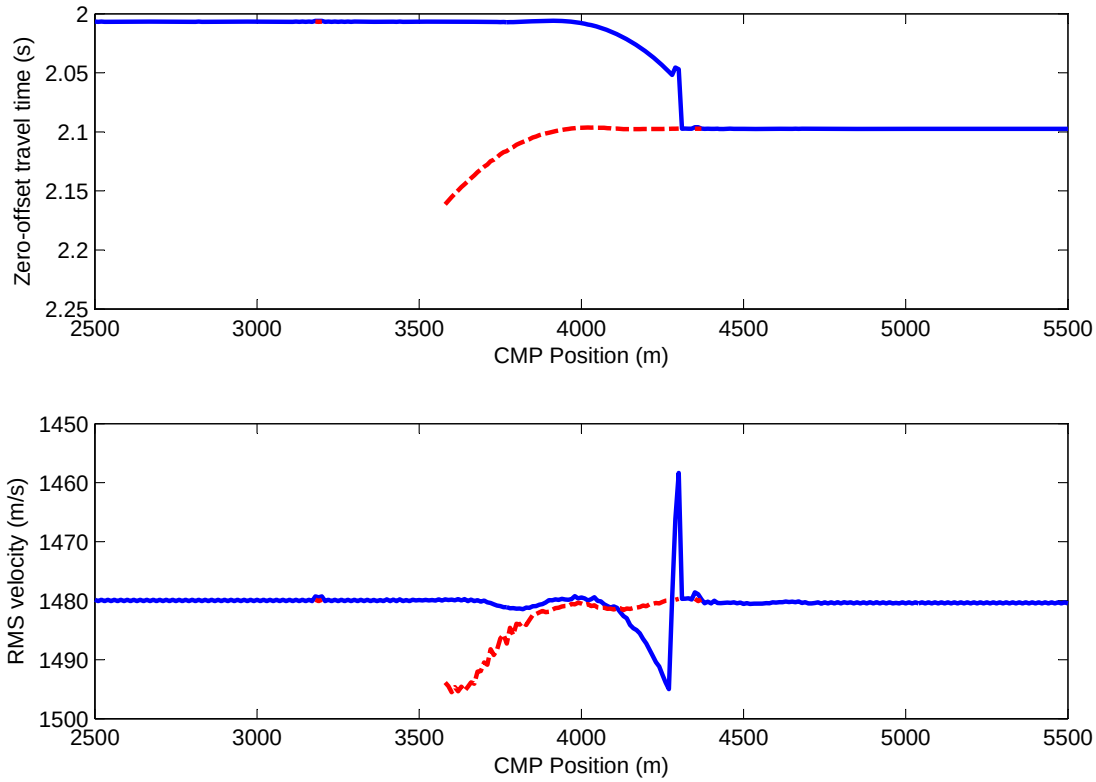

Figure 6.4: Posterior means estimated for the surface of first arrivals (dashed red) and for the surface of second arrivals (blue) using the 1D SS-BRAINS in each gather separately. Top: posterior means for the zero-offset travel times. Bottom: posterior means for the RMS velocities.

Say we assume that there is only one layer in this synthetic and we choose the layer that corresponds to the highest index of flatness, then we obtain the surface in Figure 6.8. Applying the 1D SS-BRAINS to this surface, we obtain the posterior means and 95% credibility intervals in Figure 6.9; the zero-offset travel times and RMS velocities seem to represent the subsurface fairly well almost everywhere with the exception of a few gathers around the discontinuity. This could be a result of the interference caused by the diffracting rays in that region in which case we would not be able to resolve the discontinuity completely or it could be a result of the shape of the likelihood and posterior distribution at those gathers which might contain more than one mode. If that is the case, we should find a better strategy to find a posterior estimator for these variables or we can set a prior that reflects the variation of these variables across gathers; we choose the latter and apply the

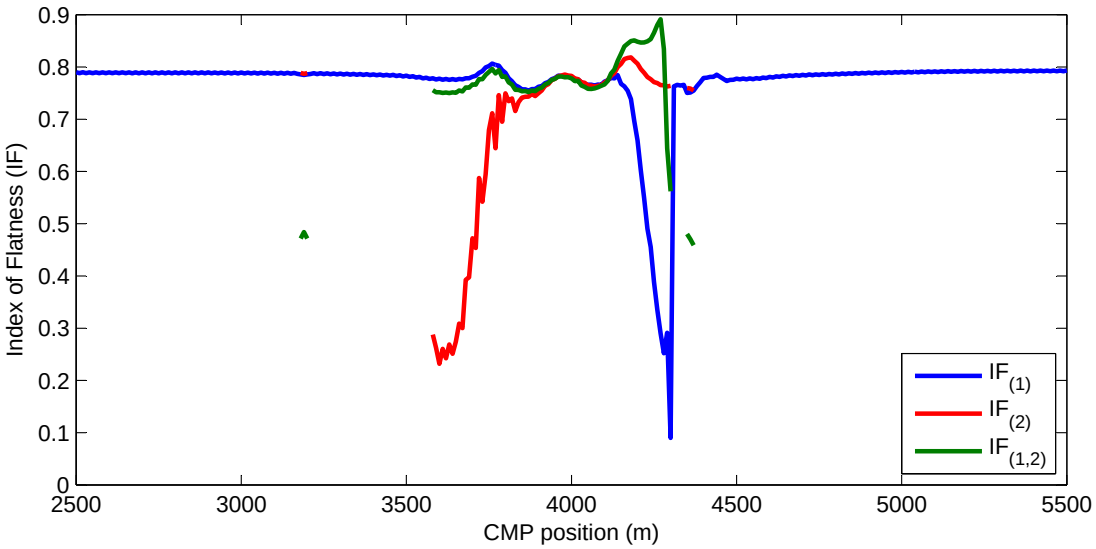

Figure 6.5: Calculated $IF_{(1)}$ (blue), $IF_{(2)}$ (red) and $IF_{(1,2)}$ (green) for each migrated gather using the picks in 6.7.

SS-BRAINS method to these picks.

If we apply the SS-BRAINS model as described in Subsection 3.4.1, we obtain the results in Figure 6.10. In comparison with the results in Figure 6.9, the SS-BRAINS model produces a smoother RMS velocity trend and seems to resolve part of the apparent slopes in the zero-offset travel time trend improving the resolution of the discontinuity; however, one of the edges of the step is still "smooth" and opting for different mean and covariance functions for the priors does not seem to resolve the edge around 4000m; this is probably caused by the effect of diffracting rays in the gathers around the discontinuity and that we are not capable to resolve without enforcing stronger priors; however, our objective here is to test the strengths and weaknesses of these models and identify situations where the results are not the same as the actual values used to create the synthetics and, while enforcing priors might improve the edge of the discontinuity, it does not reflect the effect of diffracting rays in the picking algorithm and the shape of the likelihood. When dealing with real data, it is unlikely that we would have that level of prior information to resolve this discontinuity completely without further processing. The ambiguity in this model comes from the fact that the high amplitude diffractions from the upper surface appear to be better candidates for migrating the data then the true reflections

Figure 6.6: Combined surfaces for SM-3a using the highest IF scores for each gather. Left: selected layers from the first surface in Figure 6.3. Centre: selected layers from the second surface in Figure 6.3. Right: union of these two surfaces showing the selected picks.

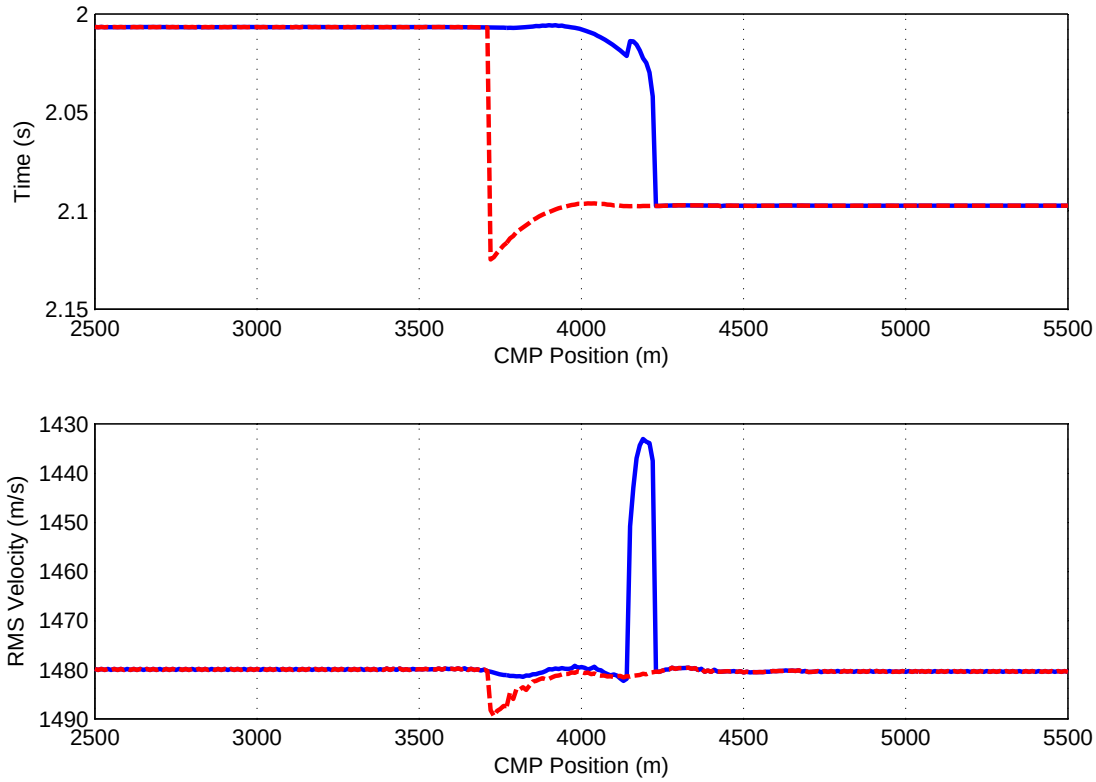

Figure 6.7: Posterior means for the zero-offset travel times and RMS velocities obtained using the 1D SS-BRAINS applied to the picks in Figure 6.6.

around the discontinuity; however, if we submit the data to a 2D prestack time migration using these posterior means, the diffractions would collapse revealing the trend of primary reflections. This prestack time migration is similar to the single-gathered transformations using the travel time equation that we use for migration in the previous Chapters and it is used to improve seismic images (Yilmaz, 2002). With clearer events, one could try to re-estimate the posterior means but that would imply that the original data was used twice and our estimates would not correctly reflect the uncertainty in the original data. However, if we have a large number of gathers as in this example, one can estimate the posterior means using a subset of these gathers as a training set and produce predictions for the remaining gathers. With these predictions, the dataset can then be processed using a prestack time migration and we can repeat the estimation process for the migrated gathers that were not used in the first step. Ideally the migrated gathers should be treated as a new dataset and the results obtained in the previous step should be discarded or

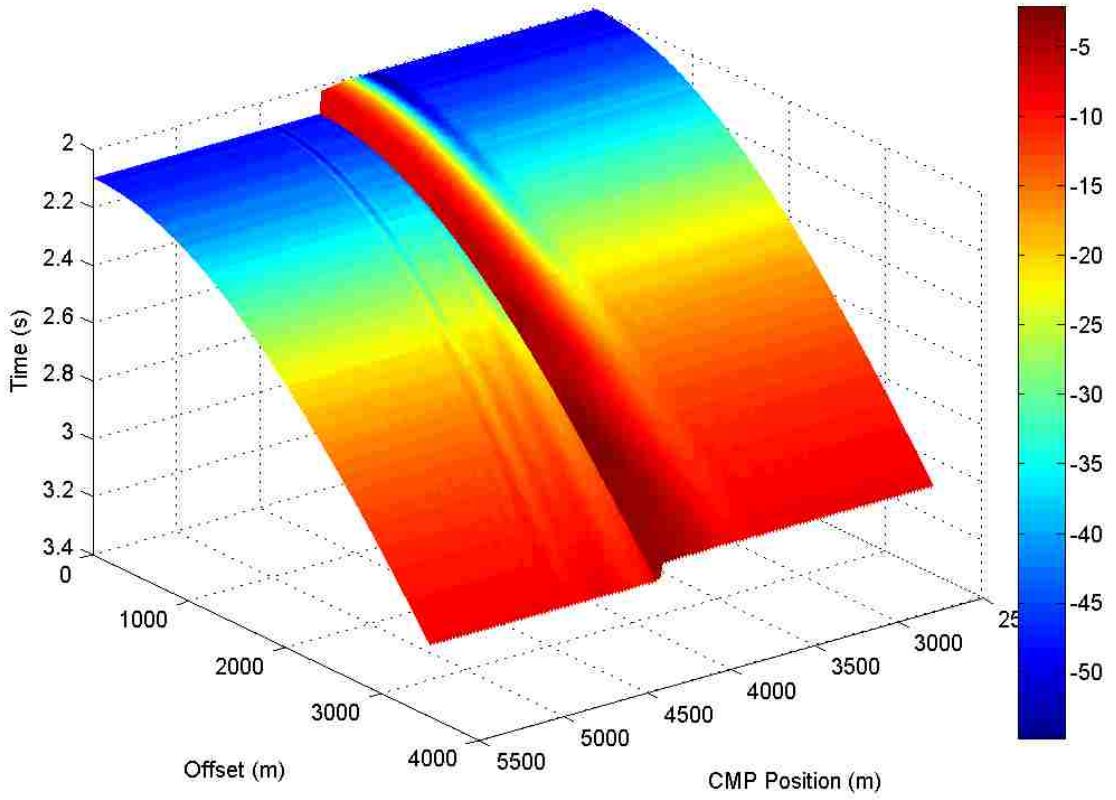

Figure 6.8: Combined surfaces for SM-3a using the highest IF values for each gather assuming that the number of layers in the model is one.

carefully used to update priors. Since seismic surveys in recent days contain large amounts of redundant data, these steps can be repeated more than once to produce a clear image. The downside is that 2D and 3D migrations are computationally expensive and the time taken to migrate a dataset depends on the complexity of the velocity model used therefore migrations should be used only with large datasets and when time is available.

Finally, if we apply the IS-BRAINS model to this data to estimate interval velocities and depths simultaneously, we obtain the results in Figure 6.11. We know that the interval velocity for this model is constant at 1480m/s and that the depth of the subsurface before the discontinuity is 1480m and after the discontinuity is 1554m. The fact that the interval velocity is lower before the discontinuity than after is possibly an anomaly caused by the algorithm used to create this dataset; this effect can also be seen in the results for the SS-BRAINS where the zero-offset travel

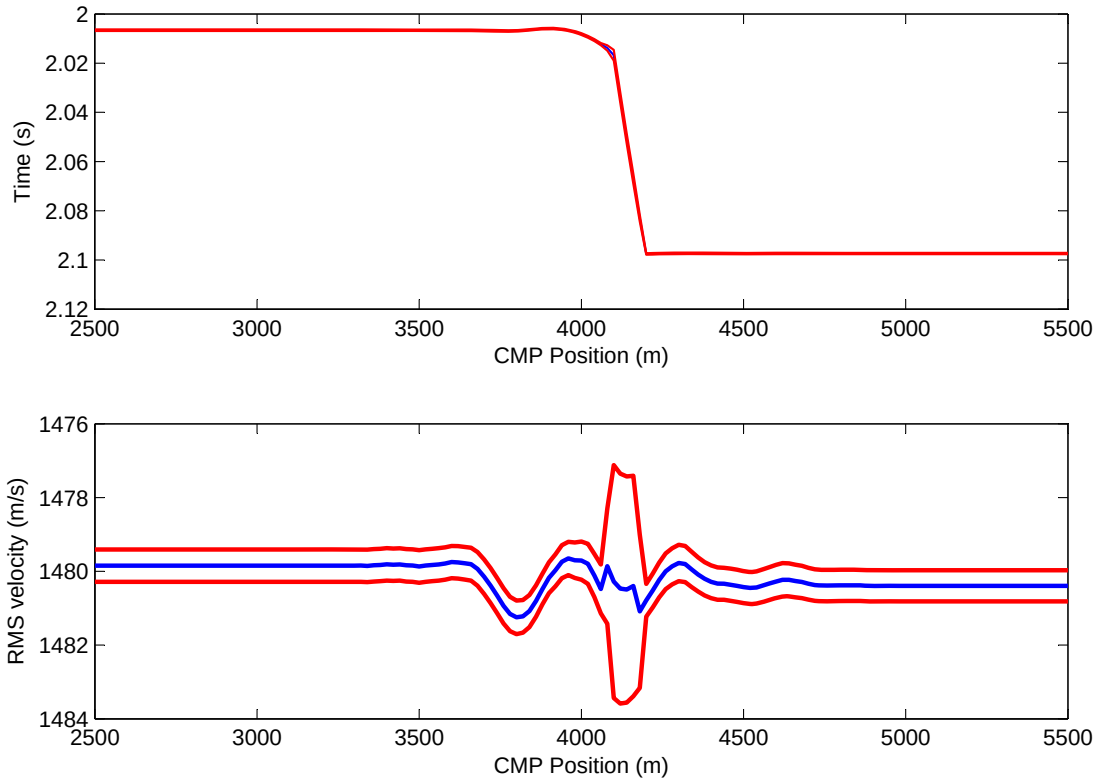

Figure 6.9: Posterior means (red) and 95% credibility intervals (blue) for the zero-offset travel times and RMS velocities for the picks in the surface represented in Figure 6.8 by modeling each gather independently using the 1D SS-BRAINS.

time estimates before the discontinuity is slightly higher than 2s and higher than 2.1s after the discontinuity and this is reflected in the RMS velocity curves as well. This effect put aside, we have a fairly smooth interval velocity trend when compared to the RMS velocity trend and a consistent depth model. If only an approximate depth model is necessary, the results from the IS-BRAINS should suffice. If better resolution of the discontinuity is necessary and only the 2D dataset in use is available, then we can include a step of time migration as described above. Alternatively, if prior information about the location of the discontinuity is available, then we can use this information to track the primary reflections and discard the diffracting events. Next we study a dipping reflector.

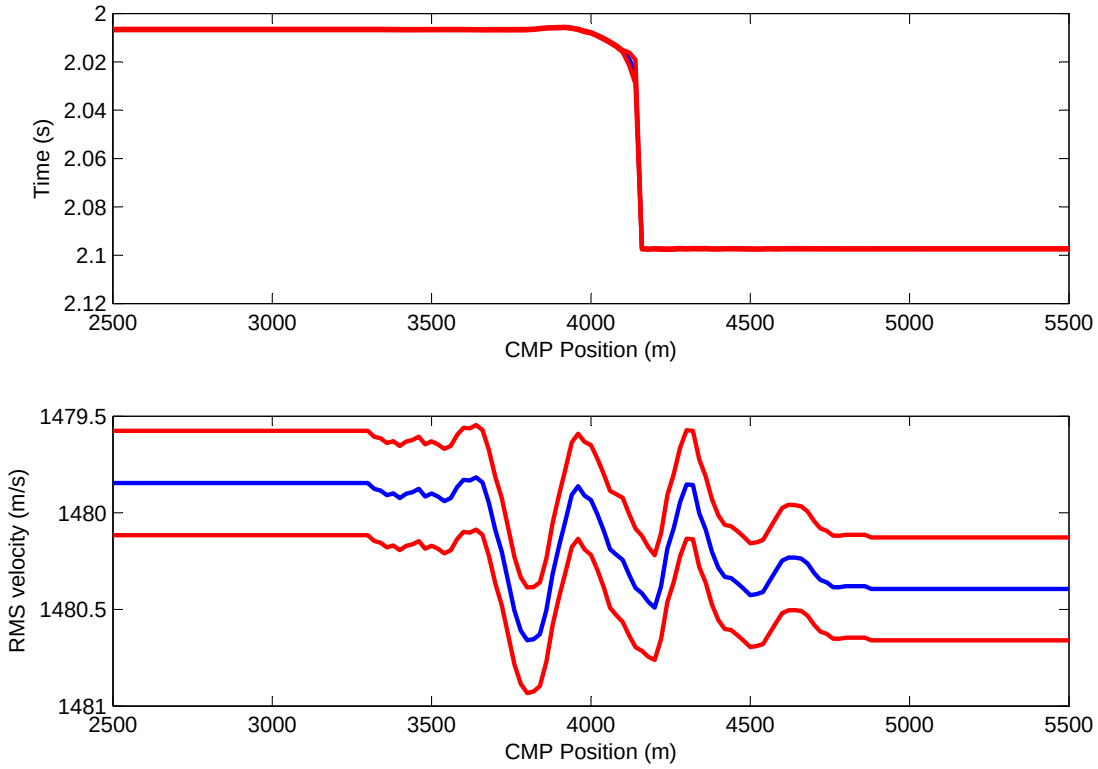

Figure 6.10: Posterior means (red) and 95% credibility intervals (blue) for the zero-offset travel times and RMS velocities for the picks in the surface represented in Figure 6.8 modelled using the SS-BRAINS.

6.2 A dipping reflector: if only the world were flat

Here we have an example of a dipping reflector which consists of a planar subsurface that is not parallel to the surface as represented in Figure 6.12. The interface forms and angle of 5 degrees with the surface in the direction of the array of sources and receivers and the layer is assumed to be isotropic and homogeneous with constant interval velocity of 1480m/s and the minimum distance of the subsurface to the surface is 1480 metres. The synthetic model referent to this example, SM-3b, was created using a cylindrical source/receiver spreading function implying that the line that contains the points of reflection is in a plane orthogonal to the surface that also contains the array or sources and receivers; this reflection line is represented by the green line in Figure 6.12. The synthetic dataset is composed by 101 CMP gathers

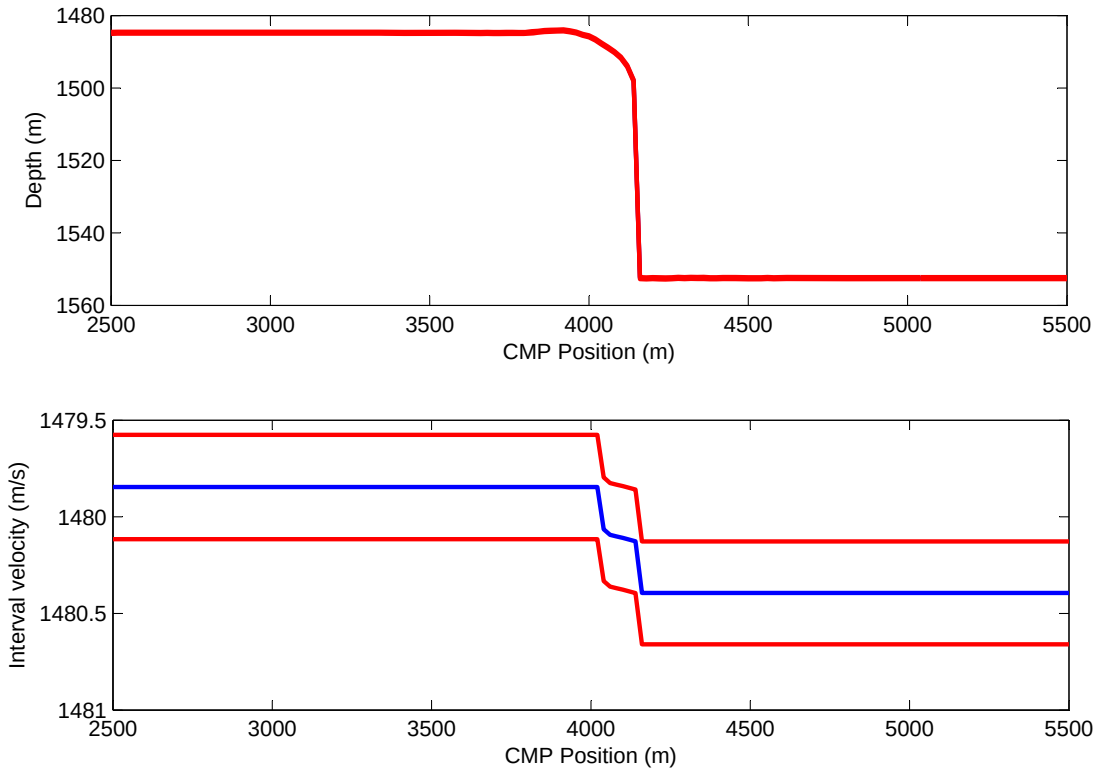

Figure 6.11: Posterior means (red) and 95% credibility intervals (blue) for the interval velocities and depths for the picks in the surface represented in Figure 6.8 modelled using the IS-BRAINS.

with common midpoints 10 metres apart and each gather contains 88 traces sampled at 2 milliseconds for 4.094 seconds. The traces in the odd-numbered gathers have offsets varying from 0 to 3480 metres with a step of 40 metres and the even-numbered gathers have offsets varying from 20 to 3500 metres with a step of 40 metres.

One of our modeling assumptions since Chapter 2 is that subsurfaces are parallel to the surface which implies that the common midpoint of a gather lays on a line orthogonal to the surface which also contains all the reflection points in the subsurfaces; when a dip is present, that is not the case. Following Huygens' (see Theorem 2.2.1) and Fermat's principles (see Theorem 2.4.1), the rays travel through the boundary following the shortest path in time and they are reflected from the boundary with the same angle of incidence relative to the line normal to the boundary which can be seen in Figure 6.14. In Figure 6.13, we have five gathers from SM-3B placed 250 metres apart; we can see that the event in the leftmost gather arrives at

around 2 seconds and this arrival time increases from gather to gather showing signs of dipping. In Figure 6.15, we have a map that represents all traces across gathers with a common offset of 0 metres, i.e. the first trace of each odd-numbered gather, and we can see the pattern formed by the dipping reflections.

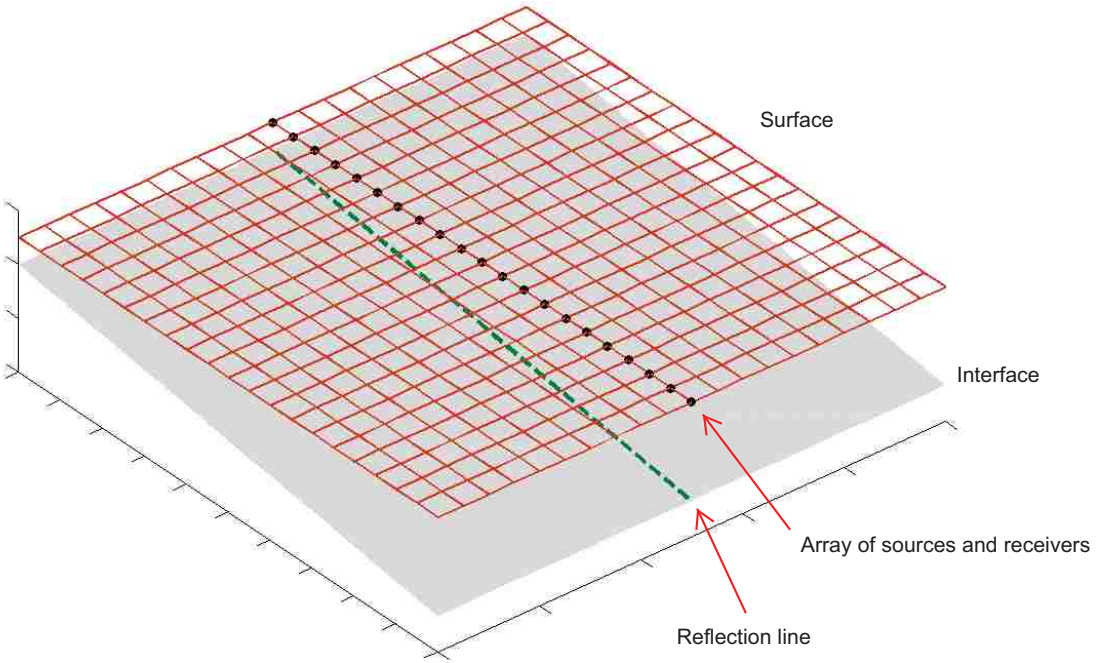

Figure 6.12: Representation of the subsurface in SM-3b; the interface contains a dipping reflector that forms an angle of 5 degrees to the plane containing the array of sources and receivers and the layer is assumed to be isotropic and homogeneous with constant interval velocity of 1480m/s; at its highest point in this plot, the subsurface is 1480m from the surface. The surface is represented by the dark red grid and the subsurface by the gray plane; the positions for the array of sources and receivers are represented by the black dots in the surface grid. We assume that the sources and receivers follow a cylindrical dispersion function so the line of reflection is on a plane orthogonal to the surface that contains the array of sources and receivers; this reflection line is represented by the green line on the graph.

If we use the SET algorithm to select the points corresponding to this event in each gather, we obtain the surface in Figure 6.16 which resembles a section of a hyperboloid slightly rotated around the offset axis; now we apply the IS-BRAINS

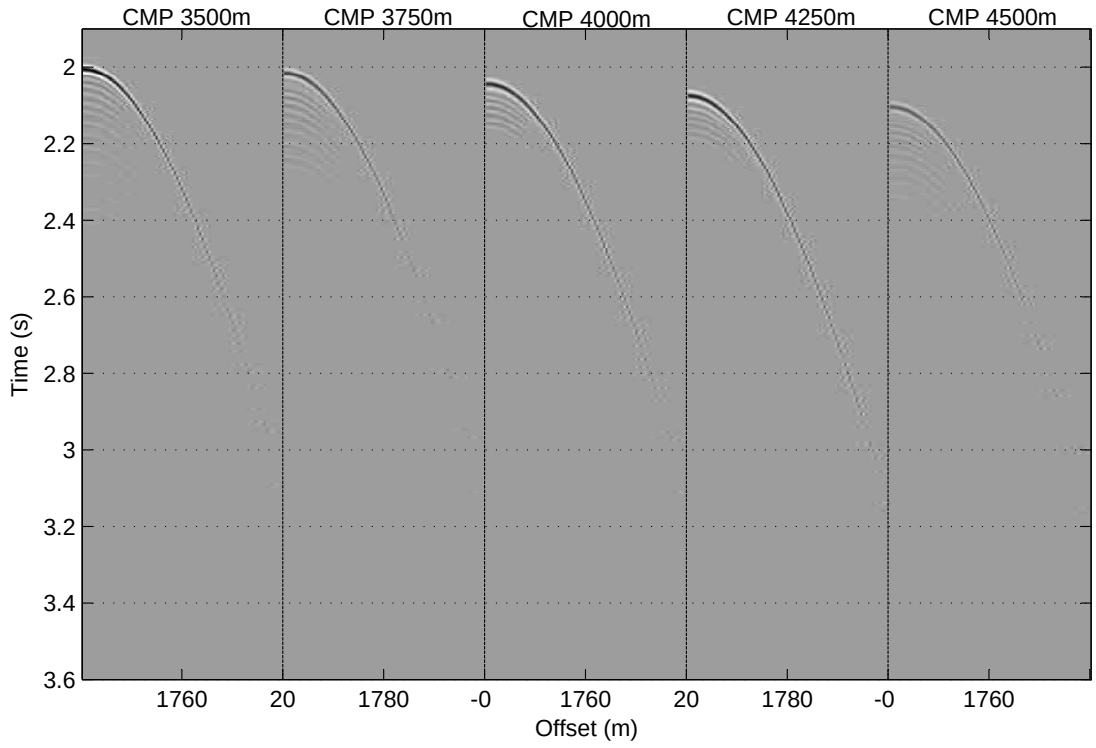

Figure 6.13: Five CMP gathers from model SM-3B placed 250 metres apart. The event arrives at around 2 seconds in the leftmost gather and dips from gather to gather with later arrivals.

to this surface and we obtain the results in Figure 6.17. The black dashed lines represent the actual values used to produce the synthetic SM-3B, the red dashed lines represent the posterior mean from IS-BRAINS and the blue lines, their corresponding credibility intervals. In each gather, we applied the 1D IS-BRAINS with a gradient search algorithm to the odd-numbered traces to point estimate the depth and interval velocities means and covariance function parameters in Equation 3.63 for each gather; we then used the posterior means for the depths and interval velocities as training points for the prior Gaussian processes in the IS-BRAINS model and we assumed that their mean functions were linear. The IS-BRAINS estimation step used the remaining traces in each gather returning the results in Figure 6.17.

It is clear that the estimated interval velocities are incorrect and this is a result of the parallelism assumption, the travel times in each CMP gather still follow a hyperbolic trend but, because the reflection and common midpoints do not coincide,

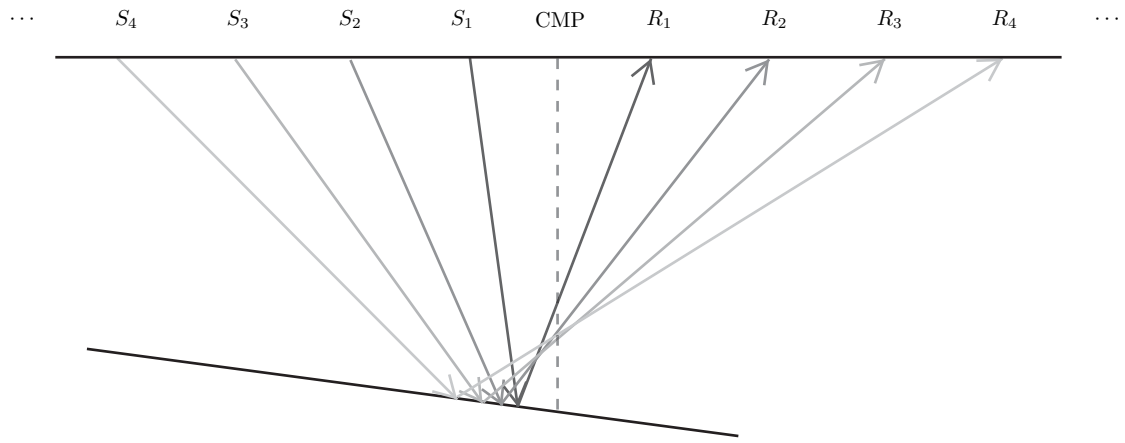

Figure 6.14: Schematic representation of the paths followed by rays fired from source S_j, reflected from a dipping boundary and recorded at receiver R_j, $j = 1, 2, \ldots$ when travelling through a homogeneous and isotropic layer. The points of reflection are not on the same line as the CMP as we would expect with the boundary were parallel to the surface.

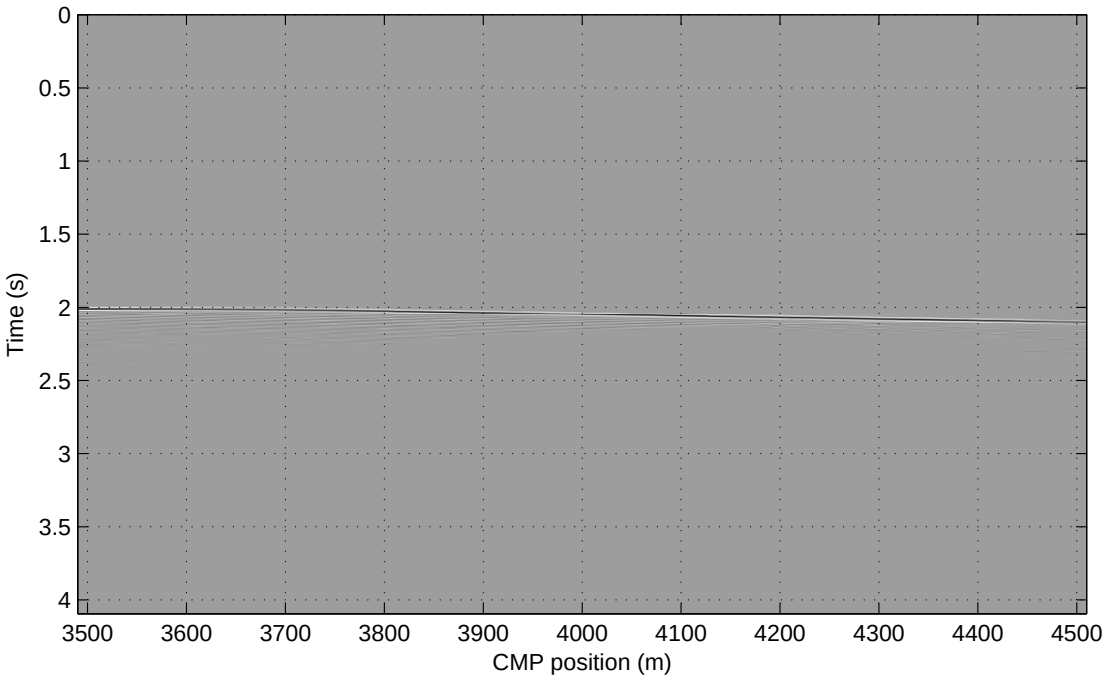

Figure 6.15: Map representing all traces across gather with a common offset of 0 metres, i.e. the plot of the first trace of all odd-numbered gathers.

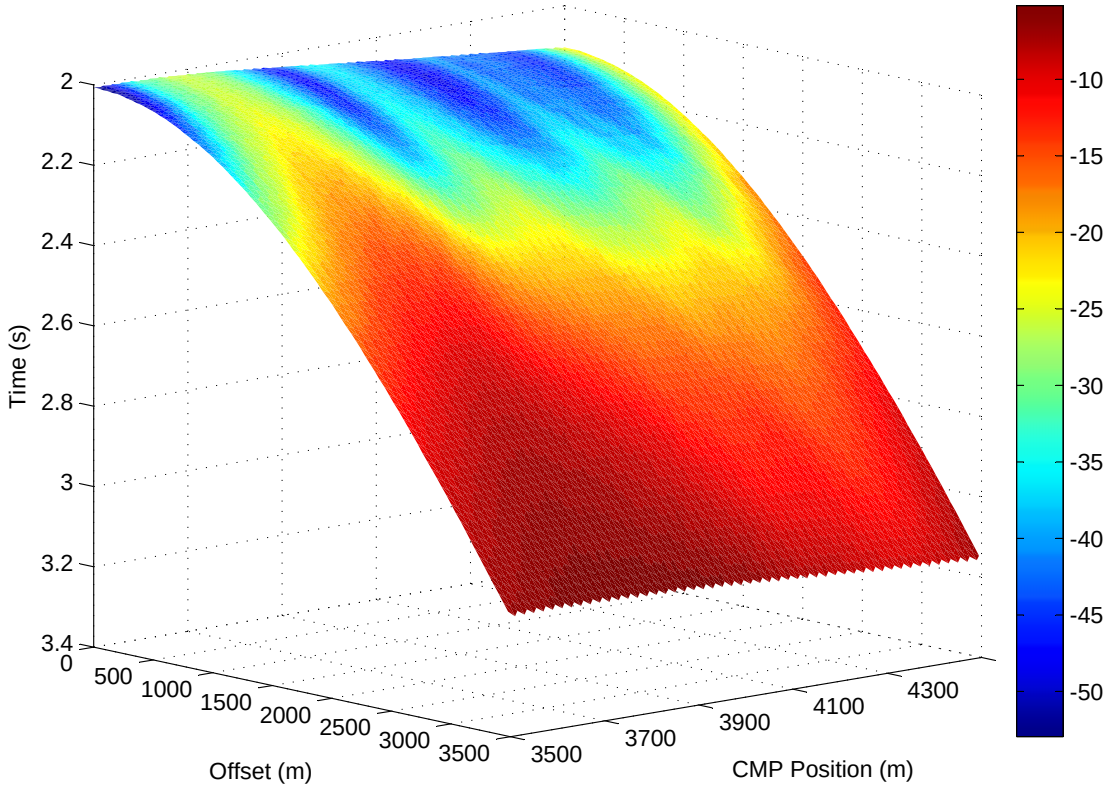

Figure 6.16: Surface of picks for SM-3B obtained using the SET algorithm; the surface colour represents the amplitude of the peaks selected.

the eccentricity of the curve is overestimated. From a single CMP gather, the dip is not obvious so a sequence of CMP gathers is necessary in order to identify the angle of inclination. In the bottom frame of Figure 6.17, we have the depth estimates for the SM-3B, the dip is clear and the depth estimates are close to the real values; if the variable that we are interested to estimate is the depth, then we have an acceptable model. However, if better velocity estimates ares necessary, then a step of seismic data processing called dip moveout (DMO) should be included before proceeding. The DMO compensates for the effects of the dipping interface correcting the CMP gathers (Yilmaz, 2002).

While the models used in this work do not correct for dipping events, they can implicitly estimate the dip in each gather which can then be used in the DMO migration. Returning to the step prior to the fit of the IS-BRAINS, we used the posterior means of the 1D IS-BRAINS to build our prior Gaussian processes. Since the derivative of a Gaussian process is also a Gaussian process (Rasmussen and

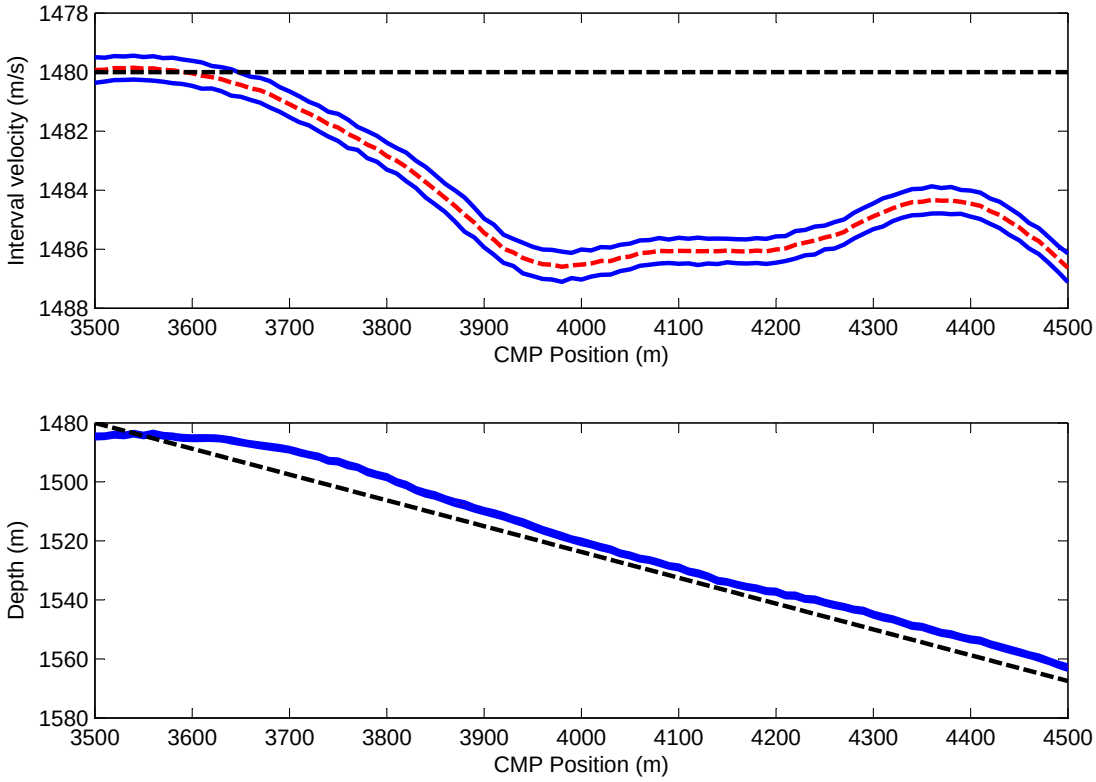

Figure 6.17: Posterior means (dashed red) and 95% credibility intervals (blue) for the surface in Figure 6.16 obtained using the IS-BRAINS and actual values used to create SM-3B in dashed black.

Williams, 2006), we can estimate the gather-by-gather dip and partially correct the depth and velocity estimates as shown in Figure 6.18. The corrected curves are plotted in green and are now closer to the actual trend when compared to the unadjusted trend. This shows that the estimated dip is suitable for a DMO migration and that, if we opt to not apply the DMO correction, we can still correct our estimates.

As discussed in the previous Section, one must be careful when including a prestack migration step to correct the seismic data and that is also the case when applying a DMO migration; it is necessary to ensure that enough data is available for each stage and that traces are not used more than once during the prior definition steps and intermediary analysis in order to avoid false overconfident results; this fallacy is ironically called double dipping (Carlin and Louis, 2000). Next we explore

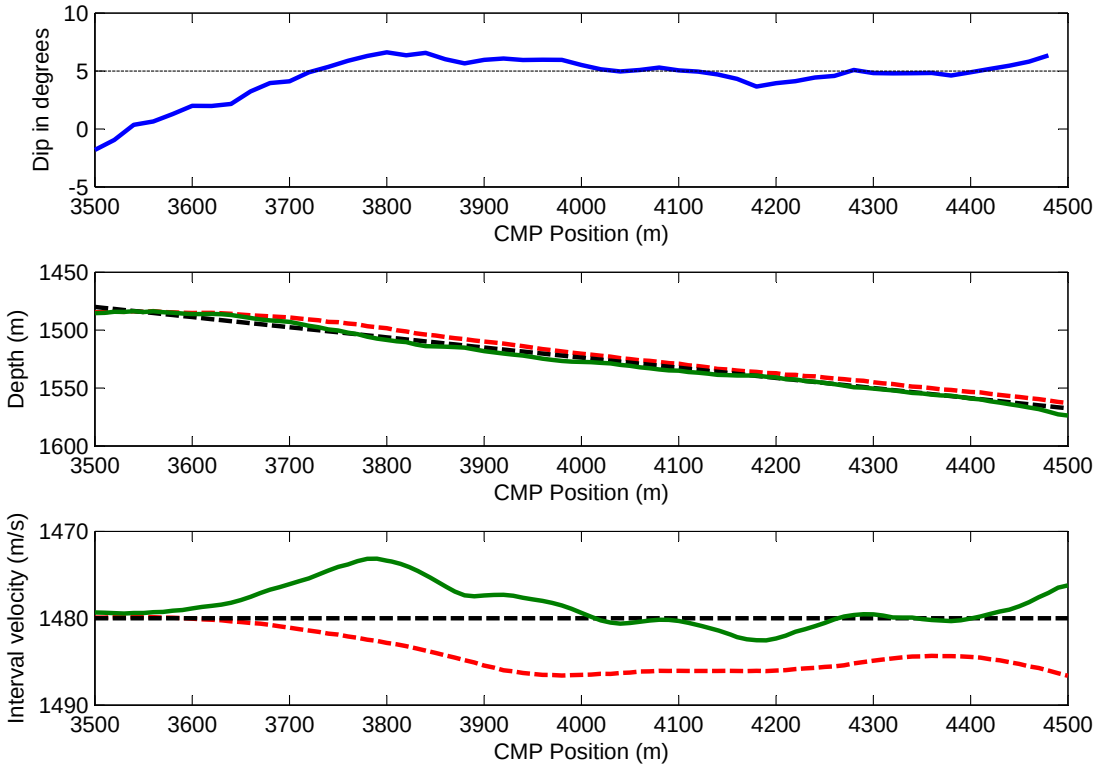

Figure 6.18: Top: estimated dip angles for SM-3b in blue. Middle: adjusted IS-BRAINS posterior depth (green) using the estimated dip angles, unadjusted IS-BRAINS posterior depth (dashed red) and actual depths (dashed black). Bottom: adjusted IS-BRAINS posterior interval velocities (green), unadjusted IS-BRAINS posterior interval velocities (dashed red) and actual interval velocities (dashed black).

an example that defies our assumption of isotropy.

6.3 Anisotropy: because the Earth is not an ideal world

As we defined in Chapter 2, a medium is called isotropic if its physical properties are the same in all directions and anisotropic otherwise; anisotropy is commonly used to refer to a directional variation in velocity. The wavefront that propagates through a media is an indicatrix in the Finsler space, i.e. a spherical image of

a curve in a smooth manifold with a Finsler metric which is a generalization of the Riemannian metric (Rund, 1959; Yajima and Nagahama, 2009). In the case of isotropic homogeneous media, the seismic ray path propagates in an Euclidean space and its corresponding wavefront is spherical. In an inhomogeneous medium, if the velocity of propagation only depends on its position then the Riemannian metric still holds. However, if the ray's velocity depends on its position and direction then its propagation follows a Finsler metric which admits curved ray paths. This implies that the moveout is hardly ever hyperbolic with the exception of spherical and unrotated elliptical wavefronts.

In geophysics, polar anisotropy or transverse isotropy is the most commonly studied case; a medium is transversely isotropic if its physical properties only vary in one direction. In most applications, the assumption of weak transverse isotropy suffices. Thomsen (1986) proposes that three parameters characterize transversely isotropic materials and these parameters are combinations of components of the stiffness tensor from Hooke's law of elasticity. In Tsvankin and Thomsen (1994), the authors rewrite the travel time equation for a parallel reflector in a polar anisotropic medium using two of Thomsen's parameters as

$$t^2 = t_0^2 + A_2 x^2 + \frac{A_4 x^4}{1 + A x^2} \tag{6.1}$$

$$A_2 = \frac{1}{\alpha_0^2 (1 + 2\delta)} \tag{6.2}$$

$$A_4 = -\frac{2(\epsilon - \delta)\left[1 + 2\delta\left(1 - \beta_0^2/\alpha_0^2\right)^{-1}\right]}{t_0^2 \alpha_0^4 \left(1 + 2\delta\right)^4} \tag{6.3}$$

$$A = \frac{A_4}{\alpha_h^2 - A_2} \tag{6.4}$$

where ϵ and δ are Thomsen's parameters, α_0 is the vertical P-wave velocity, α_0 is the horizontal P-wave velocity and β_0 is the S-wave velocity in the medium.

It is clear that the moveout is no longer hyperbolic and that this equation holds more information about travel times; however, with six unknowns and only four equations, this system does not have an unique solution resulting in ambiguity. In order to simplify this problem, Alkhalifah and Tsvankin (1995) defined

$$\eta = \frac{\epsilon - \delta}{1 + 2\delta} \tag{6.5}$$

and simplified Equation 6.2 ignoring the effect of β_0 as

$$t^2 = t_0^2 + \frac{x^2}{v_{NMO}^2} - \frac{2\eta x^4}{v_{NMO}^2\left[t_0^2 v_{NMO}^2 + (1+\eta)x^2\right]}. \tag{6.6}$$

This result solves the problem of underdetermination that we have in Equation 6.2; however, the NMO velocity is not the same as the RMS velocity or the interval velocity, in fact, Thomsen (1986) shows that

$$v_{NMO} = \alpha_0\sqrt{1+2\delta} \tag{6.7}$$

where α_0 is the vertical P-wave velocity and δ is one of Thomsen's parameters; again we have the same problem and we are unable to correctly estimate the actual velocity of the model.

While the determination of the equations that correct anisotropic events is useful, albeit complicated, it seems to go unnoticed that in cases of weak anisotropy in real datasets, the isotropic fit is likely to be in the region of credibility of an isotropic fit and, ultimately, we are interested in estimating depths and interval velocities, not only finding the "flattest" fit. If just the travel time equations are used to estimate the anisotropy parameters, the problem of underdetermination becomes common place; other solutions are available using dynamic ray tracing (Cerveny, 2001) and $\tau - p$ transforms (van der Baan and Kendall, 2002) but these are not discussed here. Now we analyse a synthetic example with vertical anisotropy.

In Figure 6.19, we have the CMP gather for synthetic SM-1c and its weighted semblance map. The reflector in this synthetic is 3000m deep with a horizontal velocity of 3000m/s and Thomsen's parameters of $\delta = -0.035$ and $\epsilon = 0.110$ which are typical values for shales (van der Baan and Kendall, 2002). Calculating the anisotropy parameter in Equation 6.5, we have that $\eta = 0.1559$. Under visual inspection of the CMP gather, there is no clear evidence of anisotropy but the semblance map appear to be asymmetric.

In Figure 6.20, we have a plot of the calculated index of flatness using the travel time equation in 3.19 for RMS velocities in the interval (2500m/s,3500m/s); the point where the IF is higher corresponds to a velocity of 3019.4m/s and we use this velocity to migrate the CMP gather in Figure 6.19 obtaining the migrated gather in Figure 6.20. If the migration were successful, we would expect the event to be

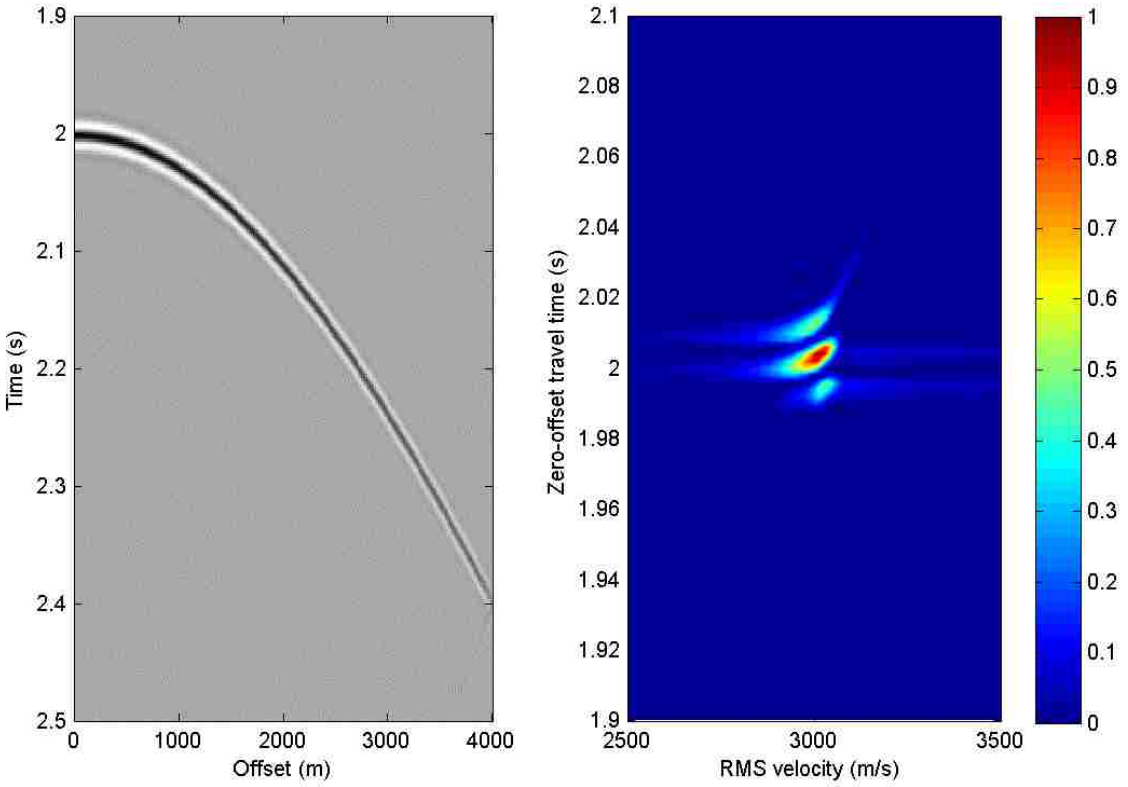

Figure 6.19: On the left, CMP gather for synthetic SM-1c and, on the right, its corresponding weighted semblance map.

completely flat but instead it seems to oscillate around a horizontal line; this is caused by the fact that we are using a hyperbolic moveout.

If we fit the moveout equation in 6.6 to the points selected using the SET algorithm plotted in black in Figure 6.21, then we have that $\eta = 0.1152$ and $v_{NMO} = 2924.3$m/s returns the highest IF. The estimate for η is affected by the assumptions made when simplifying Equation 6.2 to Equation 6.6. In Figure 6.21, we have the fit using Equation 3.19 in blue and the fit using 6.6 in red. Even though the latter offers the best fit, the first offers an estimate for interval velocities under the assumption of isotropy.

Applying the 1D IS-BRAINS to these picks, we have that the 95% credibility intervals are $(3017.9\text{m}, 3.0307\text{m})$ and $(3015.8\text{m/s}, 3027.1\text{m/s})$ for depth and interval velocity respectively which are not far from the actual values in this model. In cases where anisotropy is weak or cannot be distinguished from other types of error, we can assume that the hyperbolic approximation is suitable to estimate interval

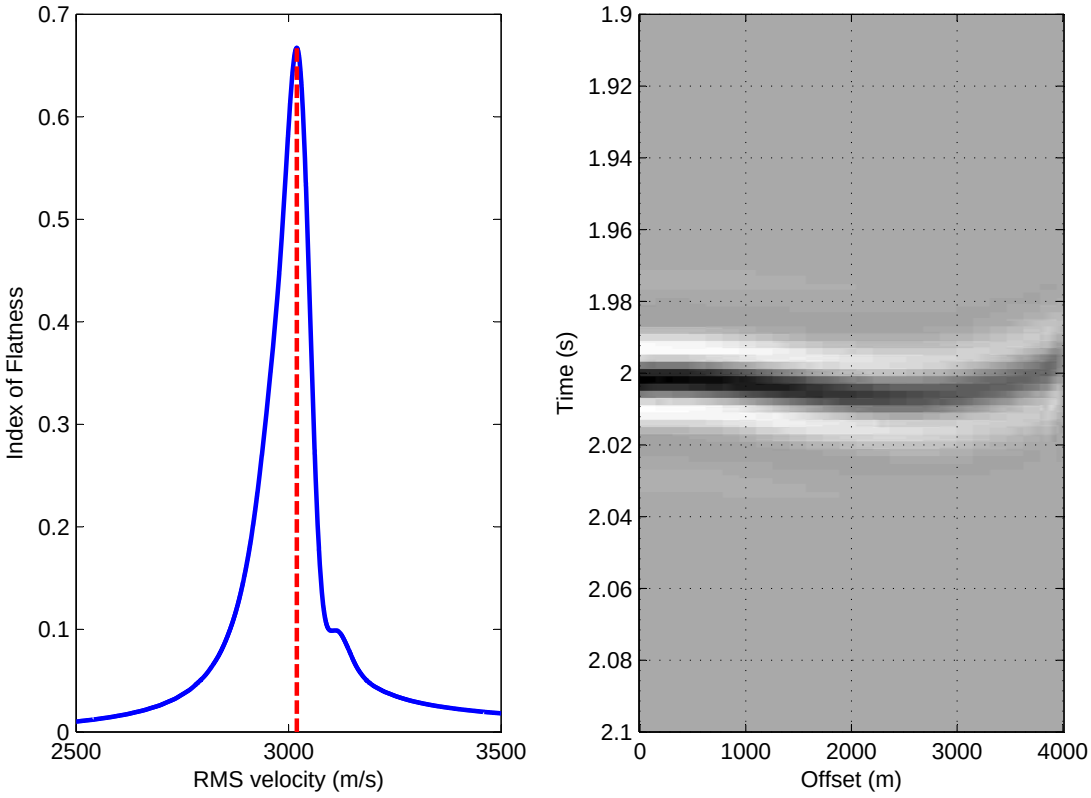

Figure 6.20: Left: plot of the calculated index of flatness for RMS velocies from 2500 to 3500 metres corresponding to model SM-1c. Right: migrated gather from SM-1c using the RMS velocity with the highest index of flatness.

velocities and depths; when stronger anisotropy is present, the incorrect fit will lower the index of flatness and our posterior estimates will be less reliable.

Here we discussed anisotropy and presented an example to illustrate the error derived from a hyperbolic fit to a non-hyperbolic trend. Because the estimation of depths and interval velocities is not trivial in these cases, we choose not to generalize the BRAINS class to include Thomsen's or Tsvankin approximations in this work; this will be a subject to be discussed further in future works. In most situations, the models based in the assumption of isotropy will be satisfactory and the error caused by such assumption will impact the estimates. Now that we studied the main weaknesses of our models based on the geophysical assumptions used to build our models, we can analyse real datasets.

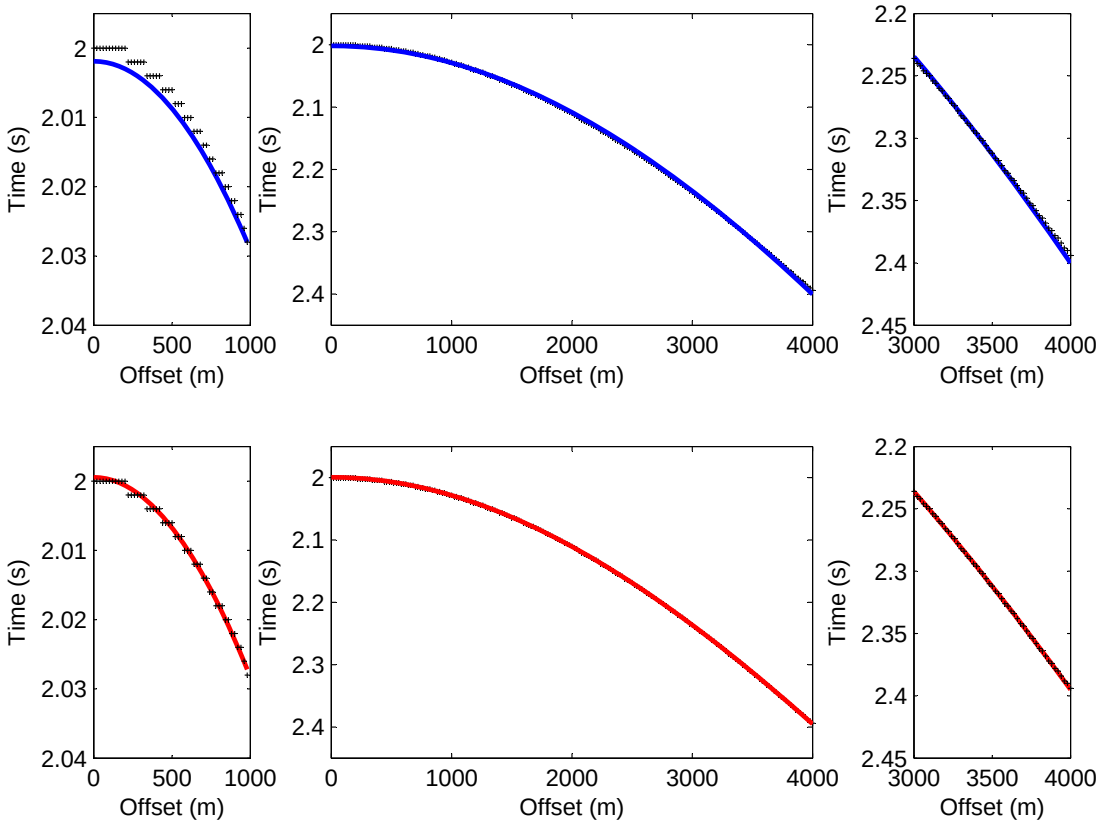

Figure 6.21: In all graphs: selected points from the CMP gather in SM-1c using the SET algorithm are represented in black. Top plots: the curves in blue represent the fit using the hyperbolic moveout in Equatil 3.19. Bottom plots: the curves in red represent the fit using the anisotropic curve in Equation 6.6.

Chapter 7

BRAINS in the real world

In this Chapter, we analyse a sequence of gathers from a real dataset which is part of a seismic survey acquired over the Naturaliste Plateau and Mentelle Basins off the south west coast of Australia (Borissova, 2002). The acquired data is from a deep water environment where complications in the received signal due to reverberation of seismic energy in the water layer (i.e. water multiples) can be ignored because the travel-time for these arrivals is longer than the target primary reflections. The research interest in this survey is the high latitude Cretaceous black shales that were deposited on this margin during a period of extreme high global temperatures which may be related to the sudden decreases in atmospheric carbon dioxide concentration (Kuypers et al., 1999) and recorded sporadic short-lived glaciations (Bornemann et al., 2008).

The seismic data analysed is part of the Geoscience Australia Survey 280 which acquired 2700 kilometres of seismic data recorded to 12 seconds using a 4900 cubic inches airgun array as source and a 636 channel digital streamer with a 12.5 metres spacing resulting in a 6.25m CMP spacing; the data was submited to basic processing including a Radon multiple suppression filter (Foster and Mosher, 1992), DMO correction and stacking (Sargent et al., 2011). The seismic lines in this survey are illustrated in Figure 7.1 and are represented by green lines; the line labeled as 280/05 represents the line we will analyse here.

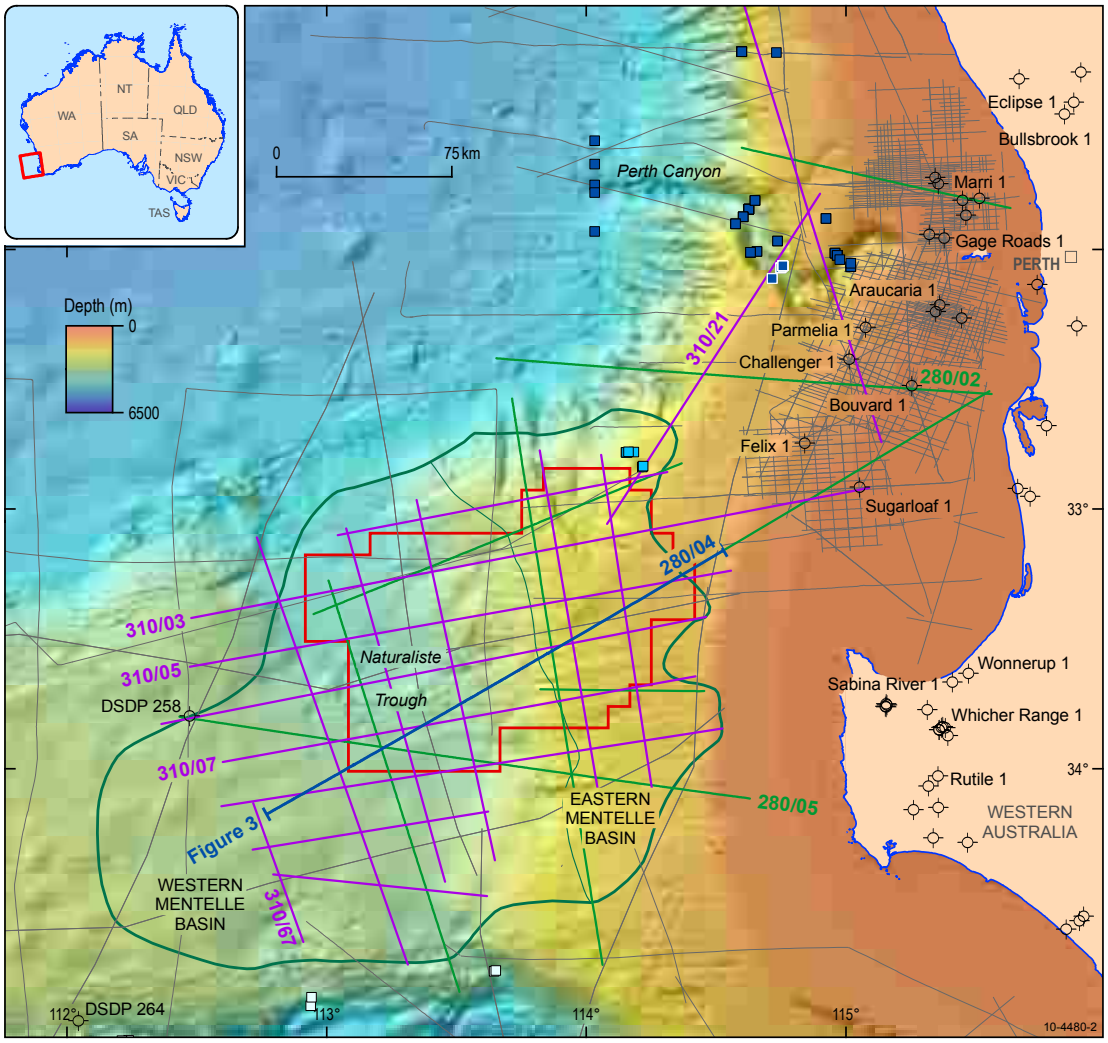

Figure 7.1: Location of seismic lines, well and dredge sampling sites in the Mentelle Basin region extracted from Borissova et al. (2010). The green line labeled as 280/05 represents the location of the collection of the data used here which passes through the drilling point DSDP-258.

7.1 Exploring the data

In Figure 7.2, we have the S280-501 seismic profile created using the pre-processed dataset; it shows how possible events vary across CMP gathers with time. The topmost reflection starting at around 3.7 s in the first CMP gather correspond to the seabed reflection forming the basin shape. In Figure 7.3, we have a plot of the gather with CMP 1240 and its semblance map. The gather is formed by 42 traces sampled at a sampling period of 2 milliseconds for 12.288 seconds each 75 metres

apart. Unlike the synthetic gathers that we analysed in Chapter 5, it is not clear which events are primaries and which are multiples and also which primaries are geologically significant; there are a number of possible models that could correspond to the true one.

First we use the seismic event tracing (SET) algorithm from Section 4.7 to identify the most likely candidates in CMP gather 1240. In the first step of the SET algorithm, we set the minimum peak height as the average of the absolute amplitudes of the last 2 seconds of recorded data. In the fifth step, we limit the search area, by assuming that RMS velocities range from 1400 and 5000 metres per second and we also set a minimum semblance value of 0.01. Using this setup, we select 61 candidates and then we apply the S-BRAINS model to each candidate individually; in this case, we used vague priors assuming that zero-offset travel times varied between 0 and 12.288 seconds and that RMS velocities varied between 1400 and 5000 metres per second. Since each layer was fitted separately, no interval velocity priors were used. From these estimates, we obtained a rough RMS velocity trend which is displayed in Figure 7.4. As we would expect, the RMS velocity trend seems to increase with time; however, the primary events are still not clear.

Selecting the best n candidates without prior information is not an easy task mainly because the number of gathers is undefined and there is a large number of models that are likely to be acceptable. If we are interested in comparing models, we can use the index of flatness defined in Section 4.6 with the criterion that models that produce complex-valued interval velocities are discarded. That would be a way to select the best model or models but that implies that we need to migrate the CMP gather and compute the index of flatness 2^n times. In this example, we have 61 candidates, that would imply that we would have to calculate this index 2^{61} times; in a top-of-the-range machine, each migration of a 42×6144 matrix and the computation of its IF takes around 0.01 seconds, so repeating this process using 8 processors in parallel 2^{61} times would take around 2.8×10^9 years. We could also opt for a backward elimination model where we start with all 61 candidates and then eliminate them one by one according to their signicance but incorporating prior information first would be more efficient. Say our expert would be satisfied with

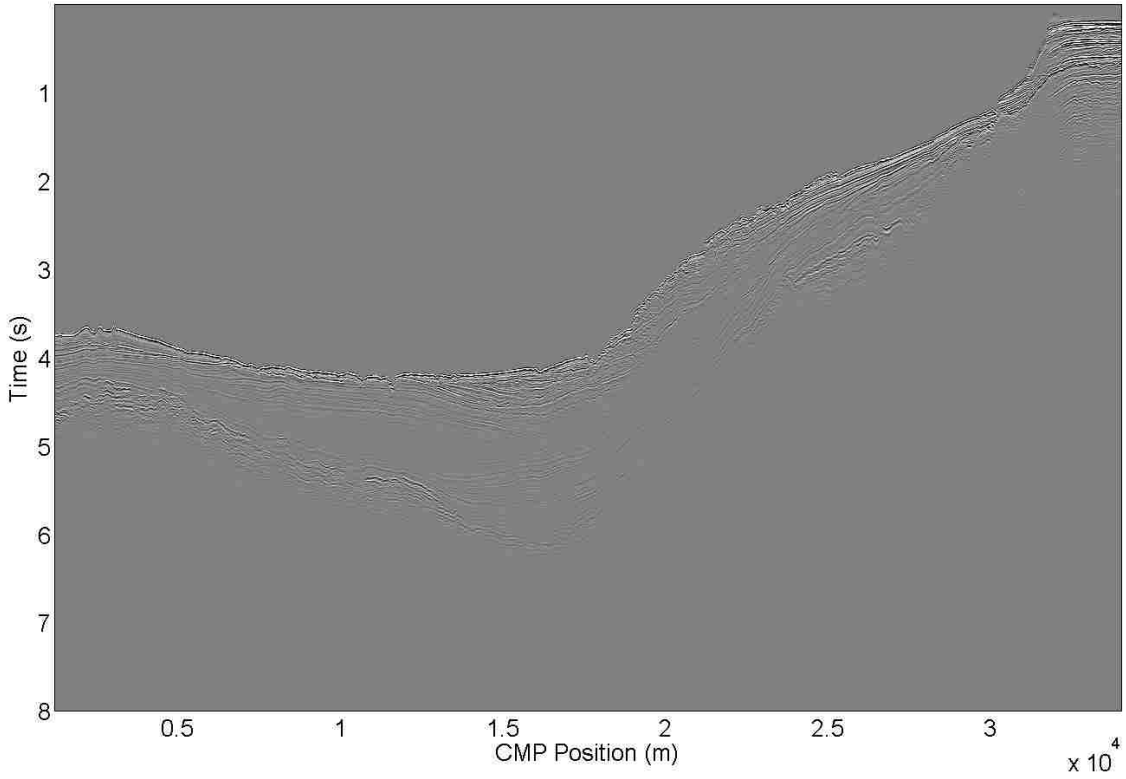

Figure 7.2: S280-501 seismic profile created using the original data.

a model that contains up to 10 layers with interval velocities limited to a certain range say, (1450m/s, 3000m/s), and minimum semblance of 0.6, then we eliminate 30 of our candidates and a handful of combinations that do not satisfy the conditions imposed on the interval velocities. This process is still quite time consuming with 31 candidates if the model remains consistent with the addition of more layers since a large number of models would still be geophysically acceptable and highly probable within these constraints. If all we have available is one gather, then we can make our conditions stricter, ask for more expert input and/or opt for a more efficient elimination criteria. If multiple gathers are available, then, for a given candidate, we can verify if there is a corresponding candidate in an adjacent gather in the neighbourhood of the current candidate.

Even if we select a plausible model across gathers in a reasonable amount of time, there is a risk that the expert will not find it geologically acceptable. A solution for this problem is to include prior information about the location of the boundaries through zero-offset travel time picks obtained from another survey, a seismic profile

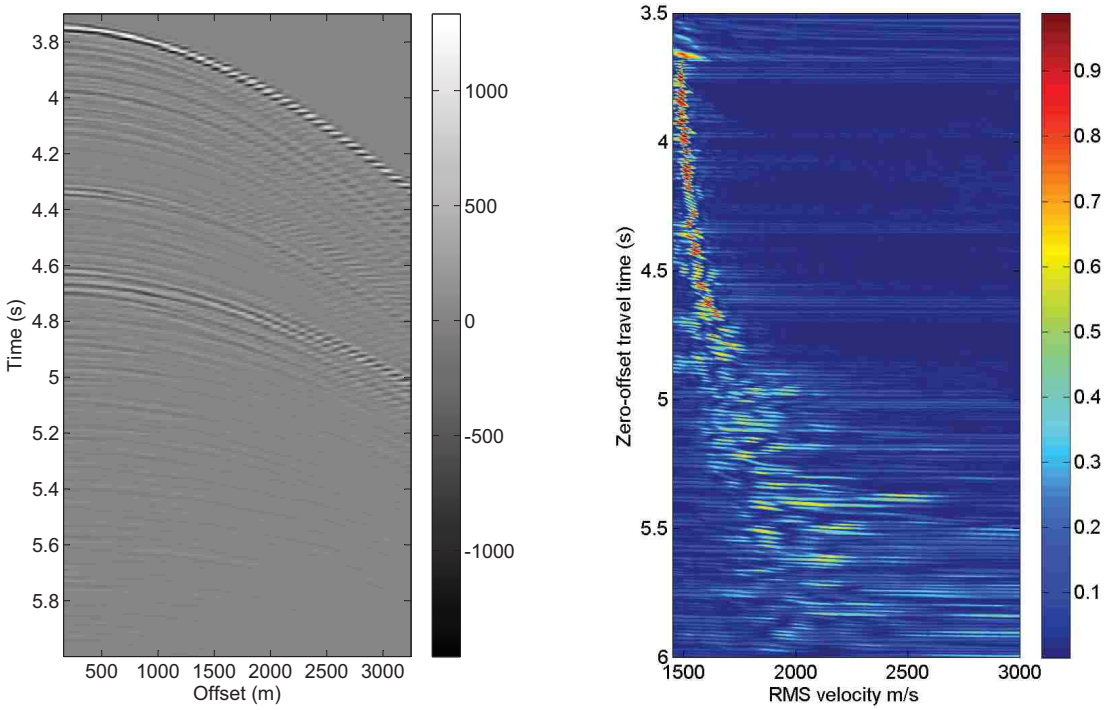

Figure 7.3: Representation of a gather of the S280-501 profile at CMP 1240 on the left where the colour bar represents the recorded amplitudes and its corresponding semblance plot on the right calculated with grid spacing of 2 milliseconds and 0.1 metres per second, here the colour bar represents the semblance at each point of the regular grid.

as in Figure 7.2 or well logs. If a seismic profile was used to update priors on the location of the horizons, then the gathers used to create this profile should not be used for further analysis; however, if that is not possible, use the selected horizons to limit the start point region with a MCMC algorithm or a gradient search algorithm to produce posterior estimates from vague priors. For this example, we started with a subset of CMP gathers sampled every 125 metres (roughly 7% of the original dataset) to create a coarsely laterally sampled seismic profile which was used to extract information about the horizons; these were then used to compose priors and start points as explained in the next Section.

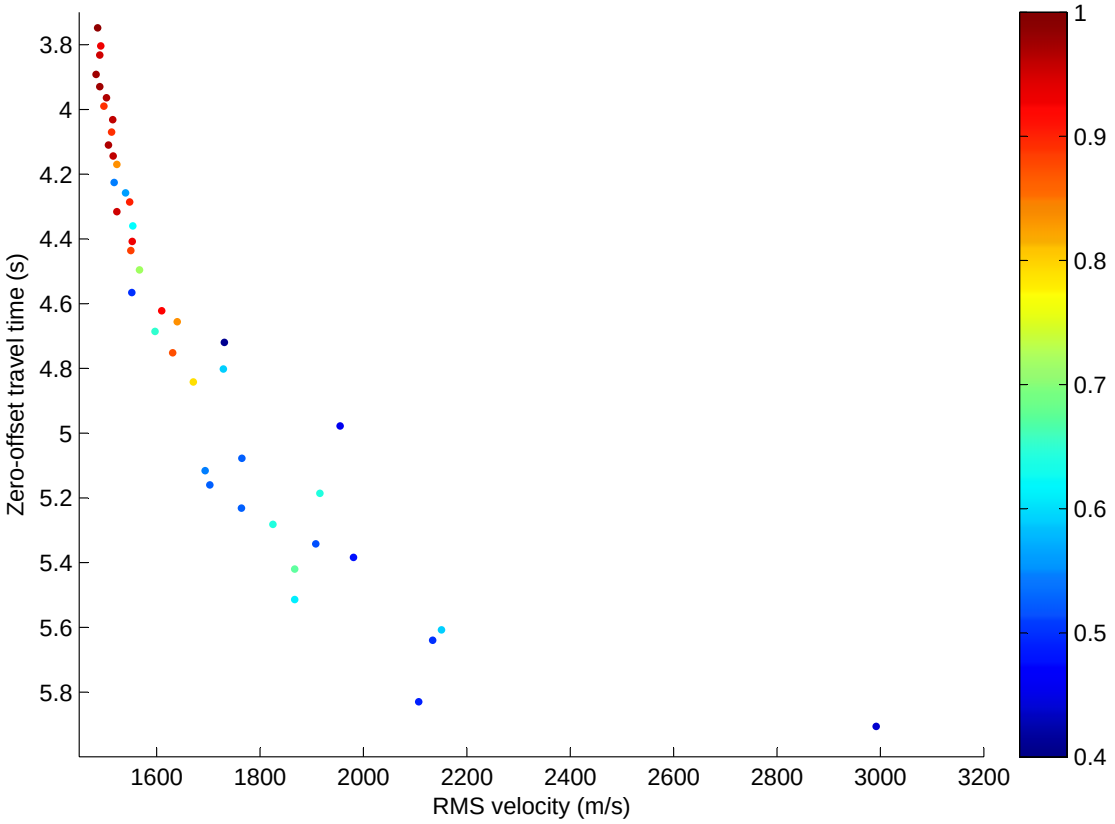

Figure 7.4: Plot of the estimates for the 61 candidates selected by the SET algorithm for the CMP gather 1240 represented in Figure 7.3, the colour bar refers to the semblance range.

7.2 Updating priors

Seismic profiles contain large amounts of information about the zero-offset travel time estimates and consequently about the number of layers around a common midpoint. In order to produce priors from a seismic profile, experts could manually pick all the horizons that they find adequate and then use these as an input for the SET algorithm; after the SET algorithm locates the sequences of events associated to these horizons, a BRAINS method can be applied according to the type and level of information desired. In the early days of BRAINS, we opted for this solution and we used the horizons published in Maloney et al. (2011) picked by Dr. Maloney. The horizons were tracked from the point of view of a geologist who seeks to tie the data to key geological boundaries; however, these picks are not the ones that best

resolve the seismic profile and the associated velocity model is not "smooth enough" for depth migration because layers may be too thin to produce a stable conversion from RMS to interval velocities.

In a second attempt to find an acceptable velocity model for this profile, we built an interactive graphical user interface which allows a trained user to pick horizons using a seismic profile and control priors as the velocity model is updated locally. So the user inputs a zero-offset travel time prior with the aid of a seismic profile and the underlying 1D BRAINS automatically updates the posterior. Since the expert or the trained user might not have any explicit information about the subsurface, double dipping (i.e. using the data more than once) is common place in seismic data analysis; while this is understable, it is not always statistically acceptable and it usually leads to over-confident results. As we demonstrated in the previous Section, there is a large number of highly probable and plausible models for each dataset and exploring all possible solutions is counterproductive; therefore, the introduction of horizon and velocity picking in prior models simplifies and speeds up the estimation process.

Rachel Chester, a second year undergraduate student at the Department of Earth Sciences at Durham University, was trained to use the toolbox written for this project and interactively update her prior beliefs using a seismic profile created for this dataset. Our trained user had full control over prior updates for both zero-offset travel times and velocities, both RMS and interval. The underlying BRAINS used was a local IS-BRAINS that spans 10 gathers for the prior Gaussian process update. When the first model was completed, one step of time migration was applied to attenuate the interference of diffracting rays and better resolve the seismic image. The seismic profile was then updated and a new set of gathers was used for the second modeling stage where the priors were once more updated. The results from this stage are displayed in Figures 7.6, 7.7 and 7.8. In Figure 7.6, we have a seismic profile with the zero-offset travel time posterior means which are fairly close to the primary horizon picks; the error in these estimates is negligible since they are mostly below the sampling interval. The fact that the error in zero-offset travel times are so small is not a result of overfitting but an effect of the nature of the hyperbolic

trend; there is almost no error linked to the location of the curves itself but mostly to their curvature so, it is expected to see small errors linked to travel times and more realistic variations in velocities and depths.

In Figure 7.7, we have the posterior RMS velocity means. In this map, we chose to represent the RMS velocity as a constant in each layer using the posterior mean at the top boundary so that the internal lateral variation was visible. Note that the RMS velocities vary smoothly across layers and increase with time as we would expect since the RMS field is expected to be smooth a priori. In Figure 7.8, we have the posterior interval velocity means. The map is consistent and plausible but shows a significant amount of lateral variation. The posteriors from this second modeling step were then used to run a depth migration resulting on the depth profile in Figure 7.9. If we jointly analyse Figures 7.8 and 7.9, we can see that the instability in the interval velocity map is directly related to the quality of the depth profile; e.g., in the CMPs between 10^4 and 1.5×10^4, the interval velocities are not as smooth as we would expect because of the presence of pinching layers and, as a result, we see a rough ripple-like patch on the depth model at around 4000 metres for these same CMPs, this pattern is unlikely to be correct and is most certainly a result of the instability of the interval velocity model. In order to minimize such effects, we can impose stricter smoothing criteria to both RMS and interval velocities; this is achieved by using the multi-gathered IS-BRAINS.

In Figure 7.10, we have the RMS velocity map obtained using IS-BRAINS, here we plotted the RMS velocity gradient to show how these velocities vary smoothly in both directions; in Figure 7.11, we have the interval velocity map and, as we can see, it is smoother than the map in Figure 7.8. In Figure 7.12, we have a scatter of the posterior zero-offset travel times where each point is coloured according to the posterior RMS velocity mean and plotted on top of its associated 95% credibility interval (in gray) using the vertical scale shown. It would be reasonable to expect that the error in the RMS velocities would increase with time or depth but it is clear from this plot that this is not always the case; because the layers are being modelled jointly, they all carry information about the overall model uncertainty and the errors are not simply additive. On the upper layers, the error estimates are

mostly associated to the hyperbolic misfit while in the deeper layers the errors are mostly explained by noise. In fact, if the signal in deeper layers were clearer, we would probably see a decrease in error with depth; in this example, a number of gathers clearly have smaller error bars in deeper layers than in their first layer. In Figure 7.13, we have a scatter of the posterior depth means where each point is coloured according to the posterior interval velocity mean and plotted on top of its associated 95% credibility interval (in gray) in a 125 metres to 500 metres per second scale. We can see that the credibility intervals are wider when layers are closer together or pinching; that is expected since in these cases the time difference is small and more affected by signal interference.

When analysing Figure 7.9, we pointed out that the instability in the CMPs between 10^4 and 1.5×10^4 is highly correlated to the instability in interval velocities in that same area, and now, the same can be seen in Figure 7.13; the wider error bars in that area reflect this same instability which can be improved with a different choice of prior or by adding more gathers to the dataset. Finally we look at the depth estimates and their associated 95% credibility intervals in Figure 7.14; the points are coloured according to the depth posterior standard deviations in metres and the gray bars indicate their corresponding 95% credibility intervals in metres in the same scale as the y axis. Overall the horizons are smooth and the errors are small and consistent laterally.

In summary, the maximum posterior depth standard deviation was of 42.93 metres and the average of these standard deviations across the profile was 9.33 metres; in comparison to the posterior mean depths, the standard deviations were in average 0.3% of the estimated mean and at most 0.4%, the highest relative variability can be found at the top layers indicating a misfit of the hyperbolic moveout or low signal to noise ratio. For the RMS estimates, the average posterior standard deviation is of 4.22 metres per second and the maximum is 36.81 metres per second, unlike one would expect, the distribution of the RMS velocities over depth is nearly flat with a few outliers exceeding 15 metres per second at depths of around 1000 metres which correspond to the rightmost part of the profile; relative to the posterior means for RMS velocities, the corresponding standard deviations are in average 0.25% of their

means and at most 1.7%.

Finally, in the interval velocity model, there is a considerably higher variation; the posterior standard deviations are 58.75 metres per second in average and at most 317.75 metres per second which correspond to average and maximum ratios relative to the posterior means of 2.4% and 23.7%, respectively. The relationship between posterior mean interval velocity and posterior standard deviations for interval velocities is not linear, it is fan shaped meaning that, even though the highest mean estimates correspond to the highest standard deviations, a high density of low standard deviations are still associated to high means. Interestingly, when looking at the distribution of standard deviations for interval velocities jointly with posterior depth means, we have that the highest standard deviations are at the layers closer to the surface at around 1000 metres corresponding to the rightmost area of the seismic profile; for depths over 1000 metres, there is no evident trend and variability is high. The result of this model is acceptable but could still be improved by including more iterations of migration and lateral smoothing of the velocity model especially in deeper layers. Potentially some of this lateral mobility is caused by miss-picking the event times which, with improved imaging, could be resolved.

Returning to Figure 7.1, we can see that the borehole DSDP-258 is on the line that we are investigating (Davies et al., 1974) and from the gathers we analysed in the second phase, the CMP gather 1240 is the closest to this borehole. A total of 525 metres of sediment were drilled and cored from borehole 258 from which five lithostratigraphic units were recognized and their depths and thicknesses were recorded. These five units are represented in Figure 7.5 near to the posterior mean depths and posterior 95% credibility intervals for the first four picked horizons. The first horizon correspnds to the seabed which is the top of the first unit, the second horizon corresponds to the lower boundary of unit 1. The observed thickness for unit 1 was of 114 metres and our posterior mean estimate was 119 metres. Horizons corresponding to the boundaries of subunit 2b and unit 3 were not picked. Both boundaries for unit 5 were picked; it was observed that these boundaries are located 514 and 525 metres below sea floor resulting in an observed thickness of 11 metres while the estimates related to the third and fourth horizons, which we believe

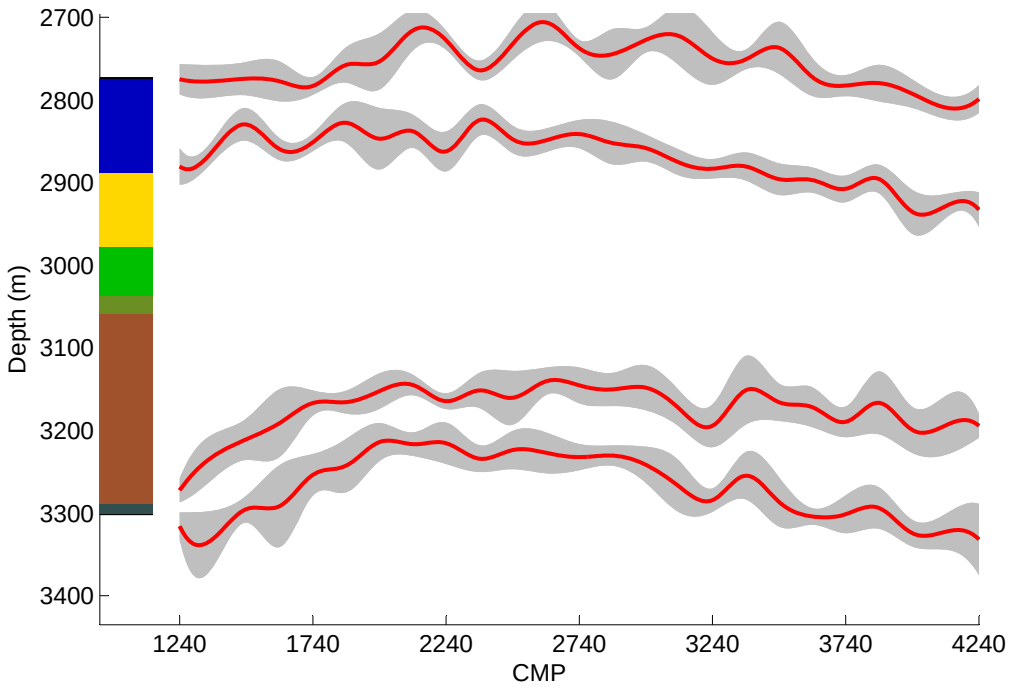

Figure 7.5: The posterior depth mean trends for the first 4 layers are represented in red with their corresponding 95% credibility patches in gray. The colour patch shows the units recovered in the borehole DSDP-258, the blue patch corresponds to Unit 1, the yellow and light green patches correspond to subunits 2a and 2b respectively, and the darker green, brown and gray patches correspond to Units 3 to 5.

correspond to these boundaries, were of 497.39 metres and 540.86 metres with an estimated thickness of 43.47 metres. However, the observed values are within all three credibility intervals as we can see in Figure 7.5. While this is not sufficient to verify that our model is correct, it shows that the depth model is consistent with the borehole.

Here we have seen an application of our methods to a real data example showing that it is a good alternative to traditional methods of seismic data processing with the benefits of prior updating and interactive uncertainty assessments of the current fit. Even though it is based in a number of well-defined physical processes, seismic data analysis is subjective and closely depends on the knowledge of the expert handling the data; for this reason, Bayesian methods provide us with a better

strategy to "simulate" the expert. In the next Chapter, we assess the impact of the methodology and models discussed so far.

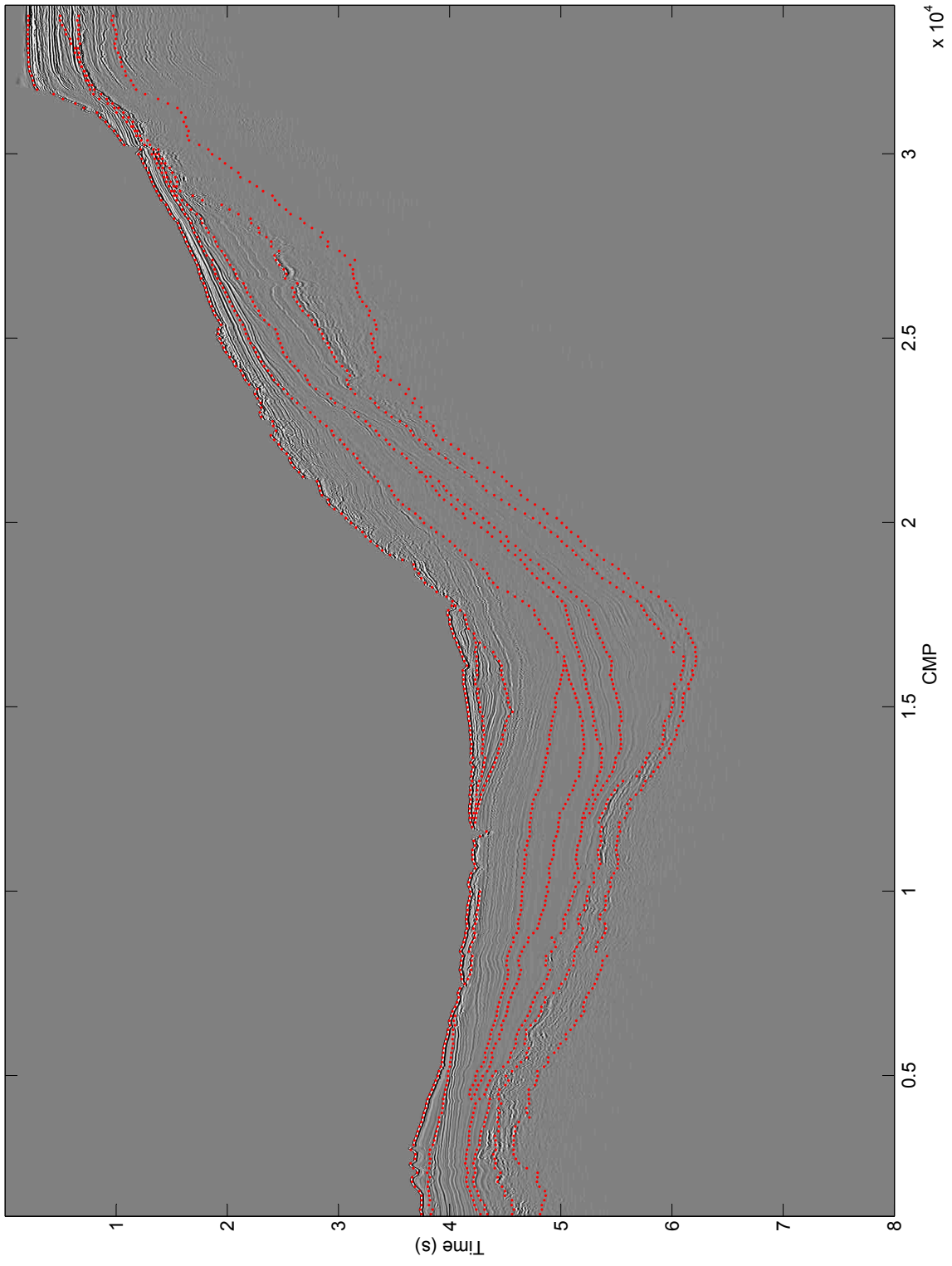

Figure 7.6: Seismic profile for S280-05 used in the second interactive step and, in red, zero-offset travel time posterior means for the picked horizons.

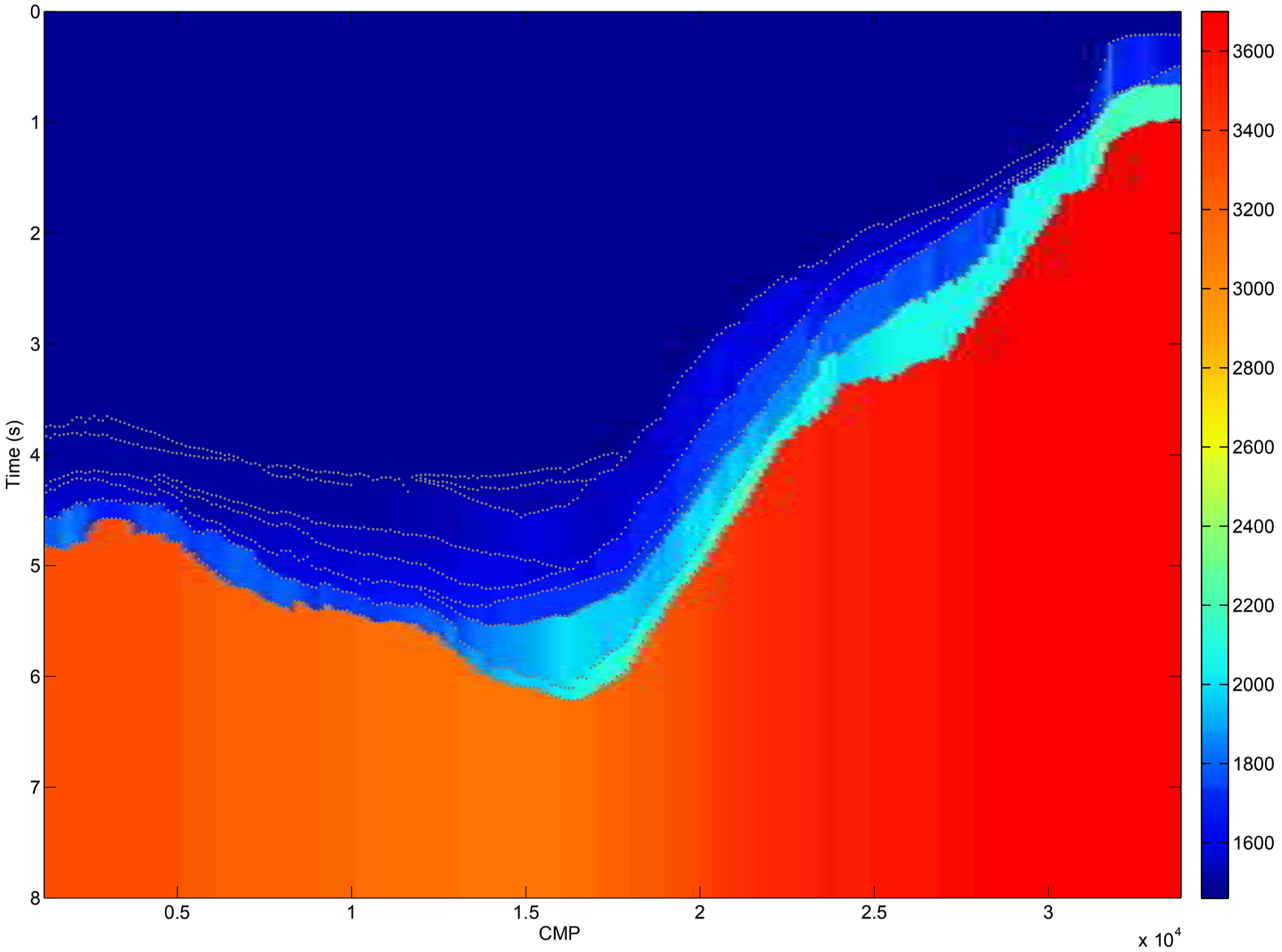

Figure 7.7: Posterior means for the RMS velocities in metres per second corresponding to the horizons in Figure 7.6 represented in a colour map according to the colour bar on the right. Here the posterior means in the boundaries are plotted as constants within layers to better show the lateral and vertical variability of these variables.

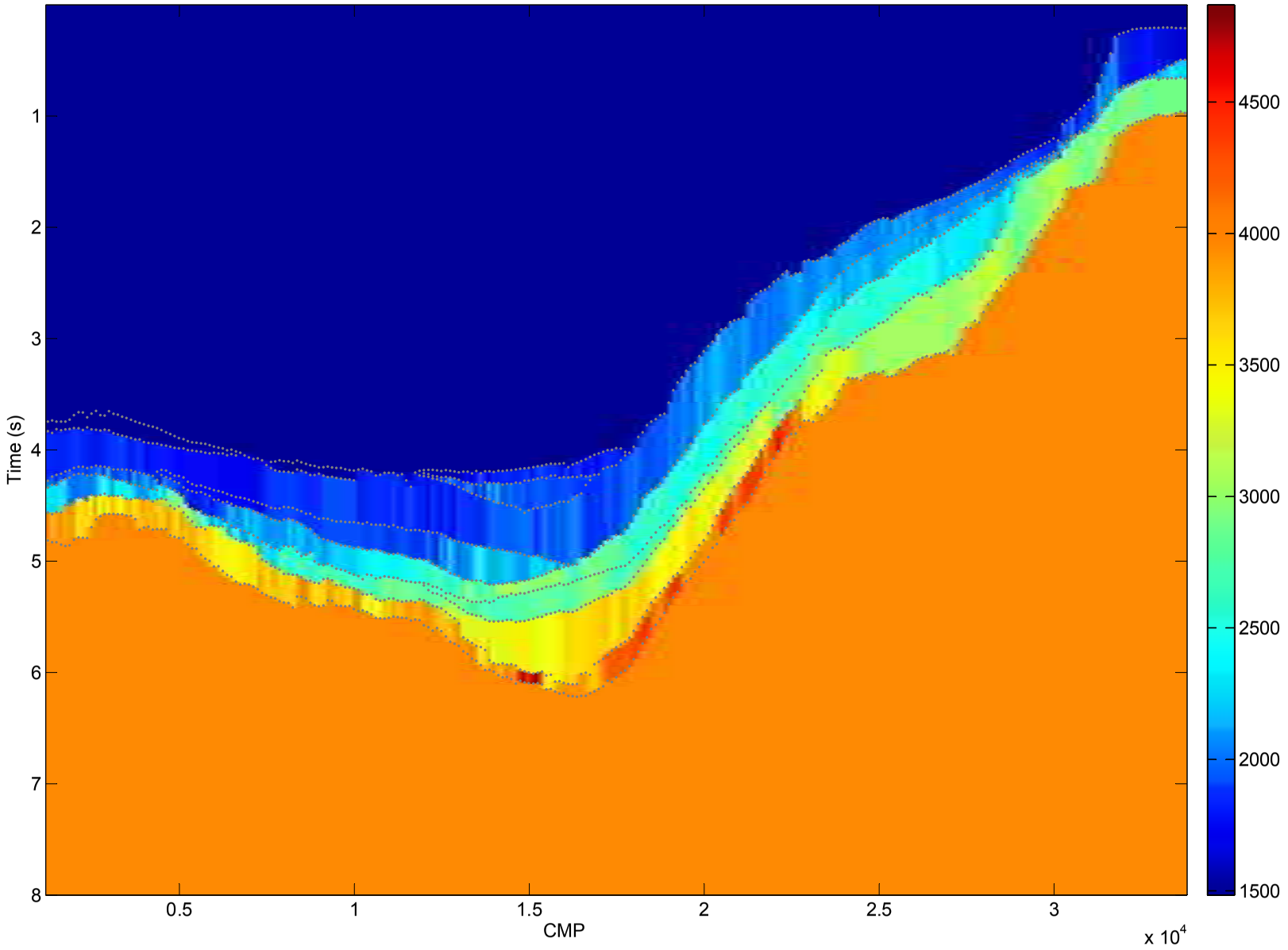

Figure 7.8: Posterior means for the interval velocities in metres per second corresponding to the estimated horizons in Figure 7.6 represented in a colour map according to the colour bar on the right.

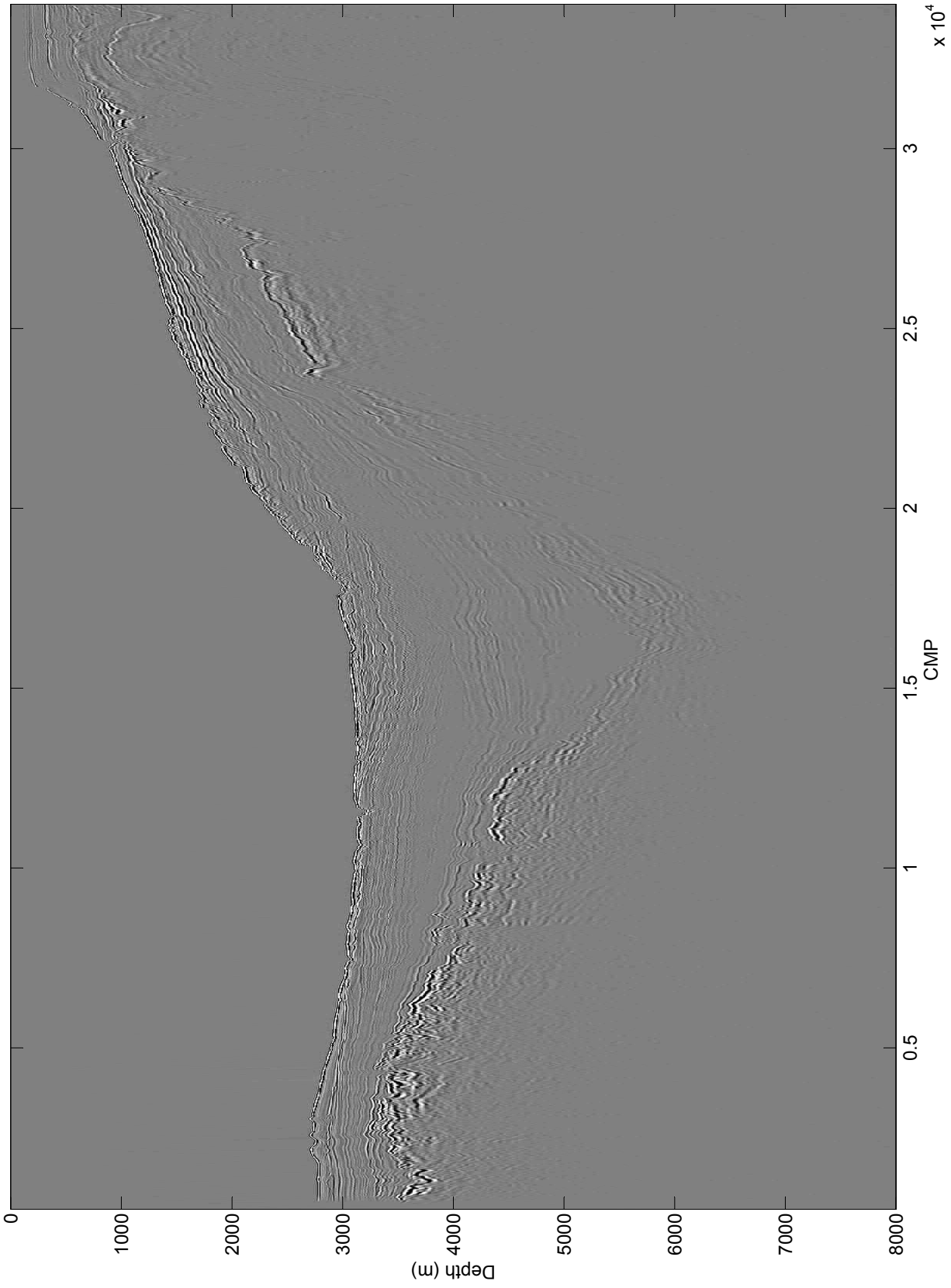

Figure 7.9: Depth migration profile obtained using the estimates in Figures 7.6 and 7.7.

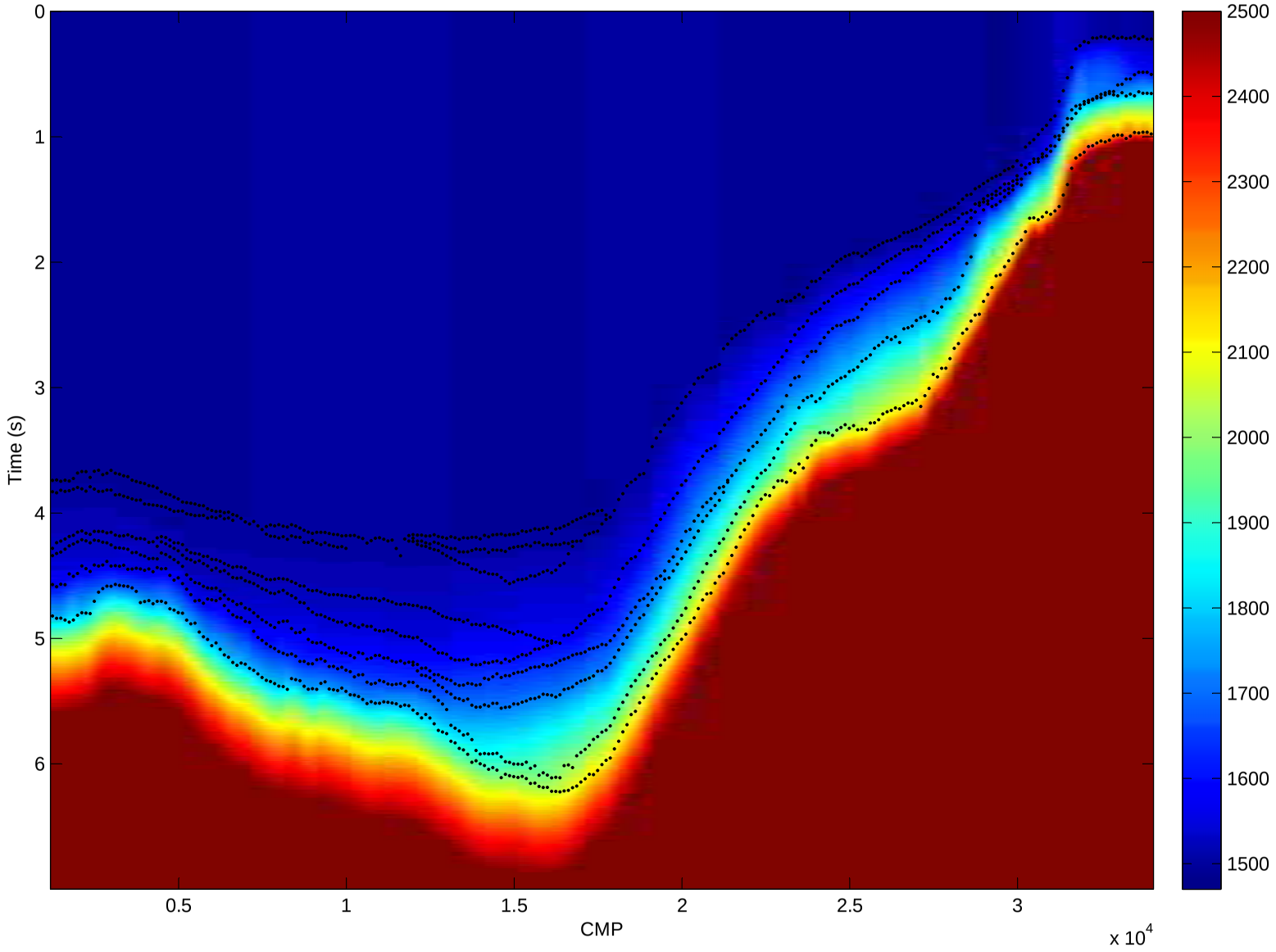

Figure 7.10: Colour map that shows the posterior means for the RMS velocities in metres per second obtained in the final model using the IS-BRAINS according to the colour bar on the right, the scatter of black dots represent the posterior mean for the zero-offset travel times.

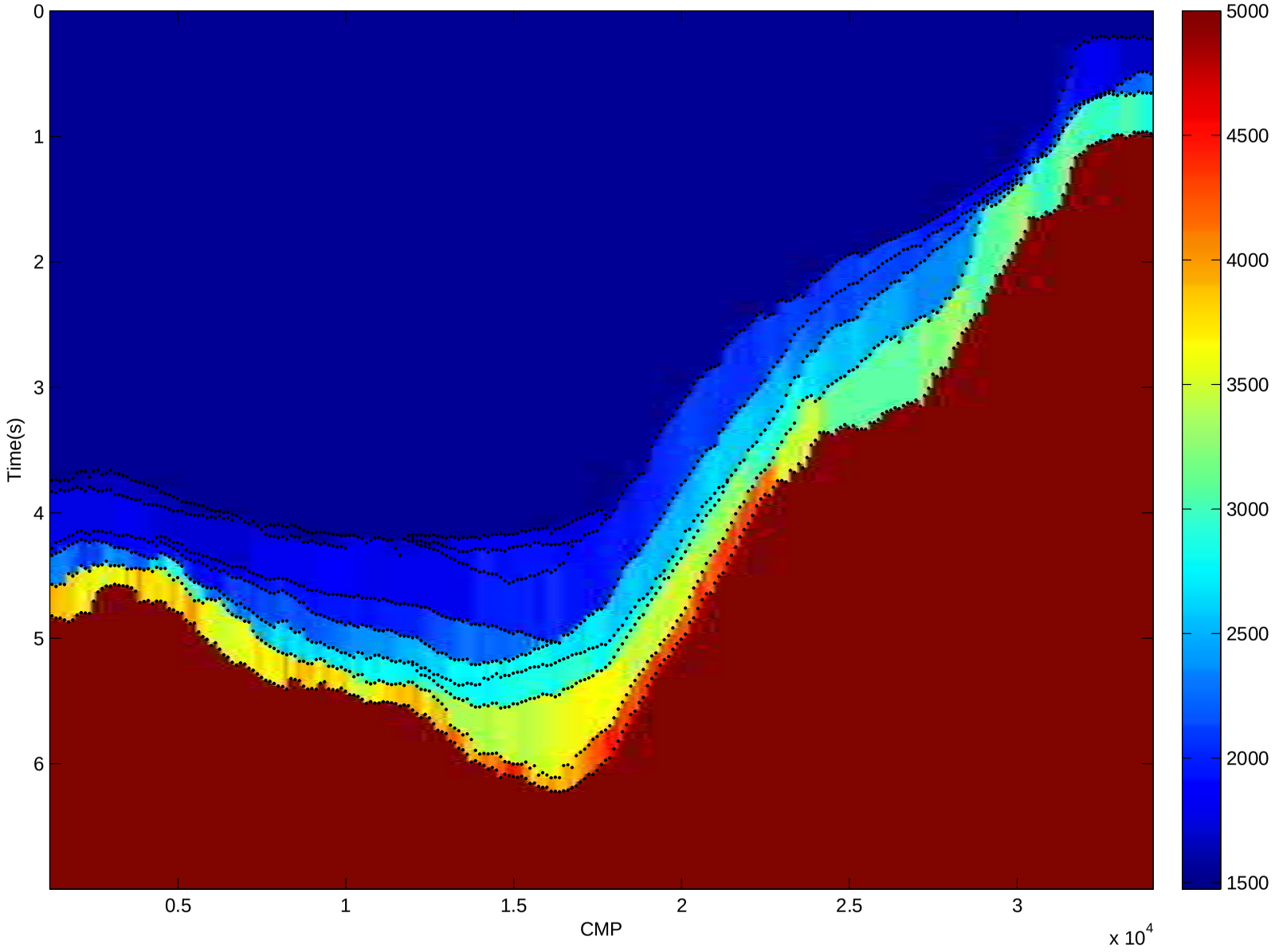

Figure 7.11: Colour map representing the posterior means for the interval velocities in metres per second obtained in the final model using the IS-BRAINS plotted using the colour bar on the right, the scatter of black dots represent the posterior mean for the zero-offset travel times.

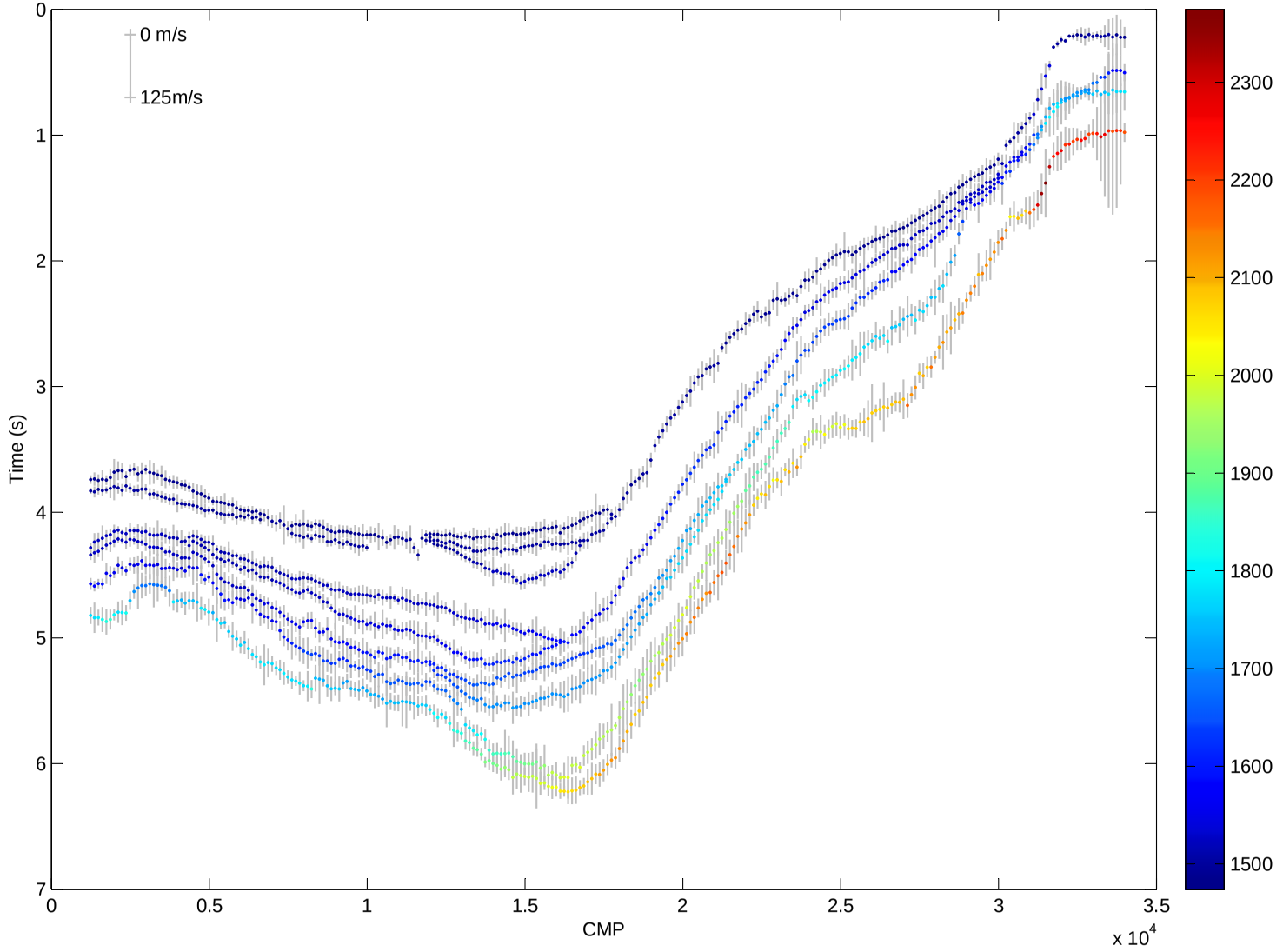

Figure 7.12: Posterior means and credibility intervals for RMS velocities (m/s). The position of the scattered points represent the posterior mean for the zero-offset travel times, these points are coloured according to the posterior mean for their corresponding RMS velocities (see colour bar on the right) and they are placed on top of their corresponding posterior 95% credibility intervals for the RMS velocity estimate. Each second in the y axis corresponds to 250 metres per second in the gray bars as indicated by the scale bar on the top left.

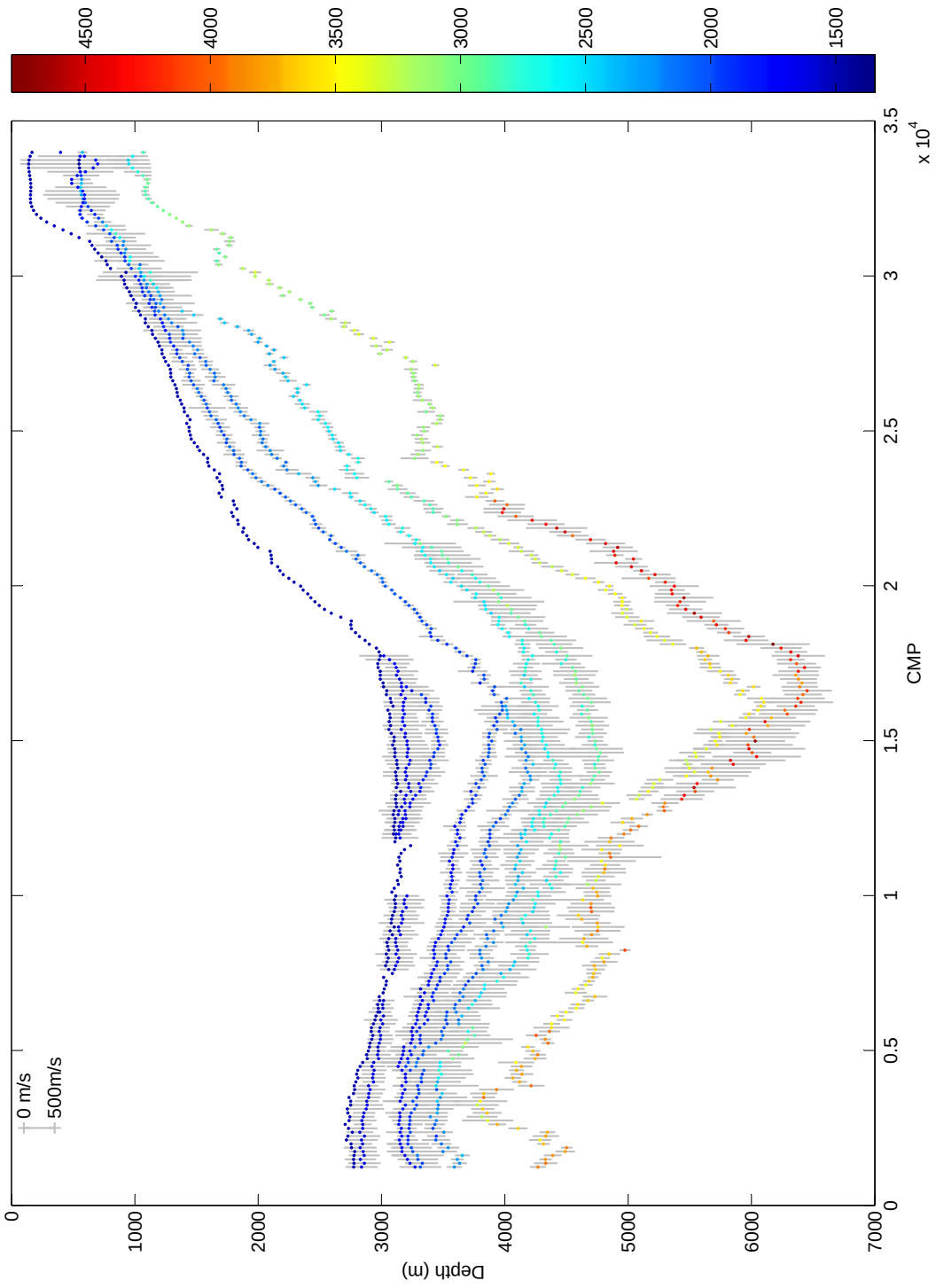

Figure 7.13: Posterior means and credibility intervals for interval velocities (m/s). The position of the scattered points represent the depth posterior means, these points are coloured according to the posterior mean for their corresponding interval velocities (see colour bar on the right) and they are placed on top of their corresponding posterior 95% credibility interval for the interval velocity estimate. Each metre in the y axis corresponds to 2 metres per second in the gray bars as indicated by the scale bar on the top left.

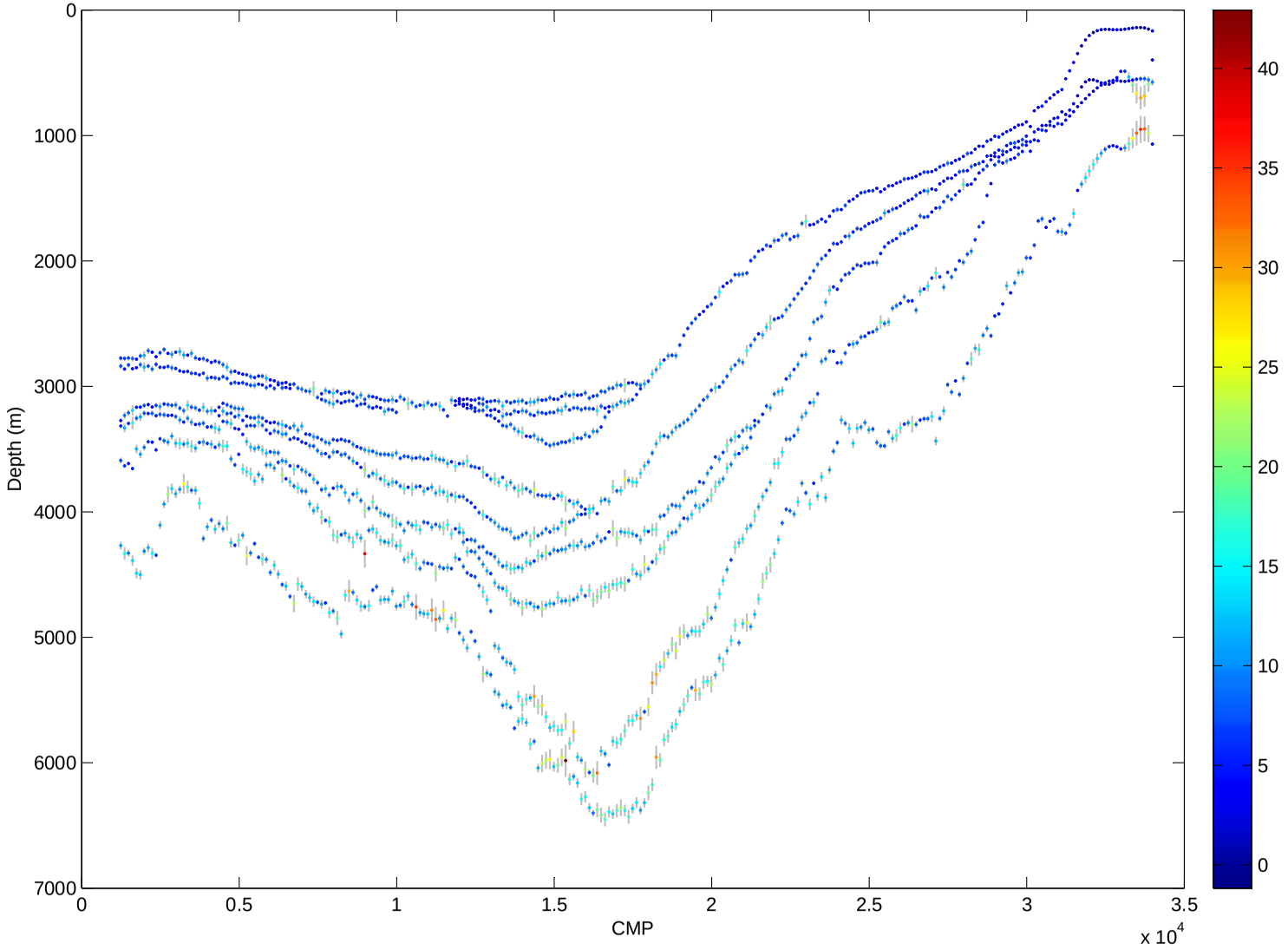

Figure 7.14: Posterior statistics for depth in metres. The position of the scattered points represent the depth posterior means, their colour represent the depth posterior standard deviations according to the coloubar on the right and the gray bars represent the 95% posterior credibility intervals for the depth estimates. All values are measured in metres and conform to the scale in the y axis.

Chapter 8

Why BRAINS?

8.1 Did BRAINS work?

Exploration geophysics is a field that is prone to constantly grow since the need to gather information about the planet where we live is limitless. The development of machines capable of processing large amounts of data in the past two decades was seen as an opportunity to gather as much data as possible; however, the methods for seismic data processing did not follow the growth in seismic data collection resulting in poor or close to no statistical analysis of the methods used to interpret these datasets. Nowadays, there is a growing group of geophysicists that understand the need to explore uncertainties and that quantity of data does not imply quality of results.

With the aim to aid these scientists, we developed the BRAINS class of models which uses the principles of ray geometry to approximate velocity models based in an ideal world. Each BRAINS model was designed to be used at a different stage of seismic data processing according to the expert's requirements and each interprets the expert's prior beliefs in a different way. In Chapters 5 to 7, we showed the BRAINS models' weaknesses and strengths and how not only the model but also the data carry a considerable amount of ambiguity. In order to assess part of this ambiguity, we proposed the index of flatness which can be used as an alternative to or in conjunction with the classic semblance analysis in order to express prior information or as a model selection tool to address the uncertainty in the number

of layers in the model.

The results obtained in the final chapters show that BRAINS did work, and that combined with other simulators like migration algorithms, it is capable of resolving seismic information to an acceptable standard. Since the expert has full control of prior inputs, the update of beliefs is easy and smooth; in this way, the expert is capable of performing an in-situ sensitivity analysis to investigate the stability and robustness of the model predictions.

8.2 Future Research

The objective of this work was to show that uncertainty analysis is compatible with seismic data exploration and that a trained user or expert can input their prior beliefs to update velocity models. While the results presented are satisfactory, the BRAINS world is still too idealized and migration steps, which are computationally expensive, are still necessary. We published a paper on the 1D SI-BRAINS method in Bayesian Analysis (Caiado et al., 2011) in order to introduce this approach to velocity modeling to the community. The next step is to further process the S280-5 line and submit our findings to a geophysical journal; finally, a second paper about the generalized BRAINS models will be submitted afterwards.

Issues that we still have to address and other points that we plan to develop are listed below:

1. Anisotropy. There are a number of models available to estimate anisotropy parameters but they are limited and underdetermined as we pointed out in Chapter 7. By approximating the geodesics in the Finsler space, we aim to develop a new class of BRAINS models to resolve anisotropy and inhomogeneity.

2. Migration. Time and depth migration are computationally expensive and rely on point estimates in a velocity model; even if used only on prior update steps, we still lose part of the information about uncertainty that we had prior to the migration step. Either by emulating migration algorithms or developing

a faster simulator, we can increase the robustness of the BRAINS class and reduce the processing time required to analyse a seismic dataset.

3. Layer Selection. The stage of layer selection in real data processing still relies on prior information about horizons. When following a horizon in a seismic profile, an expert tend to follow sequences of local maxima and minima; by generalizing the SET algorithm, we can partially reproduce this task and develop a horizon tracking algorithm.

There are a number of other generalizations that we can introduce to make BRAINS a more flexible class of models but these are the most relevant. Regarding real and synthetic data analysis, we plan to expand and test BRAINS in 3D datasets and 2D datasets with multiple lines. We also need to explore the level of resolution that we can achieve without being overconfident about the model and ultimately generate geological profiles based on velocity models.

Notation

A_{jk} Recorded amplitude at the k-th sample of trace j. 27

b_i Lower boundary of the i-th layer. 19

$b_{(i-1)}$ Upper boundary of the i-th layer. 19

Δt_{0_i} Portion of the total zero-offset two-way travel time spent in the i-th layer. 19

Δt_i Portion of the total two-way travel time spent in the i-th layer. 19

$\Delta T_{0_i}(x_c)$ Portion of the zero-offset two-way travel time spent in the i-th layer in the gather with CMP position x_c (random variable). 39

$\Delta V_{rms_i}(x_c)$ RMS velocity increment referent to the i-th layer in the gather with CMP position x_c (random variable). 39

Δx_i Horizontal distance between the source or point of refraction at boundary $b_{(i-1)}$ and the point of reflection at boundary b_i. 20

$\Delta Z_{0_i}(x_c)$ Thickness of the i-th layer in a gather with CMP position x_c (random variable). 41

Δz_i Thickness of the i-th layer. 19

i Layer index. 19

M Common midpoint. 18

m Number of traces in a gather. 23

m_s Number of sampling points in a trace. 27

n Number of layers. 19

R Receiver. 18

S Source. 18

t Travel time. 18

t_0 Zero-offset two-way travel time. Time that a ray takes to travel from a source through the media and back to a receiver in the same position as the source assuming that the ray travelled perpendicularly to the source-receiver pair through all the path. 18

θ_i Angle of refraction. Angle formed by a ray refracted from boundary $b_{(i-1)}$ with the ray's path and the vector orthogonal to the boundary cointaned in the same plane as the ray's path. 20

$\mathcal{T}_i^{(r)}$ Curve of recorded travel-times corresponding to the i-th layer. 37

$T_{0_i}(x_c)$ Zero-offset two-way travel time for the i-th layer in the gather with CMP position x_c (random variable). 39

$T_{jk}^{(o)}$ Recorded time at the k-th sample of trace j. 27

$T_i^{(r)}(x, x_c)$ Recorded two-way traveltime of a reflected ray from layer i fired by a source S and recorded by a receiver R separated by an offset x and with CMP position x_c. 39

T_{sp} Sampling time. 27

T_{tot} Total recording time of a signal. 27

v Interval velocity. Velocity with which a ray travel through an isotropic homogeneous medium. 18

v_i Interval velocity of the i-th layer. 19, 22

$V_{int_i}(x_c)$ Interval velocity of the i-th layer in a gather with CMP position x_c (random variable). 41

$V_{rms_i}(x_c)$ RMS velocity of the i-th layer in the gather with CMP position x_c (random variable). 39

x Offset. Distance that separates a source-receiver pair. 20

Bibliography

T. Alkhalifah and I. Tsvankin. Velocity analysis for transversely isotropic media. *Geophysics*, 60:1550–1566, 1995.

B. B. Baker and E. T. Copson. *Mathematical Theory of Huygen's Principle*. Oxford University Press, 2nd edition, 1950.

B. A. Berg. *Markov Chain Monte Carlo Simulations and Their Statistical Analysis*. World Scientific, 2004.

J. E. Besag. Towards bayesian image analysis. *Journal of Applied Statistics*, 16: 395–407, 1989.

I. Borissova. Geological framework of the naturaliste plateau. *Geoscience Australia Record*, 20, 2002.

I. Borissova, B. Bradshaw, C. Nicholson, D. Payne, and H. Struckmeyer. First acreage release in frontier mentelle basin: Significant potential to become new petroleum province. *Geoscience Australia Record*, 98, 2010.

A. Bornemann, R. D. Norris, O. Friedrich, S. Schouten B. Beckmann, J. S. S. Damste, J. Vogel, P. Hofmann, and T. Wagner. Isotropic evidence for glaciation during the cretaceous supergreenhouse. *Science*, 319:189–192, 2008.

E. O. Brigham. *The Fast Fourier Transform*. Prentice-Hall, 2002.

C. C. S. Caiado, R. W. Hobbs, and M. Goldstein. Bayesian strategies to assess uncertainty in velocity models. *Bayesian Analysis*, 6(3):1–28, 2011.

B. P. Carlin and T. A. Louis. *Bayes and Empirical Bayes methods for data analysis*. CRC Press, 2000.

V. Cerveny. *Seismic Ray Theory*. Cambridge University Press, 2001.

T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 2001.

P. S. Craig, M. Goldstein, J. C. Rougier, and A. H. Scheult. Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association*, 96(454), 2001.

T. A. Davies, B. P. Luyendyk, K. S. Rodolfo, D. R. C. Kempe, B. C. McKelvey, R. D. Leidy, G. J. Horvath, R. D. Hyndman, H. R. Thierstein, R. C. Herb, E. Boltovskoy, and P. Doyle. Site 258. In *Initial Reports of the Deep Sea Drilling Project*, volume 26, pages 359–414. Washington (U.S. Government Printing Office), 1974.

P. Deuflhard. *Newton methods for nonlinear problems: affine invariance and adaptive algorithms*. Springer, 2004.

C. H. Dix. *Seismic Prospection for Oil*. Harper & Brothers, 1952.

C. H. Dix. Seismic velocities from surface measurements. *Geophysics*, 20:68–86, 1955.

M. Fieldman. Non-linear system vibration analysis using hilbert transform - i. free vibration analysis method freevib. *Mechanical Systems and Signal Processing*, 8 (2):119–127, 1994.

F.H. Fisher and V.P. Simmons. Sound absorption in sea water. *Journal of the Acoustic Society of America*, 62:558–564, 1977.

D. J. Foster and C. C. Mosher. Suppression of multiple reflections using the radon transform. *Geophysics*, 57(3):386–395, 1992.

A. E. Gelfand and D. K. Dey. Bayesian model choice; aymptotics and exact calculations. *Journal of the Royal Statistical Society Series B*, 56:501–514, 1994.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6:721–741, 1984.

Schlumberger Oilfield Glossary. Schlumberger oilfield glossary, 2010. URL http://www.glossary.oilfield.slb.com.

D. Gottlieb and S. A. Orszag. *Numerical analysis of spectral methods: theory and applications*. SIAM, 1993.

P. J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82:711–732, 1995.

S. L. Hahn. *Hilbert transforms in signal processing*. Artech House, 1996.

D. F. Halliday, A. Curtis, J. O. A. Robertsson, and D. J. Manen. Interferometric surface-wave isolation and removal. *Geophysics*, 72(5):A69–A73, 2007.

W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970.

Christiaan Huygens. *Traité de la lumiere*. Pieter van der Aa, 1690.

N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*, volume 1. Wiley Blackwell, 2nd edition, 1994.

I.F. Jones, R.I. Bloor, B.L. Biondi, and J.T. Etgen. *Prestack Depth Migration and Velocity Model Building*. Society of Exploration Geophysicists, 2008.

S. Karlin and H. Taylor. *A First Course in Stochastic Processes*. Academic Press, 2nd edition, 1975.

P. Kearey, M. Brooks, and I. Hill. *An Introduction to Geophysical Exploration*. Wiley-Blackwell, 2002.

J.B. Keller. Geometrical theory of diffraction. *Journal of the Optical Society of America*, 52(2):116–130, 1962.

Marcel M. M. Kuypers, Richard D. Pancost, and Jaap S. Sinninghe Damste. A large and abrupt fall in atmospheric CO2 concentration during Cretaceous times. *Nature*, 299:342–345, 1999.

T. Lay and T. C. Wallace. *Modern global seismology*. Academic Press, 1995.

D. Maloney, C. Sargent, N. G. Direen, R. W. Hobbs, and D. R. Grocke. Re-evaluation of the mentelle basin, a polyphase rifted margin basin, offshre south-west australia: new insights from integrated regional seismic datasets. *Solid Earth*, 2:207–123, 2011.

K. L. Mengersen and R. L. Tweedie. Rates of convergence of the hastings and metropolis algorithms. *Annals of Statistics*, 24:101–121, 1996.

N. Metropolis and S. Ulam. The monte carlo method. *Journal of the Americal Statistical Association*, 44:335–341, 1949.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

D. A. B. Miller. Huygens's wave propagation principle corrected. *Optics Letters*, 16 (18):1370–1372, September 1991.

A. R. Mitchell and D. F. Griffiths. *The finite difference method in partial differential equations*. Wiley-Interscience, 1980.

R. L. Nowack. Wavefronts and solutions of the eikonal equation. *Geophysical Journal International*, 110(1):55–62, 2007.

C.B. Officer. *Introduction to the Theory of Sound Transmission*. McGraw-Hill, 1958.

A. O'Hagan. Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering and System Safety*, 91:1290–1300, 2006.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

G. O. Roberts and J. S. Rosenthal. Convergence of slice sampler markov chains. *Journal of the Royal Statistical Society B*, 61:643–660, 1999.

G. O. Roberts and J. S. Rosenthal. Coupling and ergodicity of adaptive mcmc. *Journal of Applied Probability*, 44:458–475, 2007.

G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk metropolis algorithms. *Annals of Applied Statistics*, 7:110–120, 1997.

H. Rund. *The differential geometry of Finsler spaces*. Springer, 1959.

C. Sargent, R. W. Hobbs, and D. R. Groecke. Improving the interpretability of air-gun seismic reflection data using deterministic filters: A case history from offshore cape leeuwin, southwest australia. *Geophysics*, 76(3):B71–B84, 2011.

R.E. Sheriff and L.P. Geldart. *Exploration Seismology*. Cambridge University Press, 1995.

J. Skilling. Nested sampling for general bayesian computation. *Journal of Bayesian Analysis*, 1:833–860, 2006.

A. F. M. Smith and G. O. Roberts. Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society A*, 159:309–321, 1993.

P. L. Stoffa, J. T. Fokkema, R. M. de Luna Freire, and W. P. Kessinger. Split-step fourier migration. *Geophysics*, 55:410–421, 1995.

L. Thomsen. Weak elastic anisotropy. *Geophysics*, 51(10):1954–1966, 1986.

L. Tierney. Markov chains for exploring posterior distributions. *Annals of Statistics*, 22:1701–1786, 1994.

I. Tsvankin and L. Thomsen. Nonhyperbolic reflection moveout in anisotropic media. *Geophysics*, 59(8):1290–1304, 1994.

M. van der Baan and J. M. Kendall. Estimating anisotropy parameters and traveltimes in the $\tau - p$ domain. *Geophysics*, 67(4):1076–1086, 2002.

E. Verschuur. *Seismic multiple removal techniques*. EAGE Publications, 2006.

J. C. White and R. W. Hobbs. Extension of forward modeling phase-screen code in isotropic and anisotropic media up to critical angle. *Geophysics*, 72(5):107–114, 2007.

G. B. Whitham. *Linear and Nonlinear Waves*. Wiley-Interscience, 1974.

R.-S. Wu. Wide-angle elastic wave one-way propagation in heterogenous media and an elastic wave complex-screen method. *Journal of Geophysical Research*, 99: 751–766, 1994.

R. S. Wu and K. Aki. Introduction: Seismic wave scattering in three-dimensionally heterogeneous earth. *Pure and Applied Geophysics*, 128(1-2):1–6, 1988.

T. Yajima and H. Nagahama. Finsler geometry of seismic ray path in anisotropic media. *Proc. of the Royal Society A*, 465:1763–1777, 2009.

O. Yilmaz. *Seismic Data Analysis*. Society of Exploration Geophysicists, 2002.

O. C. Zienkiewicz and R. L. Taylor. *The finite element method*. Butterworth-Heinemann, 2000.

Appendix A

Statistical topics

A.1 The Bayesian Method

In statistics, we are usually interested in analysing data by assuming that this data is subject to some form of random variation so, for the purpose of investigating this dataset, a statistical model is formulated and ultimately estimates or decisions are made based such model. In parametric statistics, the model is assumed to be described by a set of unknowns or variables. The process of analysing and drawing conclusions about these variables is called statistical inference.

In statistical inference, there are two main line of thoughts: the frequentist and the Bayesian. From the frequentist point of view, the parameters of a model are seen as unknown fixed values; given a population with an underlying distribution dependent on these parameters, a random sample is taken and used to estimate the population parameters. In the Bayesian framework, the parameters are treated as variables whose "a priori" distribution is formulated before data collection; a sample is then taken from the target population and the prior distribution is updated using Bayes' theorem resulting in a posterior distribution as described in Theorem A.1.1.

Theorem A.1.1. *Let $\pi(\theta)$ be a prior distribution and $f(\mathbf{x}|\theta)$ the likelihood function, i.e. the distribution of the sample $\mathbf{x}$ given a value of θ. Then we have that the posterior distribution, $\pi(\theta|\mathbf{x})$, is given by*

$$\pi(\theta|\mathbf{x}) = \frac{\pi(\theta)f(\mathbf{x}|\theta)}{\int_\Theta \pi(\theta)f(\mathbf{x}|\theta)d\theta} \tag{A.1.1}$$

where Θ is the parametric space.

The mean, median or mode of the posterior distribution are commonly used as point estimates for θ and each is associated to a different loss function, $L(\theta, \hat{\theta})$ that measures the cost of selecting a point estimate $\hat{\theta}$ for θ. The posterior mean is the estimate that minimizes the expected value of the quadratic loss function, the median minimizes the expected loss of the absolute deviation and the mode minimizes the indicator function.

We can also express information about the posterior distribution of θ via credible regions; the $\alpha\%$ credibility region for θ is a region for which the probability of θ falling in this region is $\alpha\%$. For a continuous θ, there are an infinite number of choices for this region and the interval selected depends on the method used, e.g. we can choose a credible interval for which the mean is the central point or we can choose to use symmetric posterior quantiles. The latter is the most commonly used method.

A.2 Statistical Distributions

In Chapter 3, we used Beta, Gamma and Normal random variables to describe our priors, posteriors and likelihoods. All three are classified as continuous random variables and each is characterized by a density function. We give a few details on these distributions next.

A.2.1 Normal Distribution

The multivariate Normal distribution of a n-dimensional random vector $(X_1, \ldots, X_n)$ is characterized by the following density function

$$f(\mathbf{X}) = (2\pi)^{-n/2} \left|\Sigma\right|^{-1/2} \exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right] \tag{A.2.2}$$

where $x \in \mathbb{R}^n$, $\mu \in \mathbb{R}^n$ is the mean vector and $\Sigma \in \mathbb{R}^{n\times n}$ is the covariance matrix of this distribution; a normal random vector is fully defined by its mean and covariance vectors.

The reason why the Normal distribution is the most common in applications is that it is the limiting distribution in the central limit theorem (CLT). The CLT

states that, given n m-dimensional random variates with a finite mean vector and positive semi-definite covariance matrix, then the sum of these n variates converges in distribution to the multivariate Normal.

A.2.2 Gamma Distribution

The Gamma distribution is an univariate continuous distribution defined on a positive support. For a Gamma variate X with a scale parameter θ and a shape parameter k both positive, we have that its probability density function is given by

$$f(x) = x^{k-1}\frac{\exp(-x/\theta)}{\Gamma(k)\theta^k},\ x \geq 0. \tag{A.2.3}$$

Unlike Normal variates, a Gamma variate is always positive making it suitable for modelling waiting times.

A.2.3 Beta Distribution

The Beta distribution is defined on the interval $(0, 1)$ and is parameterized by two shape parameters, α and β, both positive. The density function for a Beta variable X is given by

$$f(x) = \frac{x^{\alpha-1}\left(1-x\right)^{\beta-1}}{B(\alpha,\beta)} \tag{A.2.4}$$

where $B(\alpha, \beta)$ is the beta function.

This distribution is suitable for modeling events that are constrained to an interval like indicator or sign variables.