Bayesian Analysis (2012) **7**, Number 1, pp. 211–234

Bayesian Strategies to Assess Uncertainty in Velocity Models

Camila C. S. Caiado*, Richard W. Hobbs† and Michael Goldstein‡

Abstract. Quantifying uncertainty in models derived from observed seismic data is a major issue. In this research we examine the geological structure of the sub-surface using controlled source seismology which gives the data in time and the distance between the acoustic source and the receiver. Inversion tools exist to map these data into a depth model, but a full exploration of the uncertainty of the model is rarely done because robust strategies do not exist for large non-linear complex systems. There are two principal sources of uncertainty: the first comes from the input data which is noisy and band-limited; the second is from the model parameterisation and forward algorithm which approximate the physics to make the problem tractable. To address these issues we propose a Bayesian approach using the Metropolis-Hastings algorithm.

Keywords: Gaussian Processes, Metropolis-Hastings algorithm, Seismology, Velocity Modelling

1 Introduction

Seismic reflection surveying is the principal means of investigating the geological structure of the Earth to depths of about 30km. Acoustic energy from a source on the surface propagates into the Earth and is partially reflected due to a change of acoustic impedance between different layers of rock. The amplitude and phase of the reflected energy is dependent on the change of the elastic parameters at the interface, namely p-wave velocity, s-wave velocity and density (see Glossary after Section 8).

In the field, a series of acoustic sources are fired into the sub-surface at known points along the seismic profile to be investigated and each source is recorded by an array of receivers. Knowledge of the position of the source and receiver locations is used during processing to collect together recorded traces with the same geometrical mid-point between the source and receiver called a common midpoint (CMP) gather. A CMP gather consists of a number of traces with varying source-receiver offsets. In Figure 1, we can see a simple example of a CMP gather with only two traces: the recorded peak on the first trace of Figure 1a shows how long it took for the signal to travel from source S_1 to receiver R_1. Figure 2 is another example of a CMP gather, in this case with three traces. From the first trace we can see four distinct peaks each corresponding to the travel time of the signal from source S_1 to receiver R_1 after being partially reflected by four different layers with velocities v_i and depths z_i, $i = 1, \ldots, 4$.

*Department of Math. Sciences, Univ. of Durham, Durham, UK, c.c.d.s.caiado@durham.ac.uk
†Department of Earth Sciences, Univ. of Durham, Durham, UK, r.w.hobbs@durham.ac.uk
‡Department of Math. Sciences, Univ. of Durham, Durham, UK, michael.goldstein@durham.ac.uk

 DOI:10.1214/12-BA707

Once the distortion due to the variation of offsets is removed through a process called normal move-out (NMO) correction, the signals can be stacked together to enhance the required reflected signal from the background noise. For the NMO correction to be effective, it requires an estimate of the sound velocity structure of the sub-surface. In this paper we propose a Bayesian based method to address this problem.

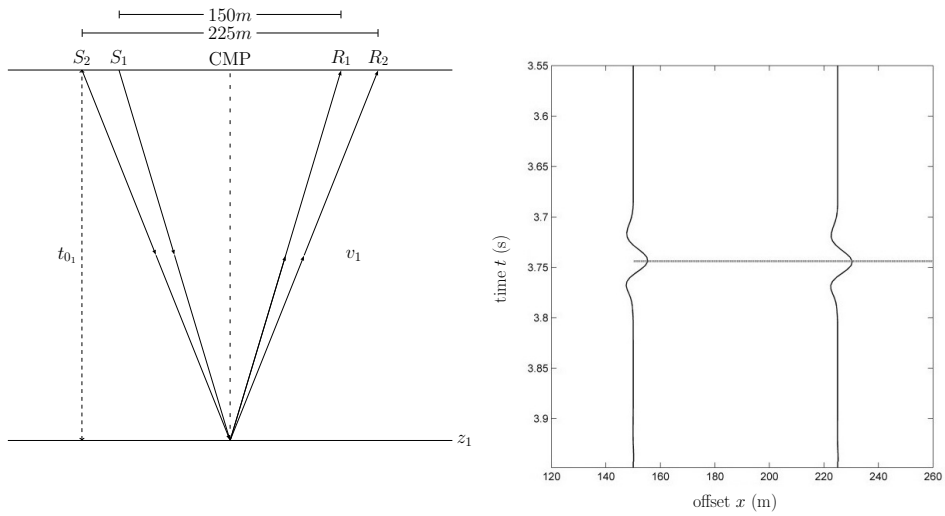
Unlike manual velocity picking by a trained operator, the Bayesian model provides us with the most probable velocity model and the uncertainty associated with it and each of its parameters. There is also the flexibility of adding prior knowledge about the structure to be analysed and control over most of the variables.

In Section 2, we introduce the preprocessed data. In order to understand how the data is sorted and filtered and the model structured, we introduce the tools used to run our analysis in Sections 3, 4 and 5. In Section 6, we explain how all these tools are combined to deliver a model for the velocity structure of the subsurface. Synthetic and real data sets are used in Section 7 to show the model efficiency compared to other methods. In Section 9, we have a small glossary of geophysics terms used here (see oil (2010)).

2 The Preprocessed Data

In order to create a velocity model, we need to understand the data and define variables. Each trace in the CMP consists of a sequence of points A_{kj} generated at source S_j and recorded at receiver R_j, which are separated by a distance x_j, at time $(\Delta(t) \cdot k)$ where $\Delta(t)$ is a fixed sampling rate.

The signal recorded is bipolar so each trace alternates between positive and negative values that correspond to the reflection strength or noise. When the signal reaches an interface with a significant change of acoustic impedance, the energy is partially reflected which is detected by the receiver and gives a peak on the data trace. Within a CMP gather, these interfaces form coherent peaks through the traces following an approximate hyperbolic curve that corresponds to a possible layer. Therefore, in order to analyse this data, these peaks need to be selected and sorted (e.g. picking the peaks on each trace of Figure 2b and matching them to their respective layer). In Section 6.1, we give an automatic method to pick and sort these points however, in the mean time we'll assume that they were already picked using the Bayesian picking method or any equivalent method, e.g., manual picking and matching of peaks in each trace. In Figure 3, we show the plot of a single trace and in Figure 4, we show the plot of a whole gather consisting of 80 such traces.



(a) (b)

Figure 1: (a) 1.5-D single-layer model (i.e., a model in which velocities are assumed to be functions of the depth only) showing the energy propagation path from the sources S_1 and S_2 (fired independently at different times to avoid mixed signals) to receivers R_1 and R_2 centred around a CMP through one layer with velocity v_1, depth z_1 and zero-offset travel time t_{0_1}. (b) The vertical traces represents the amplitude of the reflected energy arriving at $x_1 = 150m$ (distance between S_1 and R_1) and $x_2 = 225m$ (distance between S_2 and R_2) against time. The amplitude spikes on the traces correspond to the two-way travel time for the energy to travel from the sources S_1 and S_2 to the sub-surface reflectors and back to the receivers R_1 and R_2.

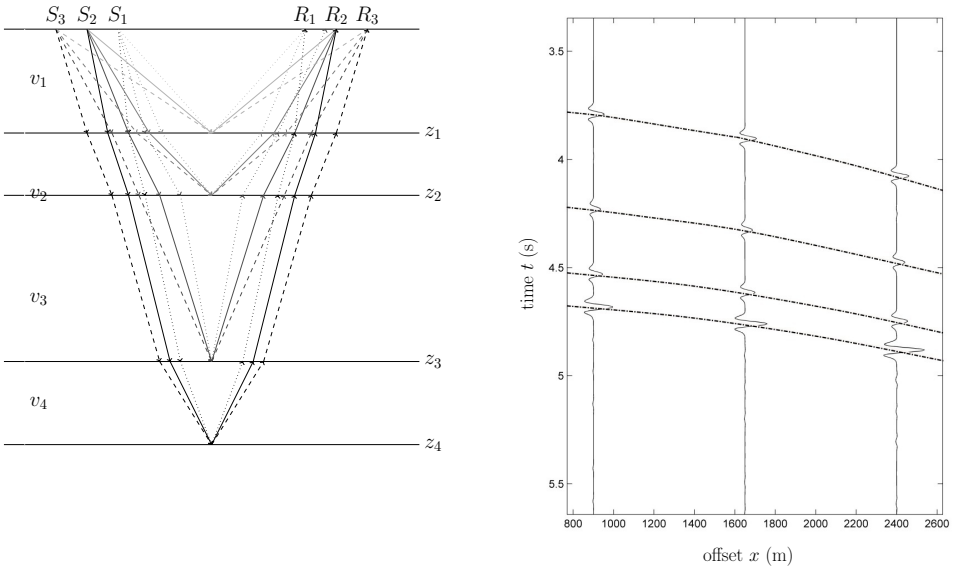
3 Velocity Models for Backscattered Seismic Reflection Data

Consider the case where a single 1.5-D layer (i.e., layer that is part of a model in which velocities are assumed to be functions of depth only) is detected and assume that it is parallel to an array of m receivers (see Figure 2a). In this case, the travel-time equation

$$t_j^2 = t_0^2 + \frac{x_j^2}{v_x^2} \tag{1}$$

applies, where $x_j \geq 0$, $j = 1, \ldots, m$, is the distance of the j-th source-receiver pair, $v_x > 0$ is the horizontal velocity for the layer, $t_j \geq 0$ is the real travel time and $t_0 \geq 0$ is the two-way travel time of a vertically reflected ray, or

$$t_0 = \frac{2z}{v_z}, \tag{2}$$



(a) (b)

Figure 2: (a) Simple 1.5-D model showing the energy propagation path from sources S_1, S_2 and S_3 to receivers R_1, R_2 and R_3 through a series of layers with velocities $v_1, \ldots, v_4$ and zero-offset travel times $t_{0_1}, \ldots, t_{0_4}$ separated by horizons at $z_1, \ldots, z_4$. (b) Each vertical trace represents the amplitude of the reflected energy arriving at a given receiver with a known source-receiver offset x_1, x_2 and x_3 against time. The amplitude peaks on the trace correspond to the two-way travel time for the energy to travel from the source to the sub-surface reflectors and back to the receiver (e.g., the first spike around $3.75s$ on the first trace at $x_1 = 900m$ corresponds to the travel time from source S_1 to receiver R_1 when the energy is reflected from horizon z_1. This is represented by the lightest grey path in (a)). The shape of the pulse is a function of the band-limited nature of the seismic source. The dashed lines on (b) show the result of the automatic picking algorithm described in Section 6.

where $z > 0$ is the layer's depth (see Figure 1a) and v_z is the vertical velocity for the layer (see Kearey and Brooks (1992) and Yilmaz (1987)).

In the case of a 1.5-D seismic model where all n layers are parallel to an array of m geophones (see Figure 2a), the travel-time equation is given by

$$t_{ij}^2 = t_{0_i}^2 + \frac{x_j^2}{v_{x_i}^2},\ i = 1, \ldots, n; j = 1, \ldots, m, \tag{3}$$

where $x_j \geq 0$ is the offset distance (or the position of the j-th geophone relative to the source), $t_{ij} \geq 0$ is the ray's real travel time reflected from the i-th layer to the j-th receiver, $v_{x_i} > 0$ is the horizontal stacking velocity for the i-th layer approximated by

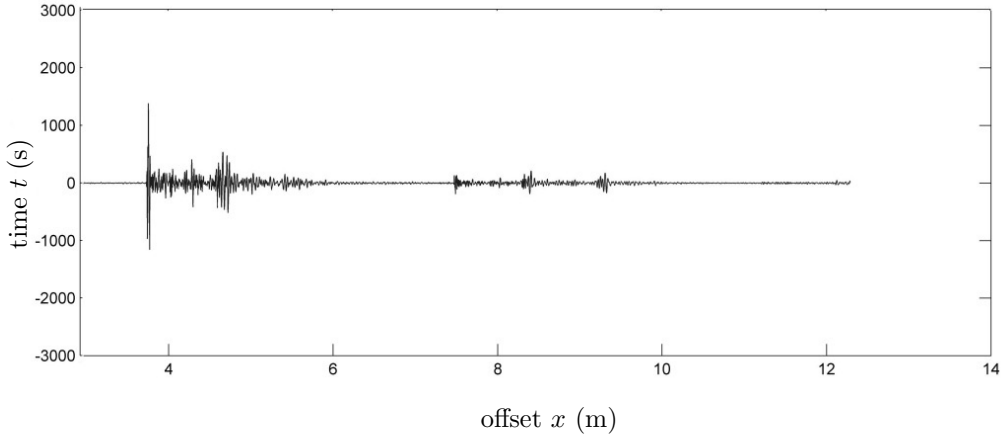


Figure 3: Plot of one recorded trace or wiggle for a given source-receiver distance x_j. The "spikes" on the trace correspond to possible layers. Even though the peaks seem clear, coherence over a set of traces is necessary in order to make a good pick. Therefore, we need to analyse gathers as a whole in order to confirm coherence.

the root mean square (RMS) velocity given by Dix's equation (see Dix (1952) and Dix (1955))

$$v_{x_k}^2 = \frac{\sum_{i=1}^{k} v_{int(i)}^2 (t_{0_i} - t_{0_{(i-1)}})}{\sum_{i=1}^{k} (t_{0_i} - t_{0_{(i-1)}})} \tag{4}$$

for $i > 1$ and $v_{x_1} = v_{int(1)}$, $v_{int(i)}$ is the real velocity for the i-th layer, $t_{0_i} \geq 0$ is the travel time of a vertically reflected ray from the i-th layer to the receiver, or

$$t_{0_i} = \frac{2z_i}{v_{z_i}}, \tag{5}$$

where z_i is the depth for the i-th layer (see Figure 2a) and v_{z_i} is the vertical velocity for the i-th layer.

As we discussed in Section 2, the only variable observed is the wave amplitude against time in a fixed offset. In order to build the model, we assume that the points forming coherent curves or layers were already picked and their travel times recorded. Usually, the data is processed using manual picking, i.e., an expert visually identifies the layers that appear to be relevant in each CMP gather. This is a time-consuming process which relies on expert judgements, where no quantitative estimate of uncertainty is made and layers near noise level are normally missed. The number of layers n is assumed to be

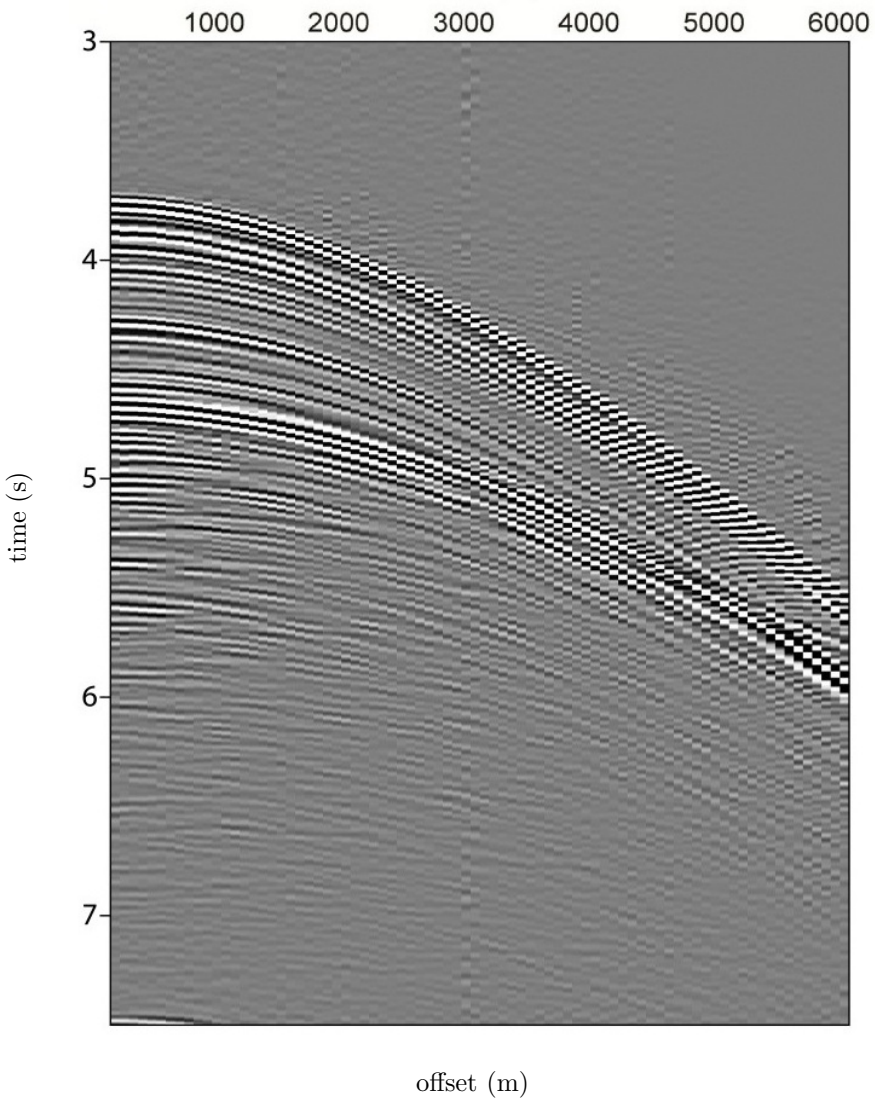


Figure 4: Grayscale map of a single gather formed by 80 recorded traces as in Figure 3. Note that the hyperbolic looking patterns are the ones we need to sort and model; the brightest ones usually correspond to significant changes in velocity and possibly a change in the sub-surfaces geology. The biggest challenges are to accurately sort the data closer to the seabed and to identify relevant velocity changes in deep water, in this case, time greater than 5s.

known and, in Section 6, we propose a method to identify layers without the need for

manual picking.

Here we described a geometrically simple problem but, even though the target horizons for real sub-surface geology are three-dimensional, multilayered and non-parallel, locally these horizons can be assumed to be a multilayered 1.5-D model. For the simple 1.5-D single layer model, estimating the corresponding velocity model is trivial; however, as more layers are added, the number of variables and the uncertainty sources grow. We also have to consider that the recorded data includes ambient noise and signal loss which degrades the ability to identify the reflected events accurately.

Therefore, we propose a Bayesian method which can produce a flexible and reliable model and an estimate of its uncertainty. Prior information for these Bayesian models include the tools used in manual picking: well-log data, outputs from other geophysical models, semblance plots and expert judgement. In the next sections, we build the models used to analyse this processed data and show how it works in synthetic and real examples.

4 Bayesian Models

Given a set of picked points, the data is analysed by two different Bayesian models: $(M1)$ isotropic model $(v_i = v_{x_i} = v_{z_i})$ estimating v_i and z_i (and subsequently t_{0_i}) and $(M2)$ anisotropic model $(v_{x_i} \neq v_{z_i})$ estimating v_{x_i}, v_{z_i}, z_i and t_{0_i}. Usually $v_{x_i} \approx v_{z_i}$, so the isotropic model $(M1)$ can be used to produce start points to the anisotropic model $(M2)$ and reduce the search area when new variables are inserted. Here the only observed variables relevant to the travel-time equations are the recorded travel times and the offsets and we want to estimate RMS velocities, depths with interval velocities, and two-way travel times as a by-product.

While the interval velocity represents the real velocity of the wave, it is the stacking velocity approximated by the RMS velocity that defines the hyperbolic shape. The interval velocity is a property intrinsic to the geology of the rock and more susceptible to variation between layers. Furthermore, its correlation with the travel-time increments is not trivial even using Dix's equation (Equation 4). The RMS velocities, on the other hand, usually create a more predictable smooth trend that handles weak priors well. When fairly accurate information about interval velocities is available (i.e., from well-logs and other surveys), priors for RMS velocities can be easily derived. Experts use RMS velocities and two-way travel times in their analyses more often than interval velocities and depths. In fact, most of the time, this information is used as an input in further modeling; interval velocities and depths are seen as a final output but rarely used as a modeling input. An alternative to the model proposed here is to use interval velocities and depth increments as our target variables but, in this research, we choose to use a combination that satisfies the industrial requirements and straightforwardly incorporates the expert information.

Define the variable $t^*_{ij} = t_{ij} + e_{ij}$ as the recorded travel time, where $e_{ij} \in \mathbb{R}$ is the error between the real and recorded travel time, assumed to be normally distributed

with zero mean and standard deviation $\sigma_{t_{ij}}$ given below. In both models, we make the assumption that depth and velocity are independent. Since the error is normally distributed, t^*_{ij} is normally distributed with mean

$$\mu_{t_{ij}} = \sqrt{4\frac{z_i^2}{v_{z_i}^2} + \frac{x_j^2}{v_{x_i}^2}} \tag{6}$$

and standard deviation

$$\sigma_{t_{ij}} = q_i \mu_{t_{ij}}, \tag{7}$$

where q_i is a Beta-distributed random variable with shape parameters α_i and β_i which can be set with the help of an expert. For example, if $\mu_{t_{ij}}$ is expected to be approximately $3.74s$ and $\sigma_{t_{ij}} \approx 0.01$ seconds, then we could set $\alpha_i = 4$ and $\beta_i = 2000$. In the case where no specific prior information is given about $\sigma_{t_{ij}}$, it is reasonable to assume that $\sigma_{t_{ij}} \leq 0.05$ seconds.

In the anisotropic model, we assume that $v_{z_i} = a_i v_{x_i}$ where a_i is normally distributed with mean around 1 and a small standard deviation $\sigma_{a_i} > 0$ which accounts for the fact that those velocities are usually similar. Now, assume that v_{x_1} is normally distributed with mean $\mu_1 \in \mathbb{R}$ and standard deviation $\sigma_1 > 0$ and write

$$v_{x_i} = v_{x_{(i-1)}} + s_{(i-1)} d_{(i-1)}, i = 2, \ldots, n, \tag{8}$$

where $d_{(i-1)}$ is Gamma distributed with shape $k_{d_{(i-1)}}$ and scale $\theta_{d_{(i-1)}}$, and $s_{(i-1)}$ indicates the change in velocity between two layers, i.e. v_{x_i} is greater than $v_{x_{(i-1)}}$ with probability γ_i. Finally, $v_{x_{(i-1)}}$ remains the same with probability δ_i and $v_{x_{(i-1)}}$ is less than v_{x_i} with probability $1 - \gamma_i - \delta_i$, where

$$s_{(i-1)} = \begin{cases} -1, & v_{x_{(i-1)}} > v_{x_i} \\ 0, & v_{x_{(i-1)}} = v_{x_i} \\ 1, & v_{x_{(i-1)}} < v_{x_i} \end{cases} \tag{9}$$

and

$$P\left(S_{(i-1)} = s_{(i-1)}\right) = \begin{cases} \gamma_i, & s = -1 \\ \delta_i, & s = 0 \\ 1 - \gamma_i - \delta_i, & s = 1. \end{cases} \tag{10}$$

In most cases, experts assume that the odds of having $s_{i-1} = -1$ or $s_{i-1} = 0$ are very low, normally around 1% each.

While it might seem unnecessary to define RMS velocities recursively, this method allows the trend to be controlled, since the curve formed by the RMS velocities is expected to be fairly smooth and mostly increasing. For example, if the water layer has a RMS velocity of 1480m/s, it is highly unlikely that the RMS velocity of the next layer will be less than that. This semi-Markovian recursive definition results in a

low correlation between jumps in RMS velocities and simplifies the detection of empty layers, i.e., layers with the same properties as the previous one. It also simplifies the diagnosis of the model and allows local prior updates.

Now assume that the thickness of the first layer, z_1, is normally distributed with mean $\mu_{z_1} > 0$ and standard deviation $\sigma_{z_1} > 0$ and that

$$z_i = z_{(i-1)} + d_{z_{(i-1)}}, \tag{11}$$

where $d_{z_{(i-1)}}$ is a Gamma-distributed random variable with parameters $k_{d_{z_{(i-1)}}}$ and $\theta_{d_{z_{(i-1)}}}$.

For similar reasons used to justify the recursiveness of the RMS velocities, defining depths with increments guarantees a monotonically increasing depth, helps detect extremely thin or repeated layers and gives us the advantage of a low correlation between layers' thicknesses.

Writing the posterior distribution for the multilayered model, we have

$$\begin{aligned}
\pi\left(\mathbf{v}_x, \mathbf{a}, \mathbf{v}_z, \mathbf{z}, \mathbf{q} | \mathbf{t}^*, \mathbf{x}\right) \propto\ & f\left(\mathbf{t}^* | \mathbf{v}_x, \mathbf{v}_z, \mathbf{z}, \mathbf{q}, \mathbf{x}\right) \pi_{(\mu_1,\sigma_1)}(v_{x_1}) \pi_{\left(\mu_{z_1},\sigma_{z_1}\right)}(z_1) \\
& \times \prod_{i=2}^{n} \pi_{\left(\gamma_{(i-1)},\delta_{(i-1)}\right)}(s_{(i-1)}) \pi_{\left(\theta_{d_{(i-1)}},k_{d_{(i-1)}}\right)}(d_{(i-1)}) \\
& \times \prod_{i=2}^{n} \pi_{\left(\theta_{d_{z_{(i-1)}}},k_{d_{z_{(i-1)}}}\right)}(d_{z_{(i-1)}}) \prod_{i=1}^{n} \pi_{(\alpha_i,\beta_i)}(q_i) \pi_{\left(\mu_{a_i},\sigma_{a_i}\right)}(a_i) \\
& \propto f\left(\mathbf{t}^* | \mathbf{v}_x, \mathbf{v}_z, \mathbf{z}, \mathbf{q}, \mathbf{x}\right) \exp\left[-\frac{\left(v_{x_1} - \mu_1\right)^2}{2\sigma_1^2}\right] \exp\left[-\frac{\left(z_1 - \mu_{z_1}\right)^2}{2\sigma_{z_1}^2}\right] \\
& \times \prod_{i=2}^{n} P\left(S_{(i-1)} = s_{(i-1)}\right) \left[d_{(i-1)}^{k_{d_{(i-1)}}} \exp\left(-\frac{d_{(i-1)}}{\theta_{d_{(i-1)}}}\right)\right] \\
& \times \prod_{i=2}^{n} \left(d_{z_{(i-1)}}\right)^{k_{d_{z_{(i-1)}}}} \exp\left(-\frac{d_{z_{(i-1)}}}{\theta_{d_{z_{(i-1)}}}}\right) \\
& \times \prod_{i=1}^{n} \left\{\left[q_i^{\alpha_i - 1}\left(1 - q_i\right)^{\beta_i - 1}\right] \exp\left[-\frac{\left(a_i - \mu_{a_i}\right)^2}{2\sigma_{a_i}^2}\right]\right\}
\end{aligned} \tag{12}$$

where $\mathbf{t}^* = \left\{t_{ij}^*\right\}_{i=1,\ldots,n;j=1,\ldots,m}$, $\mathbf{x} = \left\{x_j\right\}_{j=1,\ldots,m}$ are the observed values and $\mathbf{v}_x = \left\{v_{x_i}\right\}_{i=1,\ldots,n}$, $\mathbf{a} = \left\{a_i\right\}_{i=1,\ldots,n}$ and $\mathbf{z} = \left\{z_i\right\}_{i=1,\ldots,n}$ are the variables that we aim to estimate. As a result, $\mathbf{v}_z$, $\mathbf{t_0}$ and $\mathbf{q}$ are also estimated. The likelihood function

$f\left(\mathbf{t}^*|\mathbf{v}_x, \mathbf{v}_z, \mathbf{z}, \mathbf{q}, \mathbf{x}\right)$ is given by

$$f\left(\mathbf{t}^*|\mathbf{v}_x, \mathbf{v}_z, \mathbf{z}, \mathbf{q}, \mathbf{x}\right) \propto \prod_{i=1}^{n} \left[\frac{\exp\left[-\sum_{j=1}^{m} \frac{\left(t_{ij}^* - \sqrt{4z_i^2/v_{z_i}^2 + x_j^2/v_{x_i}^2} \right)^2}{2q_i^2 \left(4z_i^2/v_{z_i}^2 + x_j^2/v_{x_i}^2 \right)} \right]}{\prod_{j=1}^{m} \left(q_i \sqrt{4z_i^2/v_{z_i}^2 + x_j^2/v_{x_i}^2} \right)} \right]. \tag{13}$$

The number of parameters to estimate grows with the number of layers and so does the number of constants to be set by an expert using his/her own judgement and uncertainty about the model. Since we assume that $\mathbf{t}^*$ and $\mathbf{x}$ are given, the information on the amplitudes is not used. Priors for velocity and depth jumps are set with the help of an expert and the subspace where this posterior exists is bounded by the solution of the Dix equation (Equation 4) for interval velocities as follows:

$$v_{int(i)} = \begin{cases} v_{x_1}, & i = 1 \\ \sqrt{\dfrac{\left(v_{x_i}^2 t_{0_i} - v_{x_{i-1}}^2 t_{0_{i-1}} \right)}{t_{0_i} - t_{0_{i-1}}}}, & i > 1 \end{cases}. \tag{14}$$

If useful prior information on pairs of interval velocities and two-way travel times is available, then it can be transformed into a prior for RMS velocities and depths, otherwise the equation above should be used only to guarantee that the interval velocities are kept in the domain of positive real numbers. Other prior trends can be imposed using this result, say by requiring an increasing interval velocity curve.

Given such a complicated model, the Metropolis-Hastings algorithm is used to obtain estimates for the pertinent variables. The results obtained at this stage are usually accurate for practical concern but, to refine them and reduce the modelling error, a Gaussian process modelling phase based on a Taylor expansion of (3) can be applied as we describe in Section 5.2.

5 Estimation

In this section, we suppose that we observe each t_{ij}^*, $i = 1, \ldots, n; j = 1, \ldots, m$ and that the number of layers n is fixed. In the isotropic model, our aim is to estimate the depth (z_i) and the velocity (v_{x_i}) for each layer. In the anisotropic case, we also estimate $\mathbf{z}$ and $\mathbf{v}_{z_i}$ and additionally $\mathbf{a}$. In this section we introduce the tools used to estimate the model parameters. As mentioned before, we use the Metropolis-Hastings algorithm to obtain these estimates and we provide a tool to refine these values using a Gaussian process.

When informative priors exist, especially in the number of layers, and the target space is small, the steps in Section 5.1 become optional since its main purpose is to

reduce the search space, locate new layers when used jointly with the algorithm in Section 6 and retrieve quick estimates for the target parameters, especially depths and velocities. The most desirable feature of the method in Section 5.1 is computational speed and, in Section 5.2, precision.

5.1 Metropolis-Hastings Algorithm

The Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithm (M-H) (see Metropolis et al. (1953) and Hastings (1970)) generates n states of a Markov chain by drawing samples from a target probability distribution with unknown shape; in our case, the posterior distribution given by (12).

The chain is initiated using estimates from previous analyses, expert judgement or a random value inside the target search interval. Under normal circumstances, any of these options allow the chain to converge to the expected region in the sampling space. The only differences are the number of iterations necessary for the chain to converge and the number of samples discarded during the "burn-in" stage.

Given the number of variables to be estimated, the complexity of the model and the size of the search area, we use a component-wise sampling strategy. In order to generate a d-dimensional chain update X_n, we use d nested steps each consisting of an ergodic chain. While not the most elegant method, the simplest and most efficient approach is to use d uniform proposal distributions over the search interval for each variable to be estimated. As shown by Roberts and Rosenthal (2007), adaptive hybrid algorithms like this are ergodic and satisfy the law of large numbers. While it is not possible to cover the whole search space, we can easily monitor if the chain is walking towards a boundary and if the algorithm is showing signs of divergence. It is common that a few marginal posteriors, usually the ones related to the target variables, show signs of convergence to a region of the search space but, in order to assure we have a good uncertainty analysis of the whole model, the chain is run until all marginals show signs of stability. In most examples where expert prior information was used, we noticed that it is enough to sample each component around $10,000$ times in order to have convergence. When no prior information or extremely large search intervals were used (e.g. $0 \leq v \leq 15,000$), around $100,000$ samples were necessary.

It is natural to believe that there is a high correlation between all the variables involved a priori, which is true when talking about RMS velocities and depths. However, as we mentioned before, the difference between velocities and depths in each layer present a rather low correlation. Since the purpose of this stage is to produce quick rough estimates from a statistical point of view, it is unnecessary to spend resources in a more time consuming model that might not produce better results. This analysis is used later on in Section 6 with a search algorithm that determines the number of layers in the model. It would be unreasonable to create a correlation matrix or use a more complicated model while trying to sort the data efficiently. As an example, the pre-sorted model for the real data in Section 7 would produce an initial square matrix with 6144 rows, roughly 288 megabytes of memory, for each variable, which would have to be

updated and cropped through the MCMC. This would be inefficient but feasible if just a single CMP gather were to be analysed at a time, but it would be impraticable in an array of gathers. Due to the low correlation between differences and the nature of these variables, it is acceptable to assume independence to generate rough approximations and reduce the target space for the Gaussian process in Section 5.2.

When the initial search space is unfeasibly large, in order to avoid exploring around improbable regions of the parameter space, the search area is reduced by using $p\%$ credibility intervals generated by the last n runs as boundaries for the next n runs. For example, if $10,000$ samples are generated and 99% credibility intervals are estimated, than the next $10,000$ samples will be generated from points inside this subspace. This way, after repeating the process 5 times, the final intervals will cover roughly 95% of the posterior distribution. Most of the strategies suggested here, while unconventional, are fast and flexible, particularly when combined with Section 6.

5.2 Gaussian Process Modelling

The Metropolis-Hastings algorithm is a robust and time-efficient method for velocity models. When applied to synthetic examples, it returns expected values for the parameters we want to estimate close to their real values but the uncertainty related to these values is not well resolved mainly due to the assumption of a hyperbolic fit and recursive errors from the use of Dix's equation. In order to improve the accuracy of the estimates, we apply a Gaussian process model using the target space defined by the M-H intervals (see Rasmussen and Williams (2006)). In order to avoid convergence problems, the first prior from the M-H is used as the prior for this process or a uniform distribution over the M-H credibility interval.

Rewriting the travel-time equation (3), we have:

$$t_{ij} = \sqrt{t_{0_i}^2 + \frac{x_j^2}{v_i^2}},\ i = 1, \ldots, n,\ j = 1, \ldots, m \tag{15}$$

where $t_{0_i} = 2\dfrac{z_i}{v_{z_i}}$ and $v_i = v_{x_i}$ in the anisotropic case.

Expanding in a Taylor series of x_j:

$$t_{ij} = t_{0_i} + \frac{x_j^2}{2v_i^2 t_{0_i}} - \frac{x_j^4}{8v_i^4 t_{0_i}^3} + \frac{x_j^6}{16v_i^6 t_{0_i}^5} + o\left(x_j^8\right),\ i = 1, \ldots, n,\ j = 1, \ldots, m. \tag{16}$$

Now we can write

$$\begin{aligned} t_{ij}^* &= g_j\left(t_{0_i}, v_i\right) + \epsilon_{ij} \\ &= \hat{g}_j\left(t_{0_i}, v_i\right) + \epsilon_j\left(t_{0_i}, v_i\right) + \epsilon_{ij} \\ &= t_{0_i} + \frac{x_j^2}{2v_i^2 t_{0_i}} - \frac{x_j^4}{8v_i^4 t_{0_i}^3} + \frac{x_j^6}{16v_i^6 t_{0_i}^5} + \epsilon_j\left(t_{0_i}, v_i\right) + \epsilon_{ij}, \end{aligned} \tag{17}$$

where ϵ_{ij} is normally distributed with zero-mean and standard deviation σ_{ij} and $\epsilon_j\left(t_{0_i}, v_i\right)$ is a Gaussian process with zero-mean and covariance matrix

$$Cov\left(\epsilon_j(t_{0_k}, v_k), \epsilon_j(t_{0_l}, v_l)\right) = \sigma_j^2 \exp\left[-\theta_{1j}\left(t_{0k} - t_{0l}\right)^2 - \theta_{2j}\left(v_k - v_l\right)^2\right]. \tag{18}$$

The limited search space increases the rate of convergence of the chains, thus providing us with enough information to re-estimate the target parameters. Approximately 99% of the model is described by the polynomial part of the process so $\epsilon_j(\cdot,\cdot)$ and ϵ_{ij} work as tuning parameters. As expected, ϵ_{ij} will be relevant if the data is noisy or presents peculiar characteristics while $\epsilon_j(\cdot,\cdot)$ represents the modelling error.

The Taylor expansion is not a necessary step but is commonly used by geophysicists to reduce the hyperbolic effect. We therefore follow this approach to allow comparability with industry procedures. Using equation 15 instead of equation 16 in 17 causes a small increase on the modelling error but still produces the desired results. This is due to the fact that the curvature of the fit for large values of x is overcorrected in 15. By using the Taylor expansion in 16, part of this error is accounted for. Also, it is not necessary to use a polynomial of degree 6, as the only changes for different choices of degree would be on the modelling error and computational time. The chosen degree is widely used in industry and it is quite convenient given the precision and speed it offers.

It is interesting to note that, by running the model in Section 5.1 for longer, it is possible to approach the results obtained using the Gaussian process in the sense that the posterior means are almost the same as we would expect. However, as mentioned before, that model is still limited by independence assumptions which result in unresolved uncertainty around each parameter.

6 Sorting the Data

We now explain how the raw data is sorted; the peaks in the data have to be selected and matched to their respective layers and these layers have to be validated to form a coherent model. We call this procedure the Automatic Picking Algorithm which we present in detail in the next subsection. Here we will assume that a CMP gather is formed by a m-by-N matrix, A, of recorded amplitudes. Given this matrix A, we aim to end up with the number of layers $n \leq N$ and ${t^*}_{i=1\ldots n,\ j=1\ldots m}$ which are required in Section 5.

Roughly, the algorithm searches each trace for local maxima and minima, matches trends formed by these points using a correlation and semblance analysis (see Section 6.2) and saves the best candidates for further analysis using the processes discussed in Section 5. As mentioned before, the use of a semi-Markovian definition of velocity and depth jumps, in conjunction with independent samplers and a component-wise M-H allows us to locally search the dataset for layers by letting the number of layers vary without compromising the rest of the model, as we can see in the next subsection.

6.1 Automatic Picking Algorithm

It was assumed, in Sections 3 to 5, that the recorded travel-times and the number of layers were given. Here we describe how to select the points from the matrix of recorded amplitudes and match them to possible layers. We have tried to mimic some of the steps that an expert would take when presented with a raw data file. The process is a matching algorithm that detects the outlines of the picture described by the data. It is somewhat similar to a sharpening filter that enhances the main features in a picture and fades the background noise.

This statistically-based automatic peak-picking process consists of detecting peaks in each CMP gather and matching them to their respective layers using adjacent gathers to include or exclude new points or layers. Details of the method are listed here.

1. Treat the first trace of the CMP gather $A_{\cdot 1}$ as a curve, pick its extremes (points with the highest and lowest amplitudes; these amplitudes are always non-zero since the signal is bipolar) and save their recording times (t^*_{k1}) and amplitudes (a_{k1}). This sets an upper boundary p for the number of possible layers $k = 1, \ldots, p$. Minimum peak height, minimum peak distance and threshold can be set here also to reduce the number of noisy picks. A reasonable choice for minimum peak height is the average absolute height of the last 5% of points in each column or the receiver's factory standard error.

2. Start counter $k = 1, \ldots, m$ where m is the number of traces. Start counter $p = 1, \ldots, L_{2k}$ where L_{2k} is a preset lag on the number of traces to be used on the next step.

3. Given t^*_{kp}, pick the extremes at $t^*_{kp} \pm l_{1k}$ (l_{1k} is a pre-set lag on the time difference between two points in the same layer like the minimum recordable wavelength) on trace (k+1) and assign to $a_{(k+1)p}$ the extreme that retrieves the minimum absolute difference when compared to a_{kp}. For example, if $t^*_{11} = 3.74s$ and $l_{11} = 0.02s$ with amplitude $a_{11} = 1020$, then we'll search the second trace for peaks in the inverval $[3.72s, 3.76s]$ with amplitudes close to a_{11}. Alternatively, use priors (velocity and zero-offset travel time with a two-way travel time of a vertically reflected ray) or a semblance plot (discussed in 6.2) to build hyperbolic guide curves based on (3) and set lags for each parameter.

4. After the first 3 steps, there might be too many layers picked, most of them following an unreasonable pattern. If so, given the first L^*_{2k} points of each layer k, fit a least-squares non-linear model $t = \sqrt{a^2 + x^2/b^2}$, $a > 0$, $b > 0$ and reject the layers with low R^2 since that indicates a probable misfit given that we assumed that our candidates should follow an approximate hyperbolic curve.

5. Run a low-precision (around 10,000 runs) M-H algorithm based on the isotropic model $M1$ described in Section 4 to the first L^*_{2k} points of each layer k. At this point, we assume that we have a single-layer model, i.e. assume that layer k is the only layer in the model, and fit $M1$ to each layer k individually. If no prior

information is given for each layer, use a and b as estimates for the mean of t_0 and v, $\delta \cdot a$ and $\delta \cdot b$ as estimates for its standard deviation where $\delta \approx 0.1$. Use the M-H estimated velocity and t_0 to pick the next L_{3k} points.

6. Repeat steps 3-5 until all traces are used.

7. Fit a single-layer isotropic model (see Section 4) to each layer and estimate its parameters. Next build hyperbolic boundaries around the picked points using (3) and re-pick the most probable points in each layer.

0. [Optional] A filter may be used to help eliminate part of the background noise and enhance the relevant peaks if applied before step 1. We've considered three filters that can be used individually or combined with each other: Hilbert envelope (see Hahn (1996)), high-pass filter (reduces the amplitude of signals with frequencies lower than the cutoff frequency), low-pass filter (reduces the amplitude of signals with frequencies higher than the cutoff frequency).

After passing the data through this algorithm, we have two r-by-n matrices: $\mathbf{A}$ of recorded amplitudes and $\mathbf{t}^*$ of corresponding travel times where r is the number of layers picked and n, the number of traces in gather A. Both matrices are used to estimate the required parameters using a full multi-layered isotropic or anisotropic model as described in Section 4.

6.2 Semblance Plot

The semblance plot mentioned on step 3 of the Automatic Picking Algorithm is a map created by calculating the semblance for a sample of possible velocities and zero-offset travel times. Given a velocity and a zero-offset travel time, we draw a guideline on the current CMP gather using Equation (1) and pick the amplitudes a_j^+ of the closest points to this line in each trace j, $j = 1, \ldots, m$. The semblance is calculated as follows:

$$S = \frac{\sum_{j=1}^{m} \left(a_j^+\right)^2}{m \left[\sum_{j=1}^{m} a_j^+\right]^2}. \tag{19}$$

The semblance is a dimensionless measure that ranges from 0 to 1 where the most coherent picks are closer to 1. A sample semblance plot is shown in Figure 6.

6.3 The Model Structure

The model we use here can be divided in four interlinked steps: (1) create a semblance plot, (2) pick the points with highest semblance values, (3) match these possible picks to the ones selected by the automatic picking algorithm described above and discard

incoherent points and (4) run the Bayesian models (described in Section 4) to estimate the parameters involved. On step (2), we identify likely interfaces by either setting a minimum semblance threshold or using previous results. On step (3), we match the peaks on the data with their most probable layers. As mentioned before, we use the semblance plot and the M-H algorithm to create boundaries and help the automatic picking algorithm "snap" to the right points. Finally, at this stage, the data is ready to be analysed; we use the Bayesian models in Section 4 to provide the final estimates required. Here, additionally, we can use the Gaussian process stage described in Section 5 to further explain the model variability.

7 Examples

In this section, we present two data sets to test the model suggested in this research against traditional methods. The first consists of a synthetic data set created using a finite-difference algorithm that was analysed with and without added noise. In the second one, we use a real data set consisting of a single gather.

7.1 Synthetic Data

This 6-layer synthetic gather was produced based on 6 relevant layers from the real data example described in the next subsection. The actual model used in the synthetic example is given in the first two columns of Table 1. The synthetic data example is necessary in order to verify the precision of the automatic picking algorithm in data sets with a controlled noise level. In this case, we first added a small amount of uncorrelated noise sampled from a normal distribution with mean zero and variance estimated from all points on each trace. Additional noise was created by adding to each point a sample from a Normal distribution with zero mean and covariance matrix estimated using a 15×15 matrix centred in the corresponding data point. In total, the amount of noise added to points near a relevant peak is proportional to its amplitude as we would expect to happen with real data. We subject the noise-free and the noisy CMP gathers to different picking methods: manual picking by one of our trained operators (Richard Hobbs), automatic picking without prior information and automatic picking with prior information.

We used the noise-free synthetic as a control data set; the three methods picked the right 6 gathers with negligible error. Four kinds of priors were analysed: we restricted the search area to (1) points around real t_0 and v values, (2) points around real t_0 values, (3) points around expert judgement on t_0 and v and (4) vague priors on t_0 and v. The first set of priors was relaxed, i.e., the real values for the first layer are 3.743 seconds and 1480 metres per second so we assumed that $t_0 \in (3.7, 3.8)$ and $v \in (1450, 1500)$. The second method is similar to the first one except that all velocities were amply constrained in an interval between 0 and 15000 m/s. In the case where we used vague priors, no information about the number of layers was given and velocities were constrained between 0 and 15000 metres per second and time between 0 and 10 seconds.

In the cases where we fed the automatic picking algorithm with prior information on only t_0 and on both t_0 and v, the M-H converged faster to the target area than when only vague information was given. We also tested cases where a different number of layers were given and the automatic picking algorithm was able to recover or eliminate the relevant layers leading to the same results. Apart from when we omitted the picking stage, this method produced similar results consistently with the noise-free synthetic, regardless of how unreasonable the priors were. When the picking stage is removed, the search phase that finds new layers is not present. Hence, inputting a smaller number of layers than the one in the model will result in a model with the same number of layers or less. Note that it is still possible to reject layers without the picking stage since layers that are too similar are eliminated, i.e. those with a velocity or time difference below the data precision or those with negligible thicknesses.

We subjected the synthetic data with noise to similar experiments which gave results comparable to the noise-free one. The results for the case where correlated Gaussian noise was added are presented on Table 1. The results for the Automatic picking column used a non-informative prior for t_0 and v.

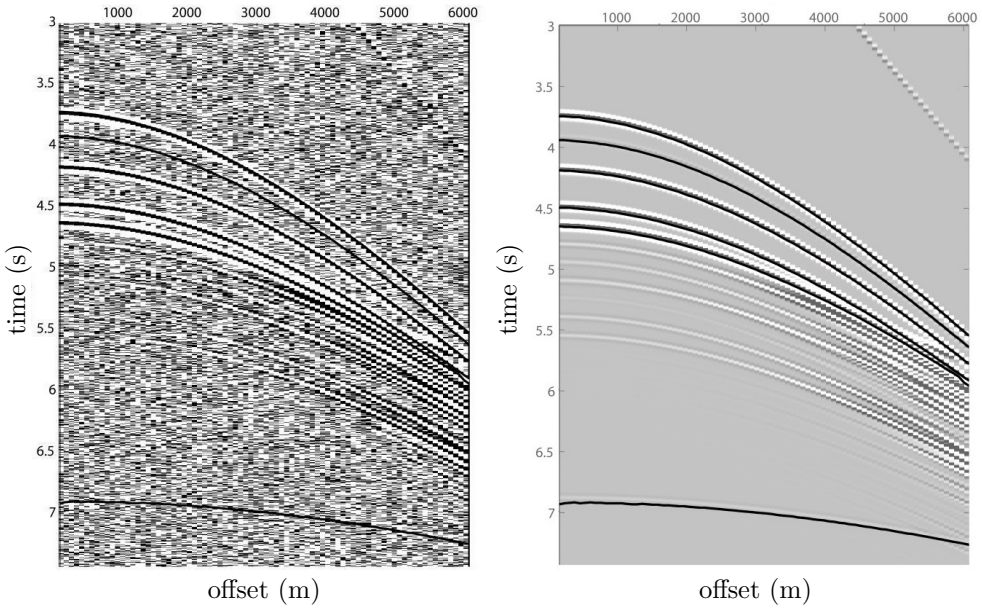


Figure 5: Left panel: noise-free synthetic data with layers picked by the automatic picking algorithm. Right panel: synthetic data with enough correlated Gaussian noise added to "hide" the second and the sixth layers, i.e. the average signal amplitude for those two layers is around noise level.

Real Values		Manual Picking		Automatic Picking		
t_0 (s)	v (m/s)	t_0 (s)	v (m/s)	t_0 (s)	v (m/s)	Sd(v)
3.743	1480	3.7055	1485.1	3.7409	1481.81	0.3799
3.934	1500	-	-	3.9393	1502.6	0.3548
4.194	1520	4.1841	1528.9	4.1889	1523.9	0.4046
4.497	1565	4.4822	1567.2	4.4959	1565.8	0.4130
4.650	1605	4.6346	1615.7	4.653	1607.5	0.8641
6.888	2630	-	-	6.8486	2629.152	2.0915
		7.8407	2849.9			

Table 1: Estimates for velocities and travel-times using the noisy synthetic data. Time is measured in seconds (s) and velocity in metres per second (m/s). Standard deviations for t_0 were negligible.

The proportion of noise added here is much higher than what would be found in a normal real data set but provides us with an example that challenges the automatic picking algorithm to identify layers that would be missed through manual picking by a trained operator. In this case, the manual picking method was unable to identify two of the layers at noise level while the automatic picking method was able to identify all of them and return estimates close to the original values. In all cases, the standard deviation of t_0 was reduced to or below the resolution of the data becoming negligible from a practical point of view. If we consider that the hyperbola in Equation 1 is a good approximation for each single layer, it is easy to check that its eccentricity is a function of the velocity only and it approaches 1 as the velocity increases. It is reasonable to expect such a small error since there are more than enough data points to offer a precise estimate for t_0. Without considering the effects of adjacent layers, we should expect proportionally higher variances for high velocities on the pre-Gaussian modelling stage as we are approaching asymptotes by using the hyperbolic approximation. The Gaussian modelling stage should resolve most of the error resulting from this assumption.

In all cases, providing the automatic picking algorithm with prior information allowed it to converge much faster to satisfactory values but the fact that priors influence only the speed of convergence shows that the model is robust enough to handle synthetic data and similar situations. During our tests, we've also fed the M-H algorithm with bad priors that are physically plausible (i.e., non-negative starting points and velocities no higher than $15000m/s$) and the estimates obtained were still within satisfactory limits.

On the layers that were meant to be hidden by noise, reasonable estimates were obtained in 100,000 runs without priors and in 10,000 runs with prior information on t_0. In all cases, the model explained around 99% of the error. The remaining variation can probably be explained by numerical errors and approximations. Given the high dimensionality of the problem, especially when multiple CMP gathers are analysed, it is unreasonable to manually check each chain for convergence so we have automated a few diagnostics strategies.

Mixing is one of the main issues in multidimensional problems and the fastest way to verify mixing problems is by checking the stability of the marginal chains. As each of the marginal chains are built, their running mean and median series are checked for stability using lags of 10 and 100. The structure of long-lag auto-correlation series is also checked in order to avoid highly correlated samples.

Another issue is related to different rates of convergence for different parameters. The chains for depths and velocities tend to take longer to stabilise when compared to the two-way travel times. To assure that all space was explored, we either tune the variance of the proposal distribution for each parameter to constrict or expand their search area or the acceptance rate with simulated annealing to assure stability for all chains.

The last problem that has to be addressed is the uncertainty around the number of layers. While we expect to find at least one layer in a given data set, there is the possibility that the recorded signal was corrupted by noise and no layers can be selected, even though there is a large number of candidates as the ones mentioned in step 4 of the picking algorithm in Section 6.1.

Given that one layer was selected, say the seabed, we might still have too many candidates. So, if we assume that the total number of candidates is the maximum number of possible layers, we can allow "bad" candidates to converge to possible good ones like the seabed. If the first candidate after the seabed was picked happened to have zero depth or a complex interval velocity, then this implies that such a layer does not exist in that specific CMP gather so we can either delete it if a single CMP gather is being analysed or replace it with the seabed estimates if we have multiple gathers to allow new layers to be created. Using this method and including an empty model candidate, the possibility of choosing a bad model because the number of layers was preset should be reduced.

7.2 Real Data

The real data example is from a seismic survey acquired over the Naturaliste Plateau and Mentelle Basins off the south west coast of Australia (see Borissova (2002)). The data set is from a deep water environment where complications in the received signal due to reverberation of the seismic energy in the water layer can be ignored because the travel-time for these arrivals is longer than the target reflections. The research interest in these data is the high latitude Cretaceous black shales that were deposited on this margin during a period of extreme high global temperature which may be related to sudden decreases in atmospheric CO2 concentrations (see Kuypers et al. (1999)) and recorded sporadic short-lived glaciations (see Bornemann et al. (2008)). Additional prior control on the geological structure is obtained from a Deep Sea Drilling Project borehole number 258 (see Davies et al. (1974)) which provides depths to key boundaries and estimates of acoustic velocity from recovered rock samples.

Currently the seismic data from this area is being reprocessed and reinterpreted. As part of this research, a robust quantified velocity model is required to enable generation

of depth sections and so provide the opportunity for a comparison of the automated Bayesian based methods described here with the traditional manual method. The seismic gather analysed here is close to the borehole and comes from a seismic survey acquired in 2004 by Geoscience Australia (Borissova (2002)).

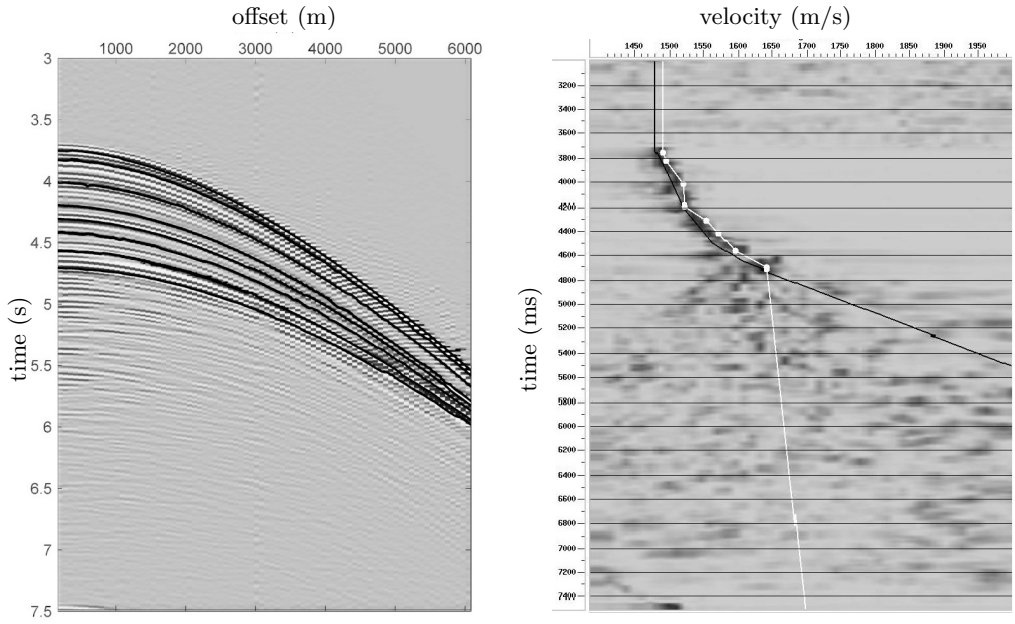


Figure 6: The most probable layers picked from CDP 1624 using the automatic picking algorithm without filters and minimum peak distance set at 100ms. On the left, we have the picked points on the original gather and on the right, the same layers picked on a semblance plot where the darkest areas correspond to high semblance points. The black line shows the picks obtained by a trained operator and the white line shows the results obtained using the Bayesian method. The direction of the lines after the last peak is irrelevant.

In this example, we used a real data set to compare the results obtained by the models and those obtained by traditional methods. At first the automatic picking algorithm without informative priors on t_0 and v produced over 200 layer candidates that were reduced to approximately 50 and later to 8. The results are listed in Table 2. The data provided was clean enough to return standard deviations lower than expected for most layers, i.e., less than one millisecond for t_0 and less than one meter per second for v.

The results obtained are consistent with the ones obtained using manual picking since the points picked (white points on the semblance plot in Figure 6) are close to

Manual Picking	
t_0 (s)	v (m/s)
3.7437	1480.5
3.9405	1498.1
4.1943	1518.9
4.4973	1563.5
4.6501	1606.0
6.8877	2632.2

Automatic Picking			
t_0 (s)	v (m/s)	Sd(v)	P(Layer)
3.7494	1490.4	0.6	0.9993
3.8228	1494.1	0.8	0.9998
4.0129	1520.2	1.0	0.9992
4.1947	1522.6	1.2	0.9998
4.3095	1554.6	0.9	0.9994
4.4198	1573.7	1.1	0.9990
4.5633	1595.3	1.6	0.9988
4.7041	1642.3	1.2	0.9880

Table 2: Velocities and travel-times obtained using manual picking and estimated velocities and travel-times and the probability that each layer is part of the model obtained using automatic picking for the real data - time in seconds and velocity in meters per second. The standard deviations for t_0 were negligible.

the trend line formed by manual picks (black line on the semblance plot in Figure 6). Even though most of the points picked manually were acceptable when compared to the candidates given by the automatic picking algorithm, they weren't selected as the most probable picks. Here the MCMC tool used, the M-H algorithm, returned 95% credibility intervals on each parameter by using a nested approach of repeated runs that reduces the search area to 99% of the current posterior distribution as the algorithm shows stability. This reduced search space is then used in a Gaussian process modelling phase, based on a more exact solution to the normal move-out equation, to further decrease the uncertainty to the noise level in the data by reassigning part of the modelling error.

In this work, the results obtained using the model described would be accepted as a good result and the point estimates selected by our method are comparable to those an experienced operator analysing the same data set could obtain. Moreover, we are able to derive an uncertainty model for our picks while an operator would be unable to do the same using traditional methods. When analysing a single CMP gather, informative priors are not necessary because we are just looking for the vertical velocity trend. To obtain better results using automatic or manual picking, it is interesting to use adjacent CMP gathers and extra information obtained from a larger data set to build models in which velocities vary on the horizontal and vertical axes and with depth. This way, it is possible to build a correlation structure between gathers and detect or delete layers more quickly and more accurately. This will be discussed in future research.

8 Conclusion

The automatic picking algorithm has proved that it can work on single gathers, both synthetic and real. It retrieves a large number of possible layers that are ranked according to their probability of pertaining to the current model, leaving space to improve the model by matching this information to adjacent gathers.

As expected, it is less stable as the signal to noise ratio decreases and tends to identify possibly spurious horizons but the same happens to our trained operators. Similar to the manual picking procedure, this will improve when dealing with real data sets where spatial oversampling provides additional data to help discrimination. Therefore, the Bayesian model proposed here proved to be more efficient than manual picking and also provided an analysis of uncertainty of the fitted solution which a trained operator cannot retrieve.

9 Glossary

1.5-D seismic model Velocity model in which velocities are assumed to be functions of the depth only. 214

acoustic impedance The product of density and seismic velocity, which varies among different rock layers. 211, 212

band-limited Fourier-transformed data that is restricted to a finite range of frequencies or wavelengths. 211

common midpoint (CMP) In multichannel seismic acquisition, the point on the surface halfway between the source and receiver that is shared by numerous source receiver pairs. Such redundancy among source-receiver pairs enhances the quality of seismic data when the data are stacked. The common midpoint is vertically above the common depth point or common reflection point. 211

geophone A device used in surface seismic acquisition, both onshore and on the seabed offshore, that detects ground velocity produced by seismic waves and transforms the motion into electrical impulses. 214

normal move-out (NMO) The effect of the separation between receiver and source on the arrival time of a reflection that does not dip. A reflection typically arrives first at the receiver nearest the source. The offset between the source and other receivers induces a delay in the arrival time of a reflection from a horizontal surface at depth. A plot of arrival times versus offset has a hyperbolic shape. 212

p-wave An elastic body wave or sound wave in which particles oscillate in the direction the wave propagates. P-waves are the waves studied in conventional seismic data. 211

receiver A device that detects seismic energy in the form of ground motion or a pressure wave in fluid and transforms it to an electrical impulse. 211

s-wave An elastic body wave in which particles oscillate perpendicular to the direction in which the wave propagates. S-waves are generated by most land seismic sources, but not by air guns. S-waves, or shear waves, travel more slowly than P-waves and cannot travel through fluids because fluids do not support shear. 211

semblance A quantitative measure of the coherence of seismic data as described in (6.2). 217

source A device that provides energy for acquisition of seismic data, such as an air gun, explosive charge or vibrator. 211

trace The seismic data recorded for one channel. A (seismic) trace represents the response of the elastic wavefield to velocity and density contrasts across interfaces of layers of rock or sediments as energy travels from a source through the subsurface to a receiver or receiver array. 211

two-way travel time The elapsed time for a seismic wave to travel from its source to a given reflector and return to a receiver at the Earth's surface. 213

References

(2010). "Schlumberger Oilfield Glossary." URL http://www.glossary.oilfield.slb.com 212

Borissova, I. (2002). "Geological framework of the Naturaliste Plateau." *Geoscience Australia Record*, 20. 229, 230

Bornemann, A., Norris, R. D., Friedrich, O., B. Beckmann, S. S., Damste, J. S. S., Vogel, J., Hofmann, P., and Wagner, T. (2008). "Isotopic evidence for glaciation during the Cretaceous supergreenhouse." *Science*, 319: 189–192. 229

Davies, T. A., Luyendyk, B. P., Rodolfo, K. S., Kempe, D. R. C., McKelvey, B. C., Leidy, R. D., Horvath, G. J., Hyndman, R. D., Thierstein, H. R., Herb, R. C., Boltovskoy, E., and Doyle, P. (1974). "Site 258." In *Initial Reports of the Deep Sea Drilling Project*, volume 26, 359–414. Washington (U.S. Government Printing Office). 229

Dix, C. H. (1952). *Seismic Prospection for Oil*. Harper & Brothers. 215

— (1955). "Seismic velocities from surface measurements." *Geophysics*, 20: 68–86. 215

Hahn, S. L. (1996). *Hilbert transforms in signal processing*. Artech House. 225

Hastings, W. K. (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrika*, 57: 97–109. 221

Kearey, P. and Brooks, M. (1992). *An Introduction to Geophysical Exploration*. Blackwell Scientific Publications. 214

Kuypers, M. M. M., Pancost, R. D., and Damste, J. S. S. (1999). "A large and abrupt fall in atmospheric CO2 concentration during Cretaceous times." *Nature*, 299: 342–345. 229

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). "Equations of State Calculations by Fast Computing Machines." *Journal of Chemical Physics*, 21(6): 1087–1092. 221

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press. 222

Roberts, G. O. and Rosenthal, J. S. (2007). "Coupling and Ergodicity of Adaptive MCMC." *Journal of Applied Probability*, 44: 458–475. 221

Yilmaz, I. (1987). *Seismic Data Processing*. Society of Exploration Geophysicists. 214

Acknowledgments

This project was fully funded by WesternGeco (Schlumberger) under the CeREES (Centre for Research into Earth Energy Systems) PhD Scholarship Programme at the Department of Earth Sciences jointly with the Department of Mathematical Sciences at Durham University. The authors thank Geoscience Australia for providing the seismic data used in this paper as part of a NERC funded project (NE/G001332/1) and Colin Sargent who, as part of the project, provided part of the manually picked data.