Geophys. J. Int. (1995) **123**, 653–664

Seismic velocity structure of oceanic crust by inversion using genetic algorithms

Guy G. Drijkoningen[1] and Robert S. White[2]

[1] *Department of Mining and Petroleum Engineering, Delft University of Technology,* 2628 RX *Delft, The Netherlands*
[2] *Bullard Laboratories, Department of Earth Sciences, University of Cambridge, Madingley Road, Cambridge* CB3 0EZ, *UK*

Accepted 1995 May 15. Received 1995 April 11; in original form 1994 November 15

SUMMARY

We determine the velocity structure along two expanding-spread seismic profiles, shot near the Blake Spur fracture zone in the western North Atlantic. We use the genetic algorithm as an optimization method in our inversion scheme. The genetic algorithm requires a forward modelling tool, for which we use kinematic ray tracing when traveltimes are required, and Chapman's (WKBJ seismogram) method when waveforms are needed. We optimize the seismic problem by first making a traveltime fit with velocity functions consisting of linear velocity gradients: these calculations are fast. Subsequently, we fit waveforms using B-splines for the velocity function. The splines give more consistent synthetic seismograms than linear velocity gradients because caustics caused by the model discretization do not introduce amplitude distortions, so we remain within the region of validity of asymptotic theory. We introduce a stopping criterion for genetic algorithms similar to the one used in crude Monte Carlo methods. Finally, we illustrate the whole procedure by applying the method to *P*- and *S*-wave refraction data, and compare the results of automatic inversion for the velocity–depth structure with results from trial-and-error forward modelling.

Key words: Chapman's method, genetic algorithms, waveform inversion.

INTRODUCTION

In this paper we present results of an automatic waveform-fitting procedure applied to two seismic data sets describing the oceanic crust. For the automatic procedure, we apply the genetic algorithm as a search method. This method starts with a randomly picked pool of models. The fitness of each model is calculated, and the method then uses the fitnesses of these models in order to produce a new set of models that have 'learned' from the previous set. We have modelled two expanding-spread multichannel seismic profiles, shot near the Blake Spur fracture zone in the north western Atlantic Ocean.

Most wide-angle seismic profiles are interpreted using trial-and-error forward modelling procedures, a time-consuming exercise. We have therefore developed an automated procedure using high-speed computers which does as good a job as trial-and-error modelling. Cary & Chapman (1988a, 1988b) described a successful inversion method using the Monte Carlo search method. Since the publication of Cary & Chapman's papers, new developments have taken place in non-linear methods, including the appearance of search methods known as the genetic algorithm (GA) and simulated annealing (SA). Simulated annealing has a firmer statistical basis than GAs, and some developments have taken place to speed it up using Cauchy distributions (Ingber & Rosen 1992; Stoffa *et al.* 1994; Varela, Stoffa & Sen 1994; Jervis, Sen & Stoffa 1993). In SA we use one model at a time, and build up a statistical knowledge of the system. The rate of convergence can be fast, although when using GAs we can make use of parallel processing, which could make it of a comparable speed. Also, in SA we need derivative information, such as how fast we are going to cool down the system, whereas this information is not necessarily needed in GAs. In this sense, the GA is a more general method than SA. Both methods, SA and GA, have their advantages and disadvantages, and, as with many of these methods, they have been combined in a hybrid fashion (Stoffa & Sen 1992). The use of the stretching of the objective function via an SA temperature scheme is problem-dependent, as illustrated in the book by Sen & Stoffa (1995). We have also used an SA temperature scheme in the GA, but we did not find any significant increase in the speed of the algorithm over using a simpler scheme, and therefore we kept to our simple scheme.

The results of applying GAs to the inversion of seismic refraction data have been discussed by Sambridge & Drijkoningen (1992). Their work was, however, focused on the performance of the GA, and not on the optimization of the seismic problem in all its aspects. We address the overall problem of optimizing inversions in this paper. We also introduce a criterion for stopping the search, similar to that

used in crude Monte Carlo techniques. We show some results of the use of this criterion, and compare them with the Monte Carlo method.

The GA needs a forward modelling scheme in order to calculate traveltimes and amplitudes. We use asymptotic ray theory to determine the local ray information, the traveltime and amplitude. In order to form the seismogram we combine the amplitudes from all the rays in the way proposed by Chapman (Chapman 1978; Chapman & Drummond 1982). Although Chapman's method is a good way of modelling signals in inhomogeneous media because it is fast, Cary & Chapman's (1988a,b) use of linear velocity gradients in their inversion scheme causes caustics to be produced at each discontinuity in gradient. When we apply a high-frequency method like Chapman's method, we must comply with the restrictions imposed by the theory: the model must be smooth. Artefacts introduced by discretization of the model can be troublesome in the inversion, unless these discontinuities are specifically desired in the results. We therefore use a 2-D B-spline for the velocity–depth function in order to smooth the models and to prevent artefacts caused by discontinuities in the velocity gradient.

Cary & Chapman (1988) implemented a Monte Carlo search method, which is computer-intensive, and consequently were limited in the number of parameters they could use (typically less than 15 depth points). An increase in the number of depth control points causes a nearly exponential increase in computation time. With the GA, this increase is not exponential but nearly linear. Therefore more parameters can be handled in the inversion for the same amount of computer time if a GA method is used.

Cary & Chapman (1988a,b) use three steps in the inversion scheme: first a Monte Carlo search, followed by a linearized inversion and finally by error analysis. The first step is needed for finding the global minimum, and the second step of linearized inversion for finding the local minimum within the region of the global minimum. Here we use the GA to replace the first two steps since, as we noted by Cary (1987), the inversion is non-linear in both the initial stage and the final stages, and the misfit around the region of the global minimum is not a smooth function. Also, the models after the first step had, in this cases, enough detail to satisfy the interpreter; therefore a fully non-linear technique can replace the first two steps. Cary & Chapman's third step, an error analysis, is used only for making an estimate of the error for the stopping criterion of the GA; a full error analysis is not implemented here because there are fundamental difficulties of doing so with a GA scheme (see Sen & Stoffa 1995).

We first outline the philosophy behind the inversion scheme, and then discuss the GA search procedure and the stopping criterion we use. We then discuss aspects of modelling real data in the presence of noise, and show the results of GA inversions of two seismic profiles from the western North Atlantic. Finally, we compare our inversion results for these profiles with results from trial-and-error forward modelling.

BAYES' RULE

In this section we start by discussing the philosophy behind the work done here; more specifically, the philosophy expressed by Bayes' rule. The GA, as used in this paper, falls within this statistical framework. Bayes' rule connects the information we have before we perform an experiment (the *a priori* information) to the information after we have completed the experiment (the *a posteriori* information). In short form:

$$\rho(\mathbf{m}\,|\,\mathbf{d}) = \frac{\rho(\mathbf{d}\,|\,\mathbf{m})\rho(\mathbf{m})}{\rho(\mathbf{d})}, \tag{1}$$

where $\mathbf{m}$ denotes the model vector, $\mathbf{d}$ denotes the data vector, $\rho(\mathbf{m})$ denotes the *a priori* joint probability density function, and $\rho(\mathbf{m}\,|\,\mathbf{d})$ the *a posteriori* joint probability function. $\rho(\mathbf{d})$ is taken as a constant, and the combination of $\rho(\mathbf{d}\,|\,\mathbf{m})/\rho(\mathbf{d})$ is known as the likelihood function, which we denote by $l(\mathbf{m}\,|\,\mathbf{d})$ in accordance with the definition by Cary (1987). In our problem, the model $\mathbf{m}$ is given by seismic velocities as a function of depth, so we use a parametrization of $\mathbf{v} = \mathbf{v}(z)$. This is a more general parametrization than $\mathbf{z} = \mathbf{z}(v)$ (i.e. depth as a function of velocity), since it includes the possibility of interfaces and of low-velocity zones. It is common practice to choose the *a priori* joint pdf (probability density function) to be uniformly distributed so as to be completely unbiased towards the solution (Box & Tiao 1973). The main objective of our formulation is that the *a posteriori* pdf should be different from the *a priori* pdf, thus ensuring that we have learned something from the data via the likelihood function $l(\mathbf{m}\,|\,\mathbf{d})$. Therefore it is always a necessary step to compare the two pdfs after the inversion has been completed.

A feature of Bayes' rule is its flexibility to include *a priori* information in the system. We introduce some *a priori* information by putting constraints on the models that are created. In order to avoid very unlikely models, we put two restrictions on the created models, namely velocity bounds and bounds on the gradient of the velocity (slope bounds), so

$$p(\mathbf{m}) = \begin{cases} \text{constant} & \text{for } \mathbf{v}_{\min} \le \mathbf{v} \le \mathbf{v}_{\max} \text{ and } \mathbf{s}_{\min} \le \mathbf{s} \le \mathbf{s}_{\max}, \\ 0 & \text{otherwise}, \end{cases} \tag{2}$$

where $\mathbf{v}$ denotes the velocities, and $\mathbf{s}$ denotes the slopes. Although the *a priori* joint pdf is uniformly distributed, the slope bounds bias us towards certain models, making the velocities correlated, causing the *a priori marginal* pdfs to be non-uniform.

We can also apply restrictions to the data, and thus to the likelihood function $l(\mathbf{m}\,|\,\mathbf{d})$. We make use of two likelihood functions, one for traveltime fitting, and one for waveform fitting. When we consider traveltime errors, we have the likelihood function

$$l_1(\mathbf{m}\,|\,\mathbf{T}) = a\phi_{\mathrm{tt}}(\mathbf{m}, \mathbf{d}) + b, \tag{3}$$

where the data is given by traveltimes $\mathbf{T}$; ϕ_{tt} is the traveltime misfit; a and b are constants to make $l_1(\mathbf{m}\,|\,\mathbf{T})$ a more suitable probability function,

$$a = \frac{-2}{(\phi_{\mathrm{tt,max}} - \phi_{\mathrm{tt,min}})^2}, \tag{4}$$

$$b = \frac{2\phi_{\mathrm{tt,max}}}{(\phi_{\mathrm{tt,max}} - \phi_{\mathrm{tt,min}})^2}; \tag{5}$$

and $\phi_{\mathrm{tt,max}}$ and $\phi_{\mathrm{tt,min}}$ are the maximum and minimum misfits for the traveltimes, respectively. We choose a linear distribution for the probability of the data, given the model, since this is what is used in the GA.

When we consider waveform errors, we wish to exclude all models that have a poor traveltime fit; this saves us computing

synthetic seismograms for unlikely models. Hence we use the likelihood function l_2:

$$l_2(\mathbf{m}\,|\,\mathbf{W}) = \begin{cases} a\phi_{\mathrm{wf}}(\mathbf{m}\,|\,\mathbf{d}) + b & \text{for } |\mathbf{T}(\mathbf{m}) - \mathbf{T}(\mathbf{d})| \le \Delta\mathbf{T}, \\ 0 & \text{otherwise}, \end{cases} \tag{6}$$

in which the data are now full waveforms, denoted by **W**; a and b are the constants as defined by eqs (4) and (5), but for waveforms; ϕ_{wf} is the waveform misfit; **T** are the traveltimes, where the traveltimes are calculated for the model, but picked for the data; and $\Delta\mathbf{T}$ denotes the traveltime bounds.

GENETIC ALGORITHMS AS A SEARCH METHOD

In this section we discuss briefly the non-linear GA optimization method. Since the GA method has been described several times in the geophysical literature (see, for example, Sen & Stoffa 1995), we discuss the method only briefly. For a full description of the method, the reader is referred to the excellent book by Goldberg (1989). The method is analogous to the way genetics works in nature. Our model parameters are translated into bitstreams and reproductions, cross-overs and mutations are used to modify these bitstreams. The first step, reproduction, takes a pool of models represented by bitstreams and gives each model a fitness value based on its misfit. These fitness values are translated into probability density functions, where good models have a higher probability than poor models. We then create a random number and when it falls within the probability interval of the model, we let the model go through. We continue this process until we have the same number of models as we started off with. We use a linear scaling function to transform the fitness values into pdfs. When using GA in combination with SA, the pdf is chosen to be an exponential function.

The next step is cross-over. For this, we choose pairs randomly. Then we randomly choose a cross-over point for each pair. The ends of the models in each pair are then swapped.

Finally, we perform mutations. Occasionally, we change one bit, randomly. This ensures that we keep some freshness, some randomness, in our pool, and prevents the GA from premature convergence. The chance that a bit is changed is called the mutation rate. This procedure of reproduction, choosing pairs, cross-over and mutation completes one complete iteration of the GA. The complete sequence of a GA is given in Fig. 1.

The reason why GAs work is not well understood, primarily because we have little statistical control over the process. However one of the reasons may be that the GA favours small building blocks that keep on returning in the new pools and are thus associated with good models. These building blocks are exploited and are combined to give good models.

An important aspect of the GA is the implicit parallelism. We work with many models simultaneously, and information from one model can be transferred to others. This ability to communicate between models is a special characteristic of the GA, and makes it very powerful.

STOPPING CRITERION FOR GAs USING BAYESIAN STATISTICS

There are several ways to compute the pdfs required for Bayes' rule. One way is to use simple Monte Carlo techniques, where

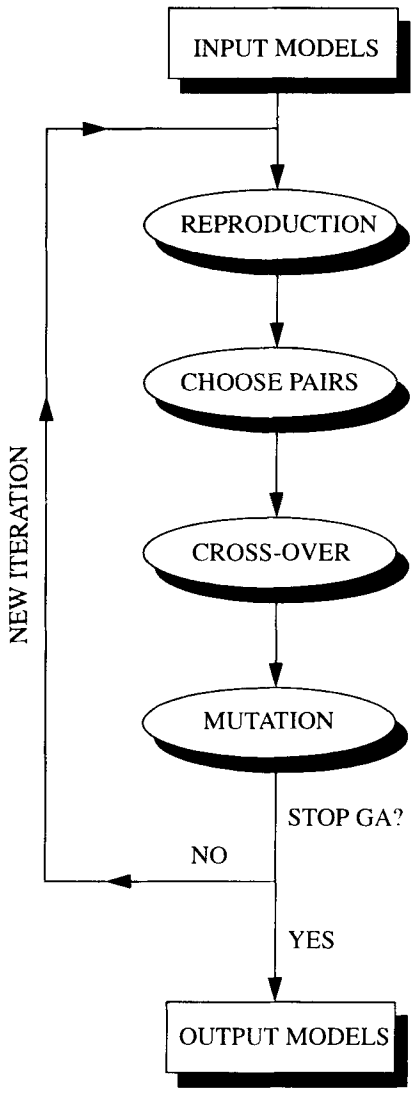

Figure 1. Flow chart of the genetic algorithm.

we continue searching in a random fashion (Cary 1987; Cary & Chapman 1988a,b). The GA is better than this because it samples model space in such a fashion that the more probable models are sampled more frequently than the less probable ones.

In a simple Monte Carlo method, we continue creating models with the same *a priori* knowledge and with the same likelihood function, since the traveltime bounds do not change during the search. This is different with GAs. In the reproduction step we apply linear scaling, so we apply the likelihood function l_1 or l_2. We then obtain the *a posteriori* pdf after reproduction:

$$\rho_{\mathrm{repro}}(\mathbf{m}\,|\,\mathbf{d}) = a\phi(\mathbf{m}, \mathbf{d}) + b$$

$$\text{for } \mathbf{v}_{\min} \le \mathbf{v} \le \mathbf{v}_{\max}, \quad \mathbf{s}_{\min} \le \mathbf{s} \le \mathbf{s}_{\max}, \quad (|\mathbf{T}(\mathbf{m}) - \mathbf{T}(\mathbf{d})| \le \Delta\mathbf{T} \quad \text{if } \phi = \phi_{\mathrm{wf}}) \tag{7}$$

and zero otherwise. For each iteration in a GA, $\phi_{\min}$ and $\phi_{\max}$ are taken as the minimum and maximum fitnesses of the present pool, not of the overall minimum and maximum

(although we usually ensure that the overall minimum remains present in the pool). This means that for each new iteration, a and b will also be different.

In the next step of cross-over there is little control on the statistical behaviour of the GA. Cross-over connects some parts of the model to other parts of the model, and thus transfers information; how this is taking place in statistical terms, is not at all clear. For cross-over the *a priori* pdf is now given by the *a posteriori* pdf of the reproduction, and the *a posteriori* pdf of the cross-over can be given only in general terms:

$$\rho_{\text{cross}}(\mathbf{m}\,|\,\mathbf{d}) = F(a\phi(\mathbf{m},\mathbf{d}) + b)$$
$$\text{for } \mathbf{v}_{\min} \le \mathbf{v} \le \mathbf{v}_{\max},$$
$$\mathbf{s}_{\min} \le \mathbf{s} \le \mathbf{s}_{\max}, \tag{8}$$
$$(|\mathbf{T}(\mathbf{m}) - \mathbf{T}(\mathbf{d})| \le \Delta\mathbf{T} \qquad \text{if } \phi = \phi_{\text{wf}})$$

and zero otherwise; in this expression F is some function, which cannot be described well, because of lack of understanding of this step.

We omit the influence of the final step of mutation since it is of minor importance, especially in statistical terms. So, with the next iteration, we start with this *a posteriori* pdf as the *a priori* distribution of the model.

An important use of the *a posteriori* pdf is to decide when to stop a search. In simple Monte Carlo methods, it is possible to devise a stopping criterion for the search, using the fact that, over a large number of experiments, the stochastic variables $\mathbf{m}$ approach a normal distribution. For large numbers the behaviour of the integral can be estimated in order to make an estimate of the integral (Tarantola 1987; Cary 1987). For the GA the situation is different, because with each iteration the distribution changes. There are two changes, one due to reproduction and the other due to crossover, expressed by the unknown function F above. The cross-over changes the overall statistical behaviour and introduces new models.

In order to devise a stopping criterion we introduce the marginal *a posteriori* pdfs for each model parameter m_j:

$$\rho(m_j\,|\,\mathbf{d}) = \int \rho(\mathbf{m}\,|\,\mathbf{d})\, dm_1\, dm_2 \ldots dm_{j-1}\, dm_{j+1} \ldots dm_J \tag{9}$$

where the subscript j refers to the parameter within the model, and J is the total number of parameters. We assume that each of these distributions, and therefore each marginal *a posteriori* pdf for each model parameter m_j (for example one velocity at one depth), approaches the normal distribution. We denote the running average of each model parameter by a bar above it, i.e.

$$\bar{m}_j = \frac{1}{N}\sum_{n=1}^{N} m_{nj}, \tag{10}$$

where m_{nj} denotes the jth parameter of the nth model, and N is the number of models taken into account. According to the mean-value theorem, for N experiments the random variable $x_j = (\bar{m}_j - E(m_j))\sqrt{N/\sigma^2}$ approaches the normal distribution when $E(m_j)$ is the expectation of m_j and σ^2 its variance. Then we can express the probability that x_j is less than $x_j^{\max}$ by

$$\Pr(x_j \le x_j^{\max}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_j^{\max}} \exp(-y^2/2)\, dy. \tag{11}$$

We can use this characteristic to put an upper bound on the relative error of the average of each model parameter (see Tarantola 1987). For large N, we have the probability that

$$\frac{\bar{m}_j - E(m_j)}{E(m_j)} = \frac{x_j^{\max}}{\bar{m}_j}\sqrt{\frac{s^2}{N}} \tag{12}$$

is given by $2\Pr(x_j^{\max}) - 1$. Here we assume that we can estimate the expectation $E(m_j)$ by the running average $\bar{m}_j$, and σ^2 by s^2:

$$s^2 = \frac{N}{N-1}\left(\frac{1}{N}\sum_{n=1}^{N} m_{nj}^2 - \bar{m}_j^2\right). \tag{13}$$

So the procedure we follow is: set the tolerable relative error to some value, which then determines the value $x_j^{\max}$ via the above integral (11); then estimate the expectation and the variance by $\bar{m}_j$ and s^2 as above; finally stop the GA search when

$$(x_j^{\max}/\bar{x}_j)\sqrt{s^2/N} < 2[1 - \Pr(x_j^{\max})].$$

Since this discussion is for each parameter of the model, we take the error $e_{\max}$ of the maximum of all these parameters, i.e.

$$e_{\max} = \max(e_1, e_2, \ldots, e_j), \tag{14}$$

where J is, again, the number of parameters in a model.

THE MISFIT FUNCTIONS

We define a misfit function which reflects whether a model is good or bad. In our problem, we apply the optimization twice, first on traveltimes, and then on waveforms. For the traveltime misfit ϕ_{tt}, we use the normalized squared error:

$$\phi_{\text{tt}} = \sum_{i=1}^{I} \frac{(T_i(\mathbf{d}) - T_i(\mathbf{m}))^2}{(\Delta T_i)^2}, \tag{15}$$

in which the summation is over the number of arrivals (I), and the normalization is with respect to the traveltime bounds as given in eq. (6).

For the waveform misfit we also use the squared error, normalized by the data seismograms:

$$\phi_{\text{wf}} = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{I}\sum_{i=1}^{I} \frac{\{\text{data}(t_i, x_j) - C\,\text{synth}(t_i, x_j)\}^2}{\{\text{data}(t_i, x_j)\}^2}, \tag{16}$$

where the first summation is over the ranges x_j of the seismograms, and the second summation is over the time samples t_i; C is the overall scaling factor, given by

$$C = \frac{1}{I}\sum_{i=1}^{I} \frac{\text{Env}_i^{\text{d}}(t_k)}{\max\{\text{Env}_i^{\text{s}}(t)\}}, \tag{17}$$

where the time t_k is the time where the envelope function Env has its maximum, i.e.

$$\text{Env}^{\text{s}}(t_k) = \max\{\text{Env}^{\text{s}}(t)\}. \tag{18}$$

Again, the summation is over all arrivals. Note that the scaling factor C which matches the data and synthetics is constant for all arrivals. The factor C is chosen as defined here because the misfit function tends to scale down the synthetics when C is chosen simply as an energy scaling factor (see Cary 1987).

MODELLING ASPECTS

Cary & Chapman (1988a,b) used layers with linear velocity gradients in their models. However, where there is an abrupt

change in gradient at a layer boundary, features are introduced. These may not necessarily be intended, or required by the data. When many linear gradients are introduced, there will be as many features. It is therefore necessary to define the model such that the model discretization does not introduce artefacts, particularly in the waveforms.

We therefore match our model parametrization to the synthetic seismogram method. For Chapman's method, the high-frequency approximation in the mixed space–slowness domain is used. The method uses the local ray attributes (angle of ray, amplitude and geometrical spreading of that ray) from asymptotic ray theory (ART) in the space–frequency domain. This

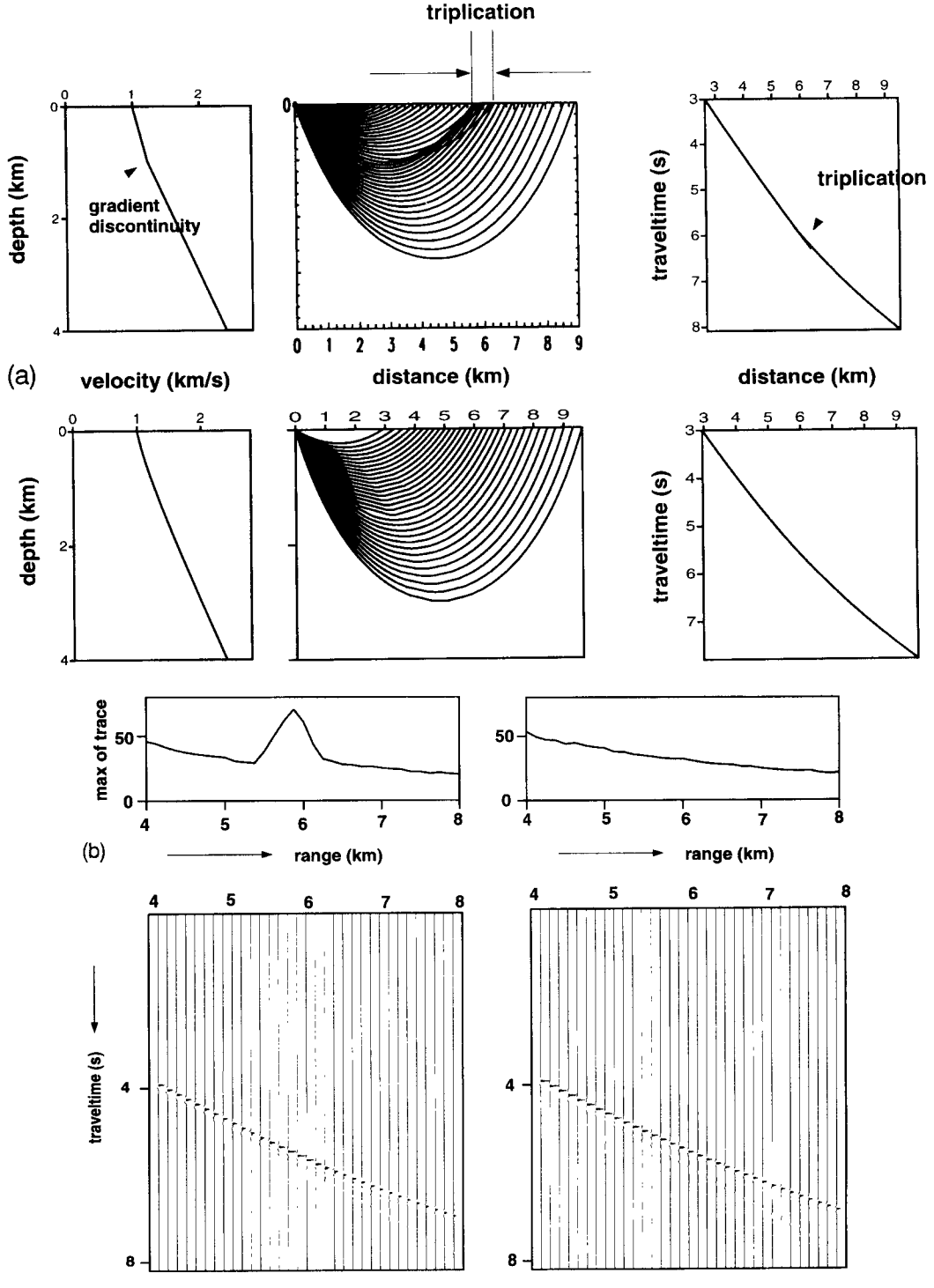

Figure 2. Effect of an increase in velocity gradient causing a triplication zone: (a) velocity model, ray plot and traveltime curve for a model with linear velocity gradients (upper) and B-spline-interpolated velocity function (lower); (b) seismograms for linear velocity gradients (left) and for B-spline-interpolated velocity model (right), together with the maximum amplitude of each trace on top.

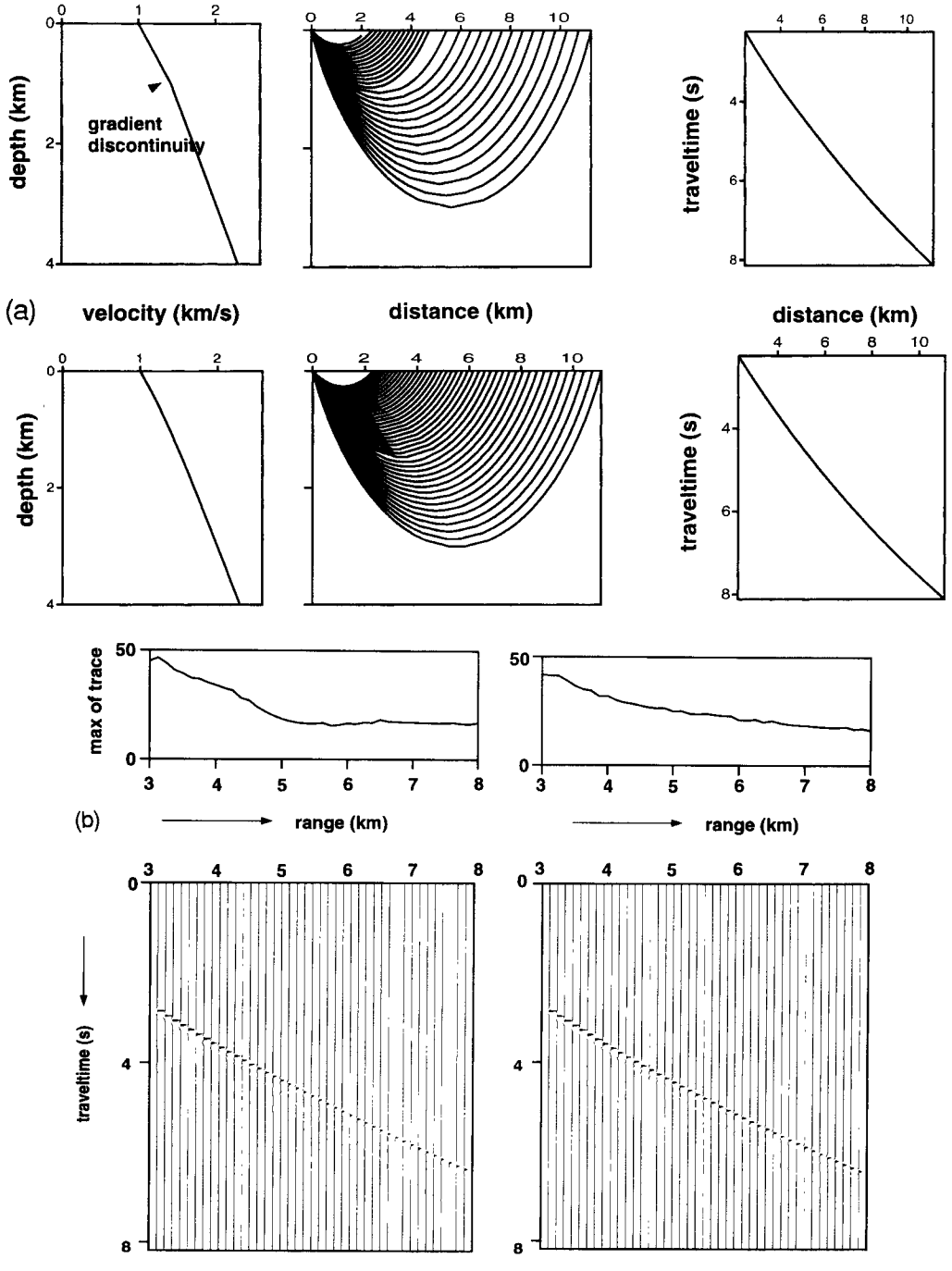

Figure 3. Effect of a decrease in velocity gradient: (a) velocity model, ray plot and traveltime curve for model with linear velocity gradients (upper) and B-spline-interpolated velocity function (lower); (b) seismograms for linear velocity gradients (left) and for B-spline-interpolated velocity model (right), together with the maximum amplitude of each trace on top.

local ray information is globalized and combined in Chapman's method to form the seismogram. The high-frequency approximation requires the model to be smooth, so we must ensure that the angular frequency ω satisfies

$$\omega \gg v\left(\frac{\nabla^2 A_0}{A_0}\right)^{1/2}, \tag{19}$$

where v is the velocity and A_0 the amplitude of the local plane wave (Kravtsov & Orlov 1990). Within one layer with a linear velocity gradient this condition is satisfied, but a discontinuity in velocity gradient at the base of the layer introduces a response. The minimum order of polynomial that does not introduce any artefacts such as these is a polynomial that is continuous up to its second derivative, namely a cubic spline. There are also other conditions that apply to ART; for these the reader is referred to Ben-Menahem & Beydoun (1985), Beydoun & Ben-Menahem (1985), Beydoun & Keho (1987) and Kravtsov & Orlov (1990).

To demonstrate the effect discussed above, we show two models in Figs 2 and 3. First consider an abrupt change in

linear velocity gradient, as depicted in the upper plots of Fig. 2(a). As the ray paths show, this produces a triplication, which is also apparent in the traveltime curve shown in the upper right-hand panel. When we compare this with results from the B-spline-interpolated velocity function, which are shown in the bottom plots of Fig. 2(a), we see that the effect on amplitudes of the seismograms is large, as shown in Fig. 2(b). On the left we show the seismograms for a model with linear velocity gradients, while on the right we show the seismograms for the B-spline-interpolated velocity function; the seismograms are calculated with Chapman's method. On the seismograms on the left the amplitudes are larger in the area of the triplication. In order to demonstrate this effect more clearly, we have plotted the maximum amplitude of each trace in the diagrams at the top of Fig. 2(b). This demonstrates that the discontinuity in velocity gradient produces a large peak in seismic amplitudes. For the B-spline-interpolated velocity function, the amplitudes behave smoothly.

There is also an effect when the linear-velocity-gradient model decreases abruptly (Fig. 3). In the ray picture we see that the ray density changes abruptly near a range of 5 km. In the traveltimes no marked effect can be seen (Fig. 3a). On the seismograms, a small effect is apparent (Fig. 3b). The amplitudes for the smooth-velocity model do not change as much as for the linear velocity gradients, as can be seen in the plot of maximum amplitudes shown above the seismograms. However this effect is much smaller than for the case of an abrupt increase in velocity gradient shown in Fig. 2.

A disadvantage of using cubic splines is that there is no analytical expression available for the traveltime and range integrals, and therefore those integrals have to be determined numerically. This takes considerable computing time when we are dealing with turning rays, since the depth integral for the range has a singularity (which is integrable) at the end point. For example, to calculate the response of a typical model with linear velocity gradients takes about 0.5 s per model, while with cubic B-splines this increases to about 30 s on a Convex C1 computer. However, since we can separate the traveltime and waveform inversions, and only the waveforms are affected greatly by caustics, we can use linear velocity gradients for the traveltime inversion and switch over to cubic-spline velocity functions for the waveform inversion.

DATA: FIELD EXPERIMENT AND PROCESSING

The seismic data were collected in 1987 November and December in the vicinity of the Blake Spur fracture zone, south-west of the Bermuda Islands (Fig. 4). Results of manual trial-and-error forward modelling are given by Minshull *et al.* (1991). Wide-angle seismic data were collected using an expanding-spread technique (Stoffa & Buhl 1979) with the shooting and receiving ships sailing in opposite directions. The shooting ship was equipped with a tuned 10-airgun array, and the receiving ship towed a 2.4 km, 48-channel streamer. During processing the traces were stacked and static corrections were applied for sea-floor topography. The field data were recorded digitally and sampled every 4 ms, then resampled at 8 ms after applying a trapezoidal band-pass filter with cutoffs at 3–6 Hz and 35–45 Hz. The same filter is applied to the synthetics. Both the synthetics and the data are smoothed with a $\mathrm{sinc}^4(f/f_N)$ function where f denotes the frequency and $f_N = (2\Delta t)^{-1}$ is the Nyquist frequency. This is required by the synthetics since they otherwise show singularities at the geometrical arrival time. Range-dependent weights are applied to both the data and the synthetics in order to ensure that the high amplitudes at near ranges do not dominate over smaller amplitudes at far ranges. To achieve this we apply a scaling factor of $x^{-1/2}$ in which x denotes the range.

We have chosen two typical lines to model. One (ESP 2, Fig. 5) is away from the fracture zone, where both refractions are prominent in the observed seismograms. The other (ESP 6, Fig. 6) is above the Blake Spur fracture zone, where only refractions can be discerned clearly. For both profiles, we used the inbound leg where the ships were sailing towards each other.

RESULTS FOR ESP 2 (INBOUND LEG)

ESP 2 was shot outside the fracture zone, and a reflection off the Moho can be discerned clearly at the far ranges (phase labelled 6 in Fig. 5; 30 km to 60 km). At nearer ranges, both *P*-wave (phases 1–5) and doubly converted *S*-wave (phases 7–10) refractions can be seen. For our analysis we assume that the velocity structure is laterally homogeneous. We first

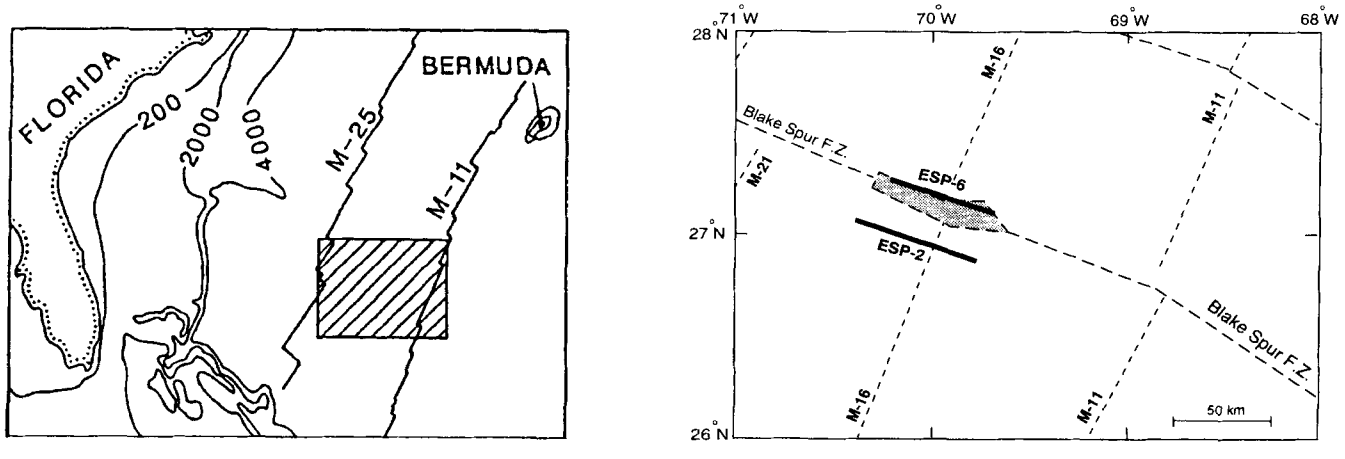

Figure 4. Location map for seismic profiles (after Minshull *et al.* 1991).

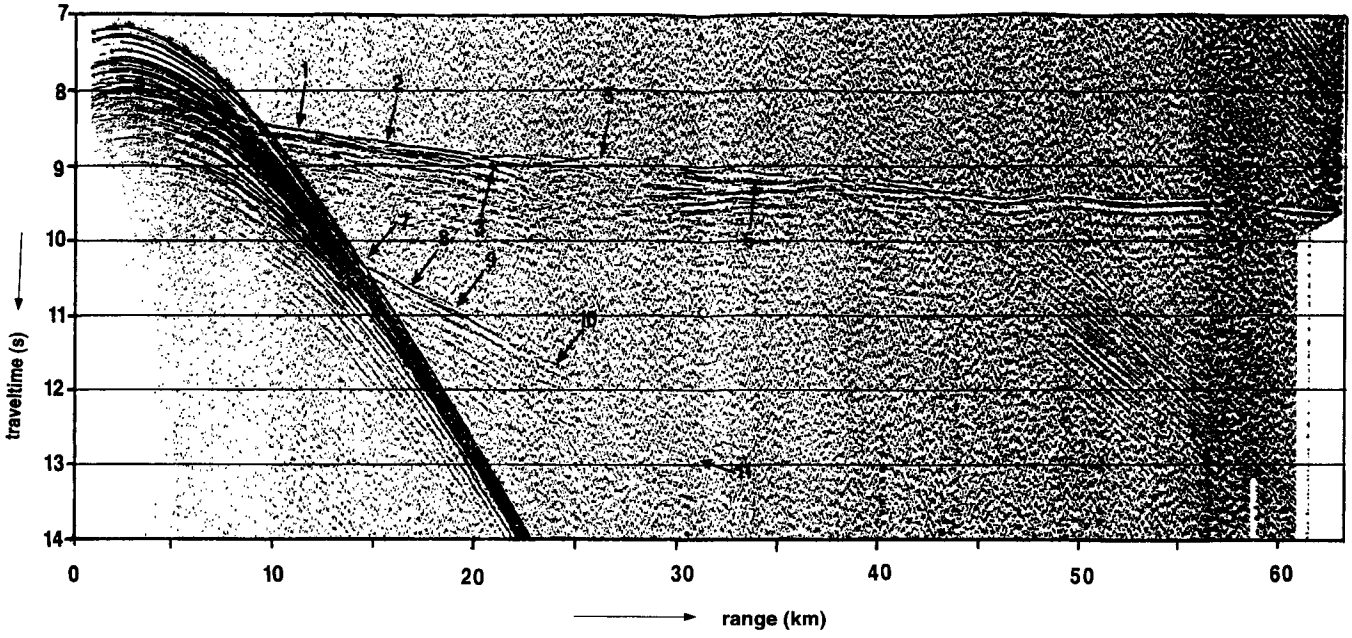

Figure 5. Original data ESP 2 inbound (after Minshull *et al.* 1991). The traveltimes are reduced with a velocity of 8 km s^{-1}. Numbered arrows refer to phases used by Minshull *et al.* (1991) in their interpretation.

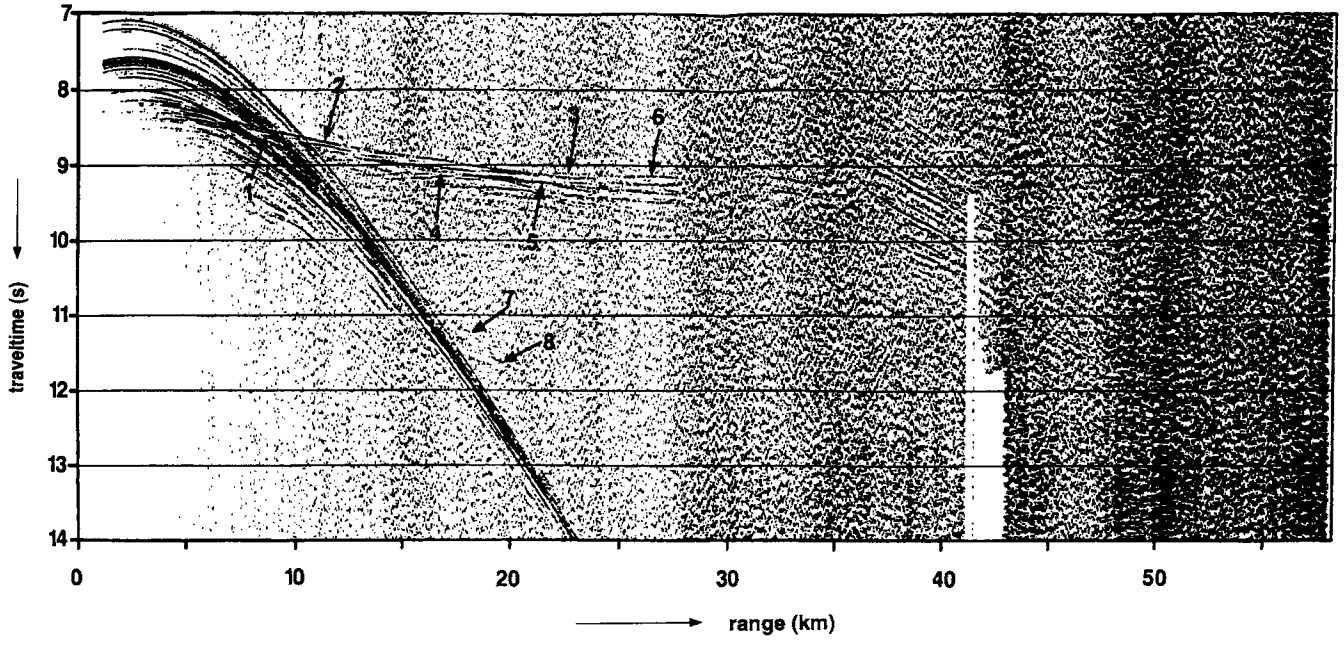

Figure 6. Original data ESP 6 inbound (after Minshull *et. al* 1991). The traveltimes are reduced with a velocity of 8 km s^{-1}. Numbered arrows refer to phases used by Minshull *et al.* (1991) in their interpretation.

consider *P*-wave data, using 19 traces with clear *P*-wave refractions and with traveltimes picked by hand. We derive a source waveform from the recorded ocean-bottom reflection, which also includes near-surface reflections and near-bottom effects. This 'source' function is therefore considered a good representation of the effective source.

In the definition of the models we have to set bounds on acceptable velocities. Usually a regional model is taken and variation in this model is allowed by including wide bounds around it. The interval between each minimum and maximum bound is divided into equally spaced points. Each node is then assigned a bit representation for input to the GA (Sambridge & Drijkoningen 1992). In our implementation, we also made the depths at which velocities are specified variable. In Cary's original work, these depths were fixed, except at the interfaces, so as to reduce computing time with the Monte Carlo method. This could cause bias in the acceptable models, so in our work we have allowed depths as well as velocities to vary.

We performed the GA search with 10 depth–velocity pairs as variables. The pool was chosen to consist of 50 models, where each model contained 82 bits; the mutation rate was set to 0.025, so every bit had this chance of being swapped. The

relative error was set at 0.01, which was reached by the traveltime GA. This caused the GA to continue for 40 iterations, and finally 795 models passed the traveltime test. The best traveltime misfit was $\phi_{tt} = 0.0221$.

For the best 50 models passing the traveltime test, the program automatically switches to waveform fitting, and the velocity functions are then B-spline-interpolated functions, where the model points serve as control nodes of the splines. We used the same GA parameters as before, but allowed a maximum of 25 iterations for the GA because of computer-time restrictions. The 25-iteration limit was reached before the error limit of 0.01 was achieved. The final misfit value was $\phi_{wf} = 0.835$.

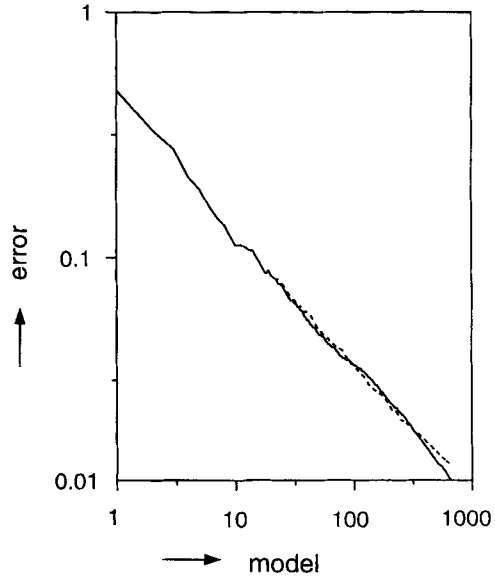

Figure 7. Error curves for a stopping criterion with a simple Monte Carlo technique (dashed line) and the genetic algorithm (solid line).

We have already mentioned the error criterion for the GA as used in the Monte Carlo technique. How does the error behave? In Fig. 7 we plot the logarithm of the error as a function of the logarithm of the number of models. The error estimate continues decreasing approximately linearly with increasing number of models. In Fig. 7 two curves are shown, one for the GA (solid line), the other for the same model bounds but using the simple Monte Carlo technique (dashed line).

Use of the error definition as described in the section on the stopping criterion seems to solve a problem with the stopping of GAs. Stoffa & Sen (1992) let the GA continue until the covariances had stabilized. Sambridge & Drijkoningen (1992) stopped the GA when a fixed number of iterations had been completed. The error curve can be used for a further purpose: from extrapolation of the curve an estimation can be made of how many iterations are needed in order to obtain a given error tolerance, and this information can, for instance, be used to estimate the computer time needed to achieve this.

The model with the best waveform misfit is plotted in Fig. 8(a), together with the associated synthetic seismograms in Fig. 8(b). At near ranges the fits are good, but they are poorer at farther ranges, where the amplitudes of the signal are smaller. The poorer control on the long-range arrivals leads to poorer control on the lower crustal part of the model: apart from this, the models from inversion and from forward modelling are very similar.

In Fig. 9 we show the (relative) errors from the GA inversions: the solid line shows those from traveltime fitting and the dashed line those from waveform fitting.

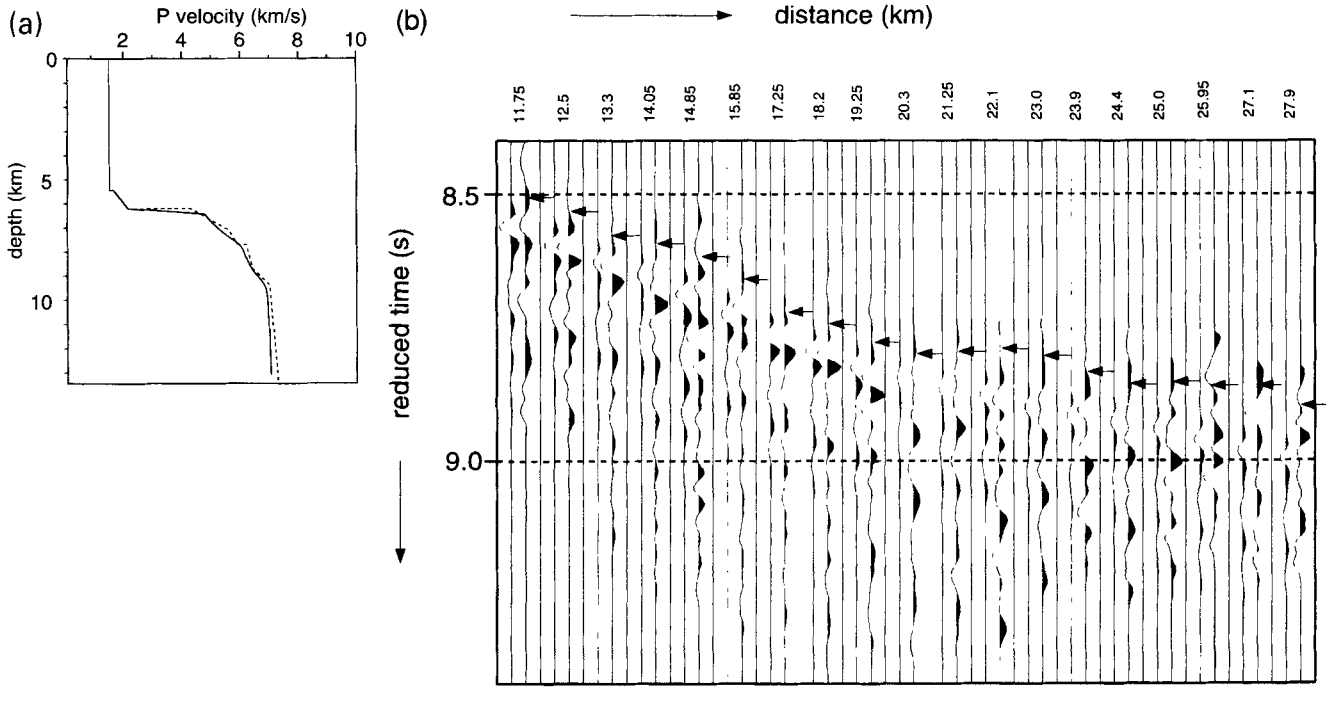

Figure 8. (a) Best model from inversion for *P*-wave refractions in ESP 2 (solid line), and from forward modelling by Minshull *et al.* (1991) (dashed line). (b) The associated seismograms: traces are shown in pairs of synthetic trace and observed data trace, each pair separated by a zero trace. Arrows indicate the hand-picked traveltimes; traveltimes are reduced with a reduction velocity of 8 km s^{-1}.

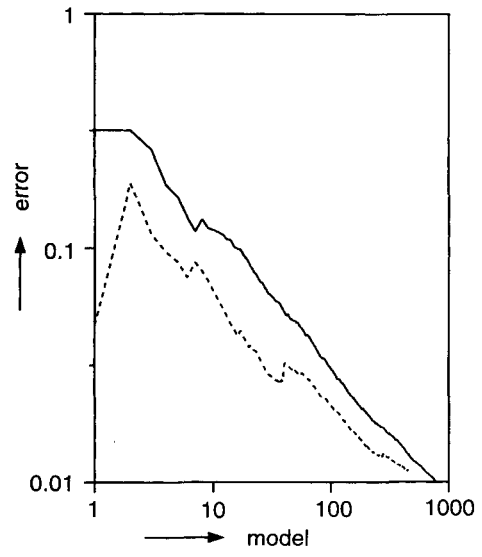

Figure 9. Error curves for fitting traveltimes (solid line) and waveforms (dashed line) of *P*-wave refraction data in ESP 2.

We will now concentrate on the shear-wave refractions on ESP 2. They can be observed at ranges of 15 to 25 km (Fig. 5). Only the refractions (phases 7–10, Fig. 5) are sufficiently clear to include in our GA search, although Minshull *et al.* (1991) also use a faint *SS* reflection (phase 11, Fig. 5). We assume that the conversion between shear and compressional energy occurs at the sediment–basement interface. We selected 10 traces for this GA search; 10 depth–velocity points were used to specify the model bounds. The pool consisted of 100 models, each with a bitstring length of 82 bits, and the mutation rate was set at 0.025. Again we allowed an error tolerance of 0.01, which required 54 iterations of the GA. The best traveltime misfit was $\phi_{tt} = 0.0427$.

After the traveltime GA with linear velocity gradients, we proceeded with B-spline-interpolated velocity functions for the calculation of amplitudes, and again a pool was chosen with 100 models, the best from 1193 models that passed the traveltime test. The best misfit obtained with the waveform fit was $\phi_{wf} = 0.778$ after 25 iterations. The results are shown in Fig. 10, which show the data and synthetic seismograms, and the associated model. The upper part of the model agrees well with that of Minshull *et al.* (1991); the lower part of the model shows less agreement, but this part of the model cannot be resolved by the traces included in the GA search. The relative errors for these runs are shown in Fig. 11: similar behaviour is observed to those from the *P*-wave inversions.

RESULTS FOR ESP 6 (INBOUND LEG)

ESP 6 lies above a fracture zone. At ranges of 10 to 35 km, the *P*-wave refractions (phases 2, 3, 6; Fig. 6) form prominent *P* arrivals in the seismograms. We picked 19 traces from these ranges and specified 10 velocity–depth pairs; the number of models in one pool was set at 100, and the mutation rate was set at 0.025. The relative error was set at 0.01, causing the GA to continue for 36 iterations. The best traveltime misfit was in this case $\phi_{tt} = 0.0645$.

We then switched to waveform fitting, and a pool of the best 100 models of the previous 1361 models passing the traveltime test were used as starting models. The GA continued for 25 iterations without reaching the error of 0.01. The best model had a waveform misfit of $\phi_{wf} = 1.0634$. We obtained the results shown in Fig. 12, which show the best model together with the associated synthetics. In Fig. 13 we show the curves for the relative error, which shows the same behaviour as for

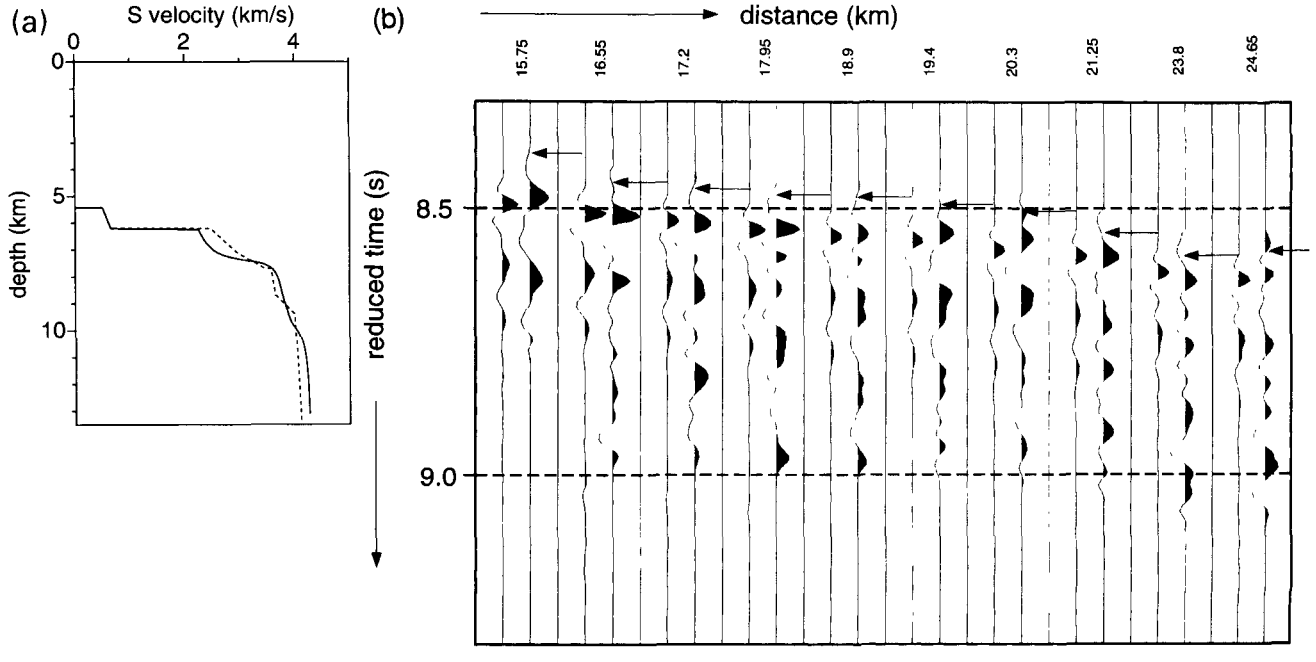

Figure 10. (a) Best model from inversion for *S*-wave refractions in ESP 2 (solid line), and from forward modelling by Minshull *et al.* (1991) (dashed line). (b) The associated seismograms: traces are shown in pairs of synthetic trace and observed data trace, each pair separated by a zero trace. Arrows indicate the hand-picked traveltimes; traveltimes are reduced with a reduction velocity of 4 km s^{-1}.

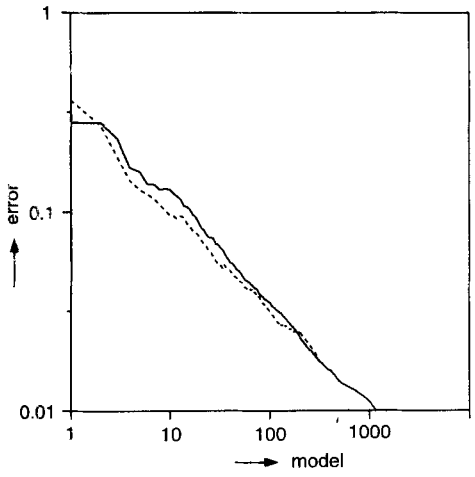

Figure 11. Error curves for fitting traveltimes (solid line) and waveforms (dashed line) of *S*-wave refraction data in ESP 2.

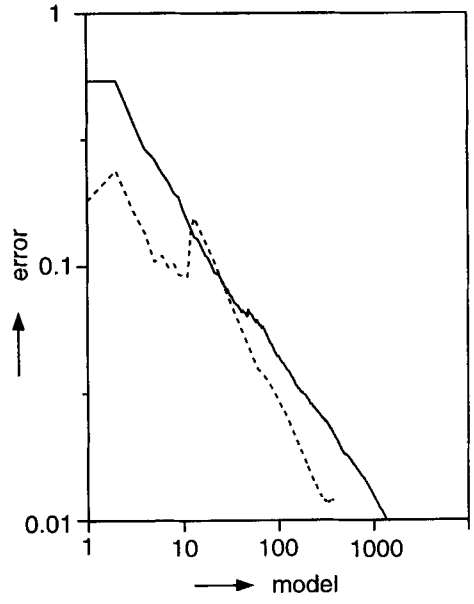

Figure 13. Error curves for fitting traveltimes (solid line) and waveforms (dotted line) of *P*-wave refraction data in ESP 6.

the other data—decreasing with increasing number of models. The best-fitting velocity–depth model from our inversion (solid line, Fig. 12a) is very similar to the best-fitting forward model of Minshull *et al.* (1991) (broken-line, Fig. 12a), except for the step in velocity at 10.5 km depth, Minshull *et al.* (1991) interpreted this as being due to a reflection from an interface in the lower crust (producing phase 5 on Fig. 6): our B-spline cannot reproduce steps in velocity if they are not explictly allowed for in the program (which is also the case for models with linear velocity gradients), although the inversion does indicate an inflection at this depth (Fig. 12a).

DISCUSSION AND CONCLUSIONS

We have demonstrated that inversion using the genetic algorithm provides as good a model as trial-and-error forward modelling when turning rays are considered. Future developments will extend the method to laterally inhomogeneous media. The inversion process using splines for the velocity–depth distribution provides a smooth velocity function, and

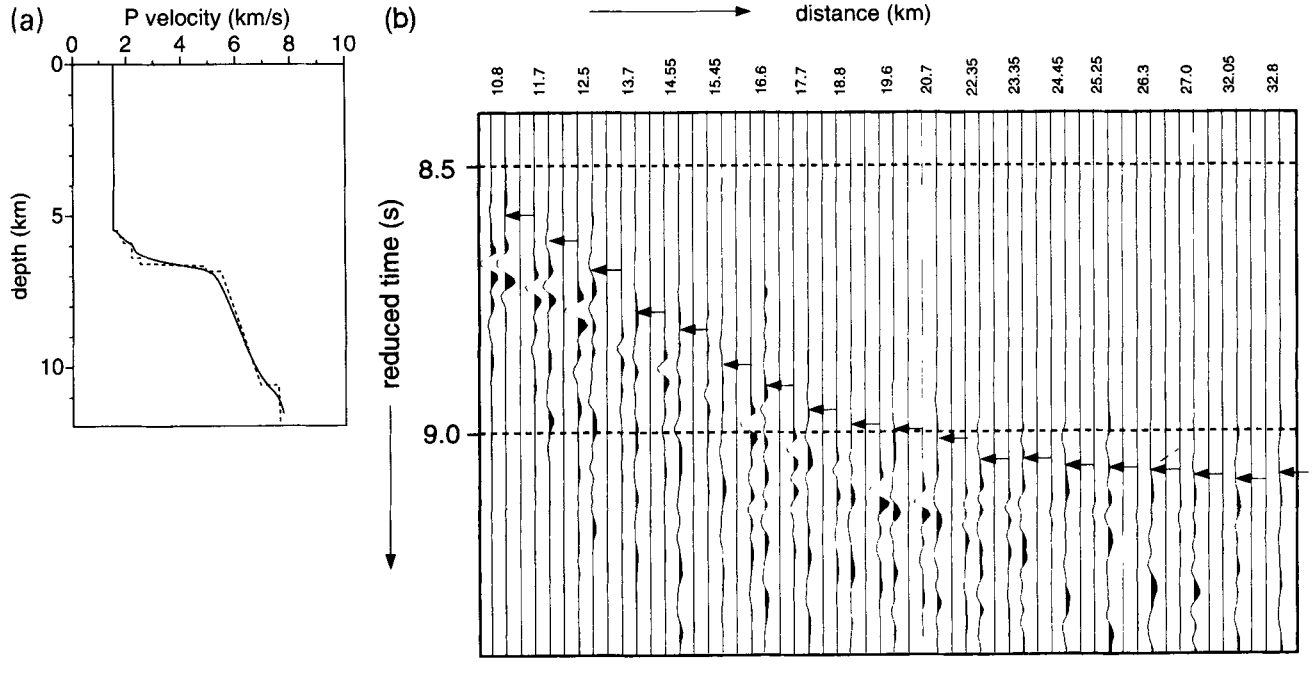

Figure 12. (a) Best model from inversion for *P*-wave refractions in ESP 6 (solid line), and from forward modelling by Minshull *et. al.* (1991) (dashed line). (b) The associated seismograms: traces are shown in pairs of synthetic trace and observed data trace, each pair separated by a zero trace. Arrows indicate the hand-picked traveltimes; traveltimes are reduced with a reduction velocity of 8 km s^{-1}.

the global minimum can be found consistently and robustly. In extending the modelling to laterally varying models, it is important to use smooth velocity models in order to allow traveltimes and amplitudes to be inverted robustly.

A more difficult problem in automatic inversion schemes is the identification and picking of arrivals. If these are done by hand, it is reasonably straightforward to identify the phases when the media are 1-D. However, if the structure varies laterally then it is not easy and sometimes not possible to identify unambiguously the paths that different arrivals have taken, particularly when different ray paths may generate crossing arrivals. Yet this is a crucial interpretive stage. The main use of inversion in laterally varying media may be the use of a high-speed inversion as a tool to aid the interpreter: if he or she is able to test rapidly a series of different interpretations with different identifications of the main arrivals, this may enable the interpreter to rule out geologically unlikely solutions, or at least to gain a feel for the range of interpretations that will satisfy the data.

(The reader can obtain the inversion code by e-mailing to: g.g.drijkoningen@ctg.tudelft.nl)

ACKNOWLEDGMENTS

The author GGD wishes to thank RSW for his one-year support while doing some of this work. He also wishes to thank the computer laboratory at Technical Geophysics for the computer time used to produce the results in this article. Dr T. A. Minshull from Cambridge University is gratefully acknowledged for his contributions to this paper. Dr Brokesova from Charles University, Prague, assisted in producing Fig. 2(a). This work was supported in part by the Natural Environment Research Council.

REFERENCES

Ben-Menahem, A. & Beydoun, W.B., 1985. Range of validity of seismic ray and beam methods in structures by general inhomogeneous media—I. General theory, *Geophys. J. R. astr. Soc.*, **82,** 207–234.

Beydoun, W.B. & Ben-Menahem, A., 1985. Range of validity of seismic ray and beam methods in structures by general inhomogeneous media—II. Canonical problem, *Geophys. J. R. astr. Soc.*, **82,** 235–262.

Beydoun, W.B. & Keho, T.H., 1987. The paraxial ray method, *Geophysics,* **52,** 1639–1653.

Box, G.E.P. & Tiao, G.C., 1987. *Bayesian inference in statistical analysis,* Addison-Wesley, New York, NY.

Cary, P.W., 1987. Waveform inversion of marine refracton data, *PhD thesis,* University of Cambridge, UK.

Cary, P.W. & Chapman, C.H., 1988a. Automatic 1D inversion of marine refraction data, *J. geophys. Res.*, **93,** 527–546.

Cary, P.W. & Chapman, C.H. 1988b. Waveform inversion of expanding spread profile 5 from the North Atlantic Transect, *J. geophys. Res.*, **93,** 13 575–13 588.

Chapman, C.H., 1978. A new method for computing synthetic seismograms, *Geophys. J. R. astr. Soc.*, **54,** 481–518.

Chapman, C.H. & Drummond, R., 1982. Body-wave seismograms in inhomogeneous media using Maslov asymptotic theory, *Bull. seism. Soc. Am.*, **72,** S277–S317.

Goldberg, D.E., 1989. *Genetic algorithms in search, optimization and machine learning,* Addision-Wesley, Reading, MA.

Hammersley, J.M. & Handscomb, J.M., 1964. *Monte Carlo Methods,* Methuen and Co. Ltd, London.

Ingber, L. & Rosen, B., 1992. Genetic algorithms and very fast simulated reannealing: a comparison, *Math. Comput. Modelling,* **16,** 87–100.

Jervis, M., Sen, M.K. & Stoffa, P.L., 1993. Optimization methods for 2D pre-stack migration velocity estimation, *Proc. 63rd Ann. Int. Meet., Soc. Expl. Geophys.,* pp. 965–968, SEG, Tulsa, OK.

Kravtsov, Y.A. & Orlov, Y.I., 1990. *Geometrical Optics of Inhomogeneous Media,* Springer-Verlag, Heidelberg.

Minshull, T.A., White, R.S., Mutter, J.C., Buhl, P., Detrick, R.S., Williams, C.A. & Morris, E., 1991. Crustal structure at the Blake Spur fracture zone from expanding spread seismic profiles, *J. geophys. Res.*, **96,** 9955–9984.

Sambridge, M.S. & Drijkoningen, G.G., 1992. Genetic algorithms in seismic waveform inversion, *Geophys. J. Int.*, **109,** 323–342.

Sen, M.K., & Stoffa, P.L., 1995. *Global Optimization Methods in Geophysical Inversion,* Advances in Exploration Geophysics series, Elseiver, Amsterdam, submitted.

Stoffa, P.L. & Buhl, P., 1979. Two ship multichannel experiments for deep crustal studies: Expanded spread and constant offset profiles, *J. geophys. Res.*, **84,** 7645–7660.

Stoffa, P.L. & Sen, M.K., 1992. Seismic waveform inversion using global optimization, *J. Seismic Expl.*, **1,** 9–27.

Stoffa, P.L., Sen, M.K., Varela, C. & Chunduru, R.K., 1994. Geophysical applications of global optimization methods, *Proc. 56th Ann. Meet., European Assoc. Expl. Geophys.,* EAEG, Zeisl, The Netherlands.

Tarantola, A., 1987. *Inverse Problem Theory,* Elsevier, Amsterdam.

Varela, C.L., Stoffa, P.L. & Sen, M.K., 1994. Migration misfit and reflection tomography: Criteria for pre-stack migration velocity estimation in laterally varying media, *Proc. 64th Ann. Int. Meet., Soc. Explor. Geophys.*