A COMPARISON OF GLOBAL OPTIMISATION METHODS FOR NEAR-OFFSET VSP INVERSION*

STEVE HORNE†, and COLIN MACBETH

Edinburgh Anisotropy Project, British Geological Survey, West Mains Road, Edinburgh, U.K.

Abstract-Three global optimisation algorithms are applied to the problem of geophysical inversion. We describe and test the methods of Tabu Search, Genetic Algorithms and Simulated Annealing. These techniques are used to invert observations of shear-wave splitting from near-offset Vertical Seismic Profiles. Each search shows distinct advantages and disadvantages so that no particular algorithm can be clearly recommended. Nonetheless, we can recommend that a global optimisation be followed by a local search. © 1998 Elsevier Science Ltd. All rights reserved

Key Words: Nonlinear inversion, Tabu, Genetic algorithm, Simulated annealing, Near-offset VSP, Shear-wave anisotropy.

INTRODUCTION

Inversion is usually cast into an optimisation problem whereby the differences between observed data and synthetic data are minimised as a function of some model parameters. Thus inversion can be seen to be a search over a range of possible models. Traditional search methods typically use gradient based methods which generally follow a path leading directly to the nearest minimum. The problem with such approaches is that the final result can represent a local minimum which is dependent upon the starting model and is not the best overall solution. It is this presence of local minima in the model space which impedes many search strategies. Ideally, a systematic search through the entire model space yielding all minima, including the global minimum, is desirable. Unfortunately, such methods must be abandoned when faced with realistic inversion problems due to the vast number of models to be tested. Instead efficient search engines able to locate the significant solutions in complicated model spaces are desirable. In addition to efficiency we also require flexibility, robustness and reliability when solving inversion problems. Several new techniques claim to possess such characteristics. In this paper we test a few of these methods known as tabu search, genetic algorithms and adaptive simulated annealing. These techniques are applied to the problem of inverting shear-wave birefringence measurements obtained from two near offset VSPs.

NEAR-OFFSET VSPS

The data that we invert are obtained from the analysis of two multicomponent near-offset VSPs conducted at the Conoco Borehole Test Facility, Oklahoma. These VSPs are referred to as the Peel VSP and the 33-1 VSP, indicating the wells about which the experiments were shot. The range of geophone depths and source offsets are similar for the two VSPs but the azimuthal locations of the source are very different (Fig. 1). In the case of the 33-1 VSP the sources were located to the west of well 33-1 (N279°E) and towards the south east (N122°E) for the Peel VSP. Vibroseis sources were used and activated in in-line, cross-line and vertical orientations. Three component recordings were obtained at regular depth intervals of 15 m. The seismic data are processed using rotation to compensate for downhole sonde rotation, $f-k$ filtering to separate up- and down-going events and near-surface correction to rectify any lateral variations in the near surface. The seismic data is then processed to obtain shear-wave birefringence measurements (Zeng and MacBeth, 1993). These take the form of a time delay measuring the arrival time difference between the fast and slow split shear-wave modes, and the polarisation direction of the fast split shear-wave. From these shear-wave measurements we construct a data vector which we invert by the application of the various optimisation methods under test.

The objective function that we use is a least squares quantity defined as

$$f(\mathbf{m}) = \frac{1}{2N} \sum \left(\frac{\tau_0 - \tau_{\mathrm{m}}}{\Delta \tau} \right)^2 + \left(\frac{p_0 - p_{\mathrm{m}}}{\Delta p} \right)^2$$

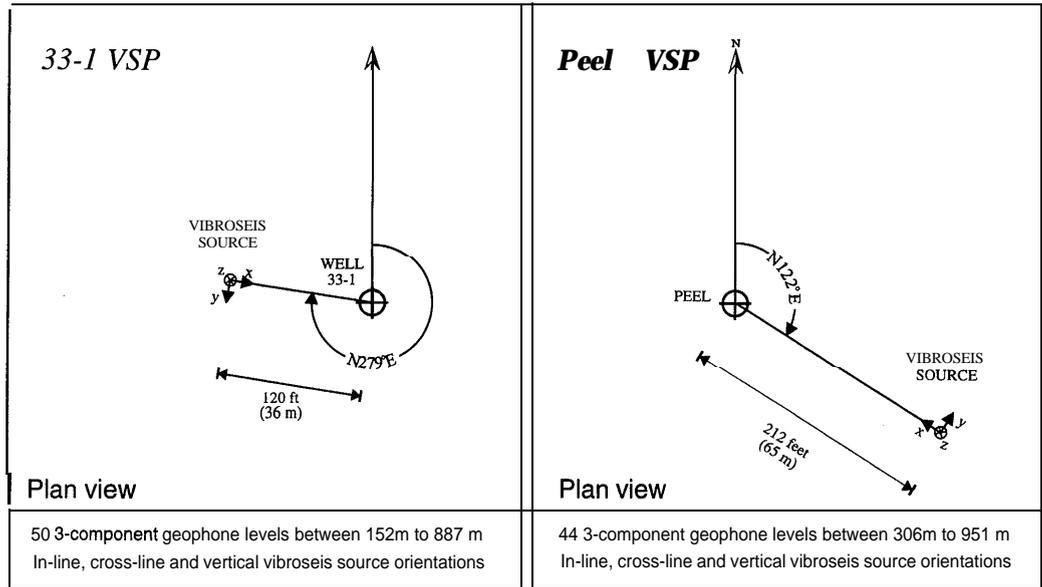where τ_0 and p_0 are the observed time delays and

Figure 1. Summary of near-offset VSP geometries

polarisations, τ_m and p_m are the modelled time delays and polarisations and $\Delta\tau$ and Δp are the estimated errors associated with the observed time delays and polarisations. The model vector, m, describes five anisotropic zones which we identify from the data. We choose to represent each zone in terms of three crack parameters which are the crack density (CD), the aspect ratio (AR) and the crack content (CT) (Hudson, 1980) In addition, we use two crack orientation parameters of strike and dip. We assume that there is no change in crack orientation with depth. The model vector therefore contains a total of seventeen parameters. Discretization is applied to each of these parameters according to the scheme in Table 1 and results in a space comprising of approximately 10^{20} models.

GLOBAL OPTIMISATION METHODS

In this study we will test the techniques of Tabu Search (TS), Genetic Algorithms (GA) and Adaptive Simulated Annealing (SA). These techniques were selected since they offer the most promising search characteristics. As a control we also use a Monte Carlo search (MC). We now review each of these techniques.

Tabu search

Tabu or Taboo searches first appeared around the late 1980s from the operational research community (Glover, 1989). This class of search derives its name from the concept that certain moves are prohibited or "taboo". TS is a deterministic method using repeated local searches which avoids entrapment in local minima and encourages exploration of unsearched regions. TS achieves this through the

storage of information relating to the previous models sampled. These historical records are referred to as "Tabu lists". In practice several "Tabu lists" can be maintained containing information such as the models previously visited or previous improving moves which have been accepted. Under some circumstances the "Tabu" condition can be ignored when a criterion known as the "Aspiration condition" is satisfied. Such conditions are included to allow "interesting moves" which would otherwise be rejected. For example, a move to an exceptionally good model lying in a "Tabu direction" might be allowed to satisfy the aspiration condition since such a move may prove to be beneficial.

Table 1. Discretization applied to model parameters used in inversion problem (this defines model space of approximately 10^{20} models)

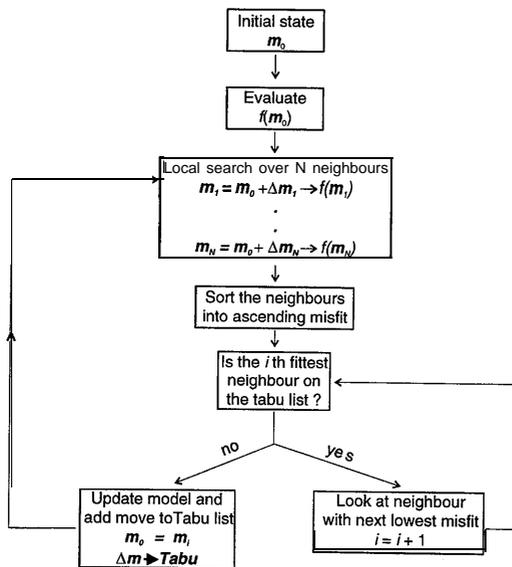| Parameter	Start	End	Number of intervals
CD_1	0.0001	0.2500	128
AR_1	10^{-1}	10^{-4}	4
CT_1	dry	saturated	2
CD_2	0.0001	0.2500	128
AR_2	10^{-1}	10^{-4}	4
CT_2	dry	saturated	2
CD_3	0.0001	0.2500	128
AR_3	10^{-1}	10^{-4}	4
CT_3	dry	saturated	2
CD_4	0.0001	0.2500	128
AR_4	10^{-1}	10^{-4}	4
CT_4	dry	saturated	2
CD_5	0.0001	0.2500	128
AR_5	10^{-1}	10^{-4}	4
CT_5	dry	saturated	2
Strike	-90" (North)	+90° (South)	256
Dip	-45"	+45"	256

TABU SEARCH

Figure 2. Flowchart showing general Tabu search scheme

To begin a TS a model is selected at random. A local search is then performed over the adjacent models. From these neighbours the best model is selected, note that this model need not be better than the starting model. This relative move from the starting model to the best neighbouring model is added to the "Tabu list". From this new model a local search is again performed. Once again the best neighbouring model is selected as a candidate for the next move. However, the move must not return the search to the previous starting model. To prevent such cycling the move is checked on the "Tabu
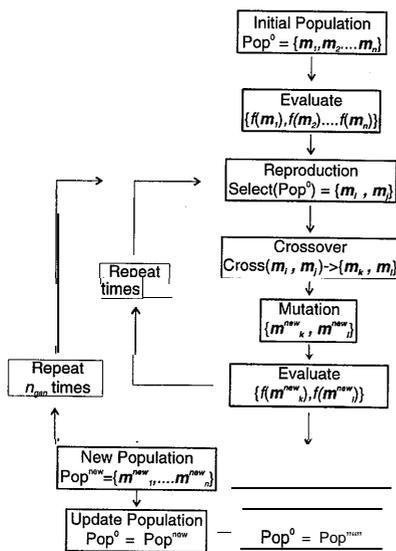
list" of previous moves. If the move is not on the list it is allowed otherwise the next best model is checked for its acceptability. This checking is repeated until an acceptable neighbour is found which does not lie in a "Tabu direction". This process of local search and selection of best non-tabu neighbour is then repeated until some termination criterion is satisfied. A flow diagram for the TS is given in Figure 2.

The two essential features of a TS which distinguishes it from other search strategies

1. the use of a "Tabu list" which prevents cycling whereby the search oscillates between previously visited states with no new areas explored;
2. an acceptance of a neighbouring solution which does not depend on the current value of the objective function, thus allowing the TS to climb out of local minima.

A few examples of TS applied to geophysical problems include Vinther and Mosegaard (1996) and Rowbotham and Pratt (1995).

Genetic algorithms

Genetic Algorithms (GA) were originally developed in the 1970s for the modelling of natural evol-

GENETIC ALGORITHM

Figure 3. Flowchart showing general GA scheme
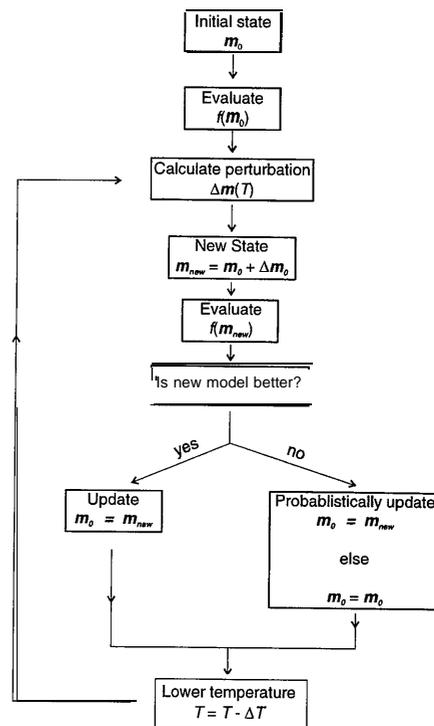
SIMULATED ANNEALING

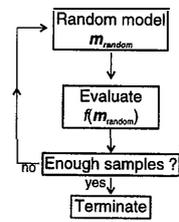Figure 4. Flowchart showing general SA scheme

MONTE CARLO

Figure 5. Flowchart showing general Monte Carlo scheme

utionary systems. Recently these techniques have been applied to many different optimisation problems including non-linear geophysical inversion (Sen and Stoffa, 1995; Horne and MacBeth, 1994).

Unlike many other optimisation techniques GAs use a pool of solutions. This is known as a population which appeals to the biological analogy. New models are then constructed from this pool by applying various genetic operations. The basic GA uses three operations known as reproduction, crossover and mutation. The reproduction operator is a stochastic operator which selects models from the current population in such a way that "better" models are more likely to be selected. A simple reproduction operator is biased roulette wheel selection where models are assigned sectors with a size related to their objective function. Models are then selected at random from this wheel. Clearly, the "better" models with larger sectors on the roulette wheel are more likely to be chosen than "poorer"

models with smaller sectors. Crossover operates on the models previously selected by reproduction and combines them together to form new models. Mutation is a random perturbation which is applied to the newly generated models. These newly generated models form the new population and this generation process is repeated until some termination criterion is satisfied. The sequence of these operations is shown in Figure 3. Further details of these types of scheme may be found in Goldberg (1989) and Horne (1995).

Simulated annealing

Simulated annealing (SA) refers to a class of optimisation methods which employ a control parameter analogous to a temperature. This temperature controls the extent of the search so that a high value allows a more extensive (global) search whilst a low value encourages a localised search. As SA progresses this temperature is lowered according to an annealing schedule so that the character of the search changes from a global to a local search. SA also includes a temperature controlled acceptance rule which determines whether a newly generated state is accepted or not. Typically if a newly generated state is better it will be accepted otherwise it is accepted with a probability which is a function of the temperature and the difference in the objective function values. At the start of a SA search the acceptance temperature is high so that non-improving moves are more likely to be tolerated compared with later stages when the
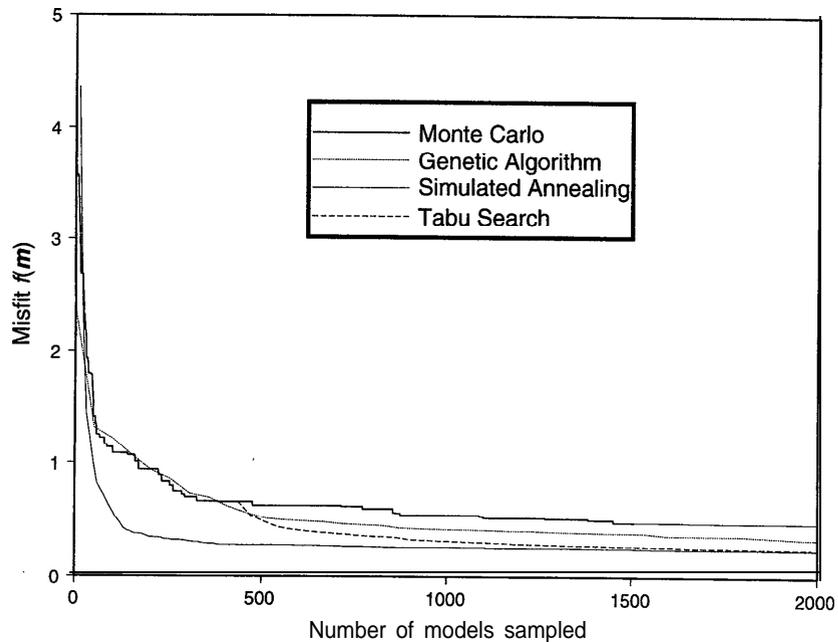

Figure 6. Convergence curves showing best model misfit obtained as function of number of models sampled
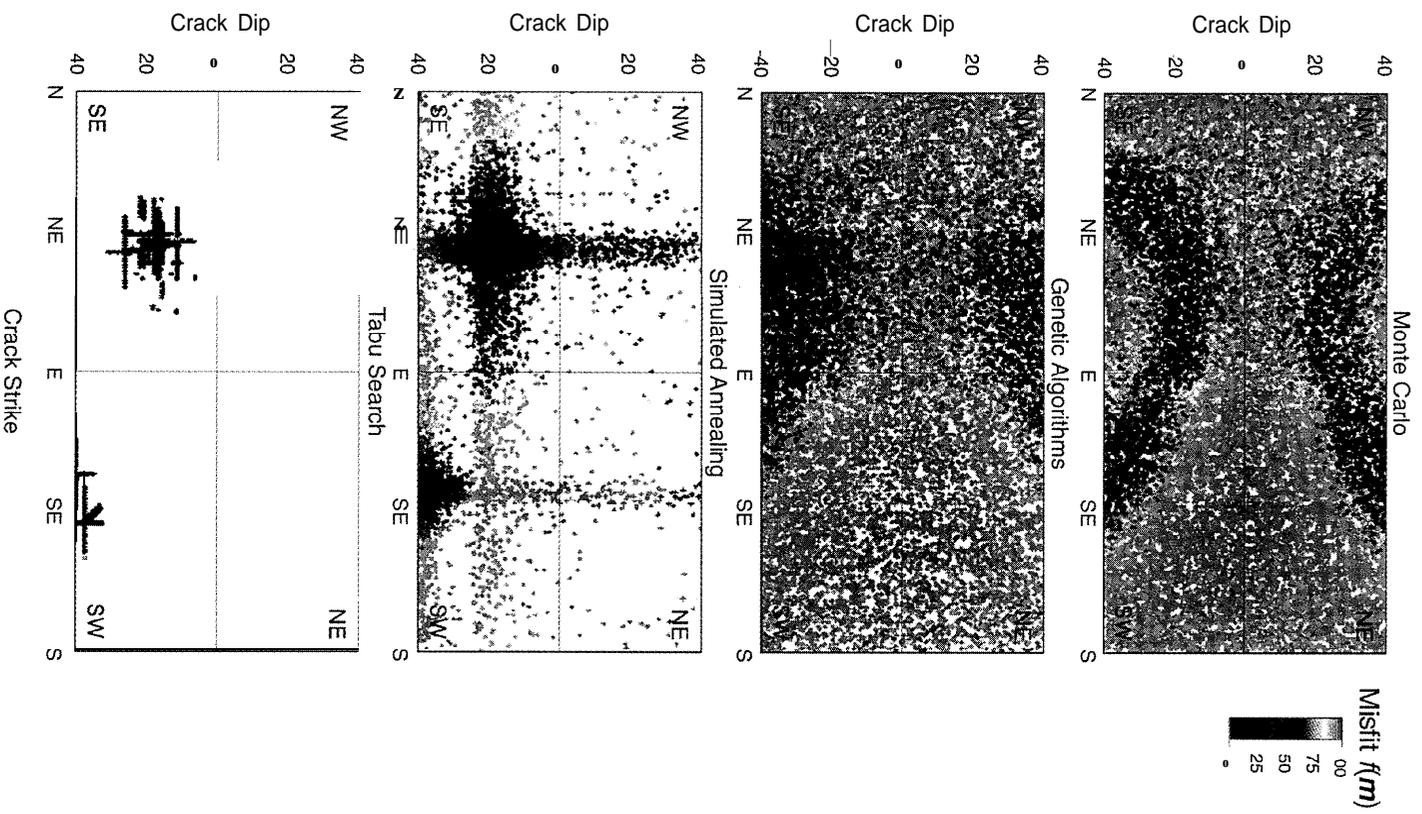
Figure 7. Scater plots showing model parameters of crack strike and crack dip sampled by different algorithms. Eah dot represents model sampled and is coloured according to corrsponding misfit value
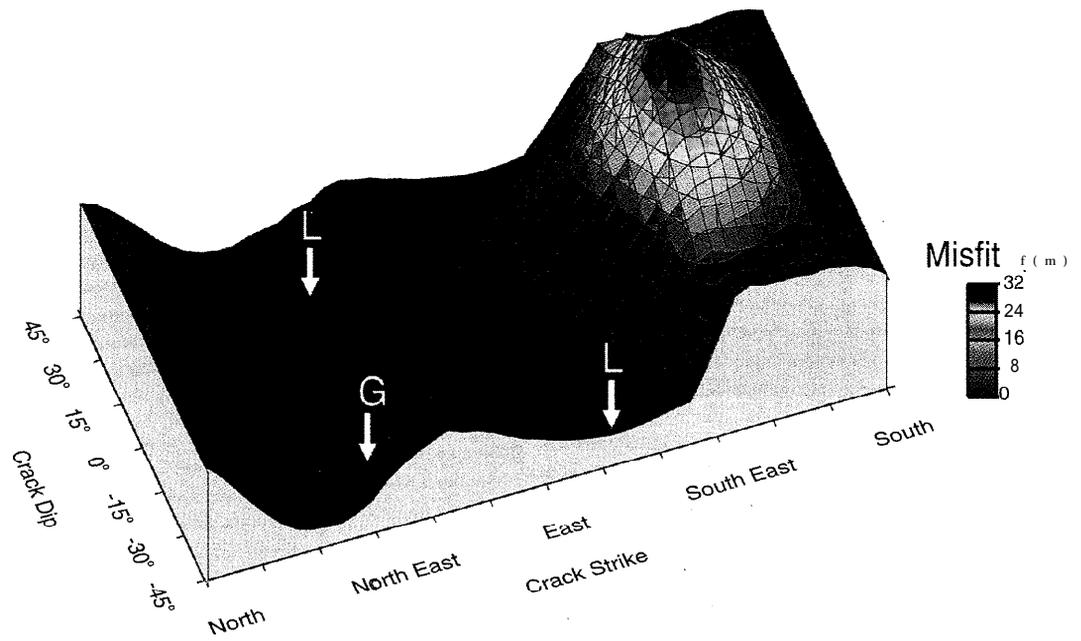
Figure 8. Two dimensional cross-section of model space about optimal solution. Ls denote local
minima and G global solution

system is cooler. A general flowchart for a SA search is shown in Figure 4.

There are many variants of SA schemes such as Boltzmann annealing, Fast (Cauchy) annealing or the Heat Bath algorithm (Sen and Stoffa, 1995). The method that we use in our inversion scheme is based upon the Adaptive Simulated Annealing algorithm (ASA) (Ingber, 1989). ASA differs from other schemes in several ways:

1. each model parameter is assigned a temperature;
2. the probability distribution function used in the generation of new models has a "fatter" tail than the Boltzmann or Cauchy distributions which are typically used in SA. This has the advantage that "outliers" may be sampled even in the closing stages of the search;
3. the annealing schedule is a dynamic function which is periodically rescaled.

Monte Carlo

This search is used as a control against which the effectiveness of the other techniques are compared. Monte Carlo (MC) methods are random searches where the model parameters are selected from a uniform distribution. There is no underlying search mechanism and models are continuously sampled until a sufficient number of samples are generated from which a statistical estimate can be made (Press, 1968). This scheme is shown in Figure 5.

RESULTS AND DISCUSSION

Because of the stochastic nature of these optimisation methods we show results that are obtained from the average of several runs initiated with different starting models.

The most common comparison test applied to optimisation techniques is the convergence rate. This is shown in Figure 6 where the horizontal axis represents the number of models sampled and the vertical axis the lowest objective function value found. The MC search shows the poorest convergent behaviour which is expected owing to the absence of any search mechanism. The TS is initiated after an initial random sample of 400 models so as to obtain a good starting model. As a result the TS shows the same behaviour as the MC search up to 400 samples. After this there is an apparently asymptotic convergence to a terminal value of $f(\mathbf{m}) = 0.250$. SA converges to a similar value within approximately 500 samples with little improvement subsequent to this. The convergent behaviour of the GA is essentially the same as the MC search within the first 400 samples. After this the GA continues to converge to a final misfit value off(m) = 0.300. The failure of the GA to converge to the absolute Global solution is not unexpected.

This is because GAs are designed to preferentially sample the more significant regions of model space. A crude ranking of these methods in terms of convergent behaviour would place first SA followed by TS, GA and MC in respective order. However, it should be noted that this ranking is based upon the number of models sampled and not the number of algorithmic iterations. In the case of the TS and GA each iteration required the sampling of 34 and 40 models respectively whereas SA is a sequential search. Had the algorithms been implemented on a parallel machine the GA and TS could naturally exploit such an architecture using a task farm approach (each processor evaluating a single model). The same cannot be said for SA which involves a step by step search.

Although the convergent behaviour of these methods is important a more thorough comparison involves the consideration of reliability. When viewed in these terms the GA and SA show less variance in the final misfit values compared to the TS (Table 2). This variability is due to the dependence of TS on the initial starting model. TS descends from this point to the nearest local minimum and then thoroughly explores the surrounding region. As a result TS can expend many function evaluations before escaping this minimum. This behaviour is apparent in Figure 7. Each point in this plot represents a sampled model and is coloured according to the model misfit. The TS can be seen to be far more localised in comparison with the other methods. In some of the runs the TS is trapped in local minima located around crack strikes of south east and crack dips of 40" to the south west. To climb out of this local minimum to the global minimum (crack strike north east and crack dip 20" to the south east) would involve a trek along an approximately circular shaped valley in two of the dimensions as shown in Figure 8. In reality the remaining model parameters complicate this valley structure introducing many more local minima thus compounding the TS method.

All these optimisation schemes are controlled by user defined search parameters. If these parameters are found to critically control the algorithm's success and cannot be estimated efficiently then the usefulness of the algorithm is limited. An example of such critical behaviour is associated with SAs. If the cooling is too rapid there is premature conver-

Table 2. Statistical results of terminal misfit values obtained from 11 repeated runs using different search methods

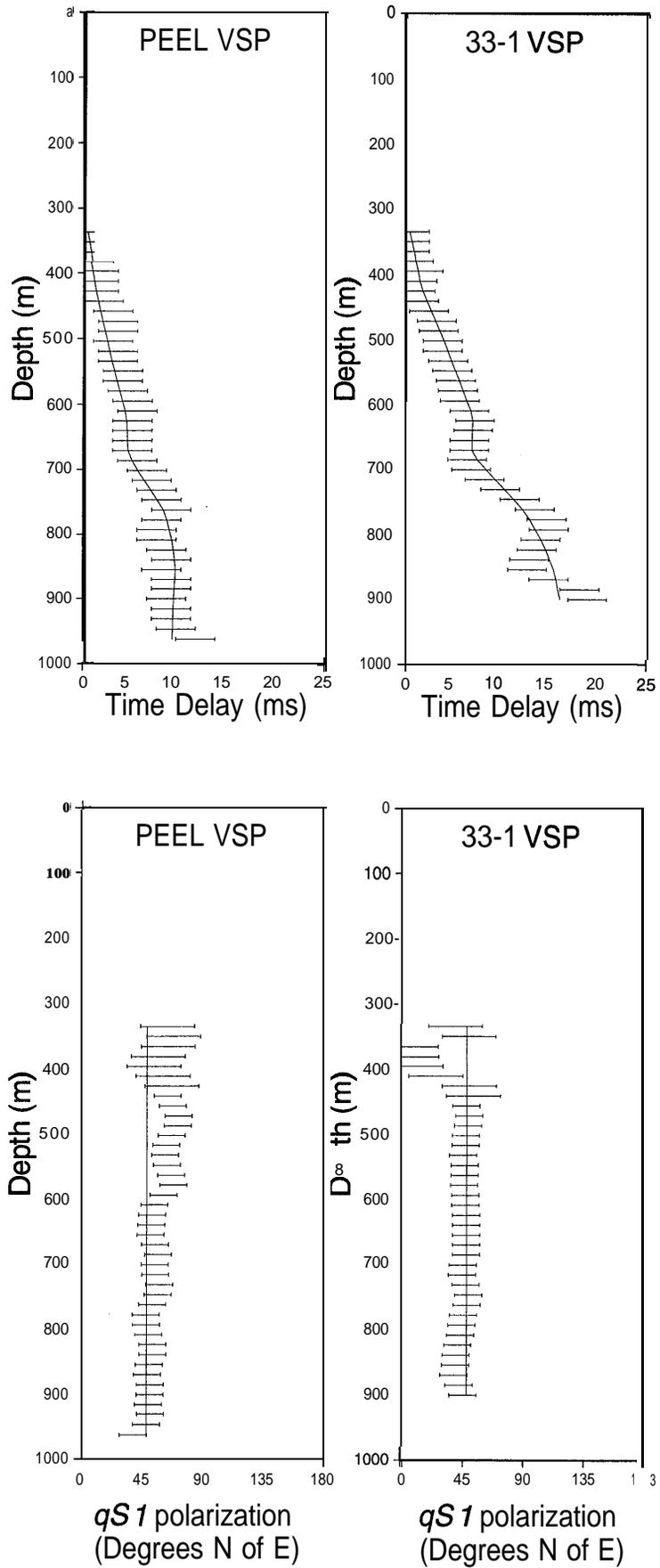|	MC	GA	SA	TS
f_{max}	0.67	0.37	0.36	0.51
f_{min}	0.33	0.21	0.20	0.18
f_{avg}	0.47	0.31	0.24	0.25
σ	0.10	0.05	0.05	0.09

Figure 9. Comparison of observed and modelled data corresponding to best model

Table 3. Control parameters used in different algorithms

GA	SA	TS
Population size	Initial parameter temperature x NP	Tabu list size
Mutation probability	Terminal parameter temperature x NP	Neighbourhood size
Crossover probability	Initial acceptance temperature	
	Terminal acceptance temperature	
	Number of steps	

NP: number of model parameters.

gence to a sub-optimal solution and if cooling is too slow the algorithm converges too slowly. Our implementation of SA is the ASA method which is relatively insensitive to the annealing schedule since temperatures are periodically rescaled in relation to the local curvature of model space. The tabu search which we use requires a choice of the Tabu list size and also the "size" of the neighbourhood. The neighbourhood size is defined in terms of the number of discretization intervals for each model parameter. Our implementation of TS defines a neighbourhood which decreases in size to one discretization unit for each model parameter as the search progresses. This is necessary since a neighbourhood using a single discretization unit for each parameter does not allow a sufficiently large step size for the TS to reach the minima within a reasonable number of iterations. This decreasing neighbourhood size is similar to the role of the model temperature parameters employed in SA methods. Simple GAs employ control parameters whose values are generally not critical to the search. (Table 3 lists the control parameters used in the different algorithms.)

As a further consideration when choosing different algorithms is the ease of implementation for each algorithm. A simple measure of this is given by the number of lines occupied by each algorithm. SA and TS, required approximately the same number of lines of code whereas the GA required more than twice as much.

The overall best model is obtained using the TS. This model has a misfit value of $f(m) = 0.180$ which represents a good fit as can be seen in Figure 9. Perhaps the most interesting feature of the global solution is the suggestion of a sub-vertical crack system. The crack dip parameter indicates cracks dipping at 20" from the vertical towards the south east. The crack strike parameter is in good agreement with previous studies relating to the crack system observed at the test site (Queen and Rizer, 1990).

CONCLUSIONS

A frequent pitfall with comparative studies of this type is to generalise the results to other pro-

blems. Each search method has its own characteristics which make the various search techniques more suited to different problem domains. This implies that a preliminary investigation of the likely nature of the model space may prove to be useful in the choice of search algorithm. However, we are able to suggest that an "aggressive" local search such as TS is a useful final step in refining the best model found by any other means.

REFERENCES

Glover, F. (1989) Tabu search-Part I. *Operations Research Society of America Journal on Computing* **1**(3), 190–206.

Goldberg, D. E. (1989) *Genetic Algorithms in Search, Optimisation, and Machine Learning.* Addison-Wesley Publishing Company, New York, 412 pp.

Horne, S. (1995) Applications of genetic algorithms to problems in seismic anisotropy. Ph.D dissertation, University of Edinburgh, Scotland, 229 pp.

Horne, S. and MacBeth, C. (1994) Inversion for seismic anisotropy using genetic algorithms. *Geophysical Prospecting* **42**(8), *953-974.*

Hudson, J. A. (1980) Overall properties of a cracked solid. *Mathematical Proceedings of the Cumbridge Philosophical Society 88, 371-384.*

Ingber, L. (1989) Very fast simulated re-annealing. *Journal of Mathematical Computer Modelling 12, 967-973.*

Press, F. (1968) Earth models obtained by Monte Carlo inversion. *Journal of Geophysical Research* **73**(16), *5223-5234.*

Sen, M. K. and Stoffa, P. L. (1995) *Global Optimisation Methods in Geophysical Inversion.* Elsevier, Amsterdam, 294 pp.

Queen, J. H. and Rizer, W. D. (1990) An integrated study of seismic anisotropy and the natural fracture system at the Conoco Borehole Test Facility, Kay County, Oklahoma. *Journal of Geophysical Research B* **95**(7), 11255-l 1273.

Rowbotham, P. S. and Pratt, R. G. (1995) Optimisation of anisotropic velocity fields. In *57th Annual Meeting of the European Association of Geoscientists and Engineers, Glasgow.* Extended Abstracts 1, E045.

Vinther, R. and Mosegaard, K. (1996) Seismic inversion through Tabu search. *Geophysical Prospecting* **44**(4), **555-570.**

Zeng, X. and MacBeth, C. (1993) Algebraic processing techniques for estimating shear-wave splitting in zero-offset VSPs—theory. *Geophysical Prospecting* **41**(8), 1033-1066.