

# Non-linear multiparameter inversion using a hybrid global search algorithm: applications in reflection seismology

Pengcheng Liu,<sup>1</sup> Stephen Hartzell<sup>2</sup> and William Stephenson<sup>2</sup>

<sup>1</sup>Institute of Geophysics, Chinese Academy of Sciences, Box 9701, Beijing 100101, People's Republic of China

<sup>2</sup>US Geological Survey, Branch of Earthquake and Landslide Hazards, Box 25046, MS 966, Denver, CO 80225, USA

Accepted 1995 March 23. Received 1995 February 17; in original form 1993 December 22

## SUMMARY

Many interesting inverse problems in geophysics are non-linear and multimodal. Parametrization of these problems leads to an objective function, or measure of agreement between data and model predictions, that has a complex topography with many local minima. Optimization algorithms that rely on local gradients in the objective function or that search the model space locally may become trapped in these local minima. By combining simulated annealing with the downhill simplex method, a hybrid global search algorithm is presented in this paper for non-linear, multimodal, inverse problems. The hybrid algorithm shares the advantages of both local search methods that perform well if the local model is suitable, and global methods that are able to explore efficiently the full model space. The hybrid algorithm also utilizes a larger and more complex memory to store information on the objective function than simulated annealing algorithms. The effectiveness of this new scheme is evaluated in three problems: minimization of the multidimensional Rosenbrock function, non-linear, 1-D, acoustic waveform inversion, and residual statics. The performance of the hybrid algorithm is compared with simulated annealing and genetic algorithms and is shown to converge more rapidly and to have a higher success rate of locating the global minimum for the cases investigated.

**Key words:** inversion, inverse problem, seismic reflection.

## INTRODUCTION

Techniques for non-linear, multiparameter optimization may be conveniently classified into two groups, local and global (Sambridge & Drijkoningen 1992). Methods in the first group rely on using local information about the gradient of an objective function to improve upon some initial model in an iterative fashion. The objective function is defined as a quantitative measure of the agreement between the model predictions and the observed data and may incorporate *a priori* geophysical or geological constraints on the model parameters. At each iterative step, the local method examines only those models within the immediate vicinity of the current solution in order to choose its next step. No attempt is made to reconnoitre distant portions of the model space. The gradient of the objective function is often utilized in this process. The second class of methods requires only the value of the objective function, not its derivative, and uses random processes to search the model space to find the best solution. In so doing, the global method makes an explicit effort to search more widely in model space.

In practice, many geophysically interesting inverse problems have objective functions that are non-quadratic

and possess many local minima. As a consequence, the solution obtained from the local method may depend strongly on the starting model, and may become trapped in a local, rather than a global, minimum. In addition, the calculation of derivative information can become difficult and costly. But, if the starting model is chosen appropriately, local methods can be very efficient in locating a global minimum of the objective function. An alternative to the local method is the so-called 'exhaustive scheme' (Smith, Scales & Fischer 1992), in which each point in the model space is tested sequentially. However, the size of many practical model spaces is far too large to allow for a solution via an exhaustive search. Global methods, such as 'simulated annealing' and 'genetic algorithms', are often more attractive alternatives.

Simulated annealing is analogous with physical annealing in thermodynamic systems (Scales, Smith & Fischer 1992), in the way that liquids freeze and crystallize, or metals cool and anneal. For slowly cooled systems, nature is able to find minimum energy states. As the temperature is carefully lowered, a system tends to freeze into its lowest energy state, which corresponds to the global minimum of an objective function. Genetic algorithms operate on the basis of an

analogy between function optimization and a biological system, in which the principle of 'survival of the fittest' is exploited. Genetic algorithms consist of a set of operations that are applied to a group of models to produce a new group with the property that the members of the new group have a greater expected average fitness than their predecessors. Both simulated annealing and genetic algorithms are non-linear, random processes which require no derivative information and have the potential for a significant increase in efficiency over a random walk strategy. A critical and practical problem encountered by both simulated annealing and genetic algorithms, however, is the efficient production of an updated model from the current model. Moreover, simulated annealing algorithms also experience the phenomenon known as 'critical slowing down' (Brower *et al.* 1989), in which the computation time is dominated by a grid size or step length that is essentially unrelated to the underlying structure of the objective function (Scales *et al.* 1992). In other words, as the number of parameters in the problem increases, the size of the model space increases dramatically, and searching the objective function space with a single, independent, random process or Markov chain becomes very time consuming.

By combining simulated annealing with the downhill simplex algorithm, we suggest a hybrid global search algorithm for non-linear, multiparameter, inverse problems. The simplex method belongs to the class of local methods in which an  $N$ -dimensional simplex crawls its way downhill in the objective function space. This new scheme shares the advantages of both a local search method that performs well if the local model is suitable, and a global method that is able to explore efficiently the entire model space. Our hybrid method also has a greatly increased capacity to store its accumulated experience of objective function evaluations over simulated annealing algorithms. To illustrate the effectiveness of this hybrid global search algorithm, we consider three problems: determination of the minimum of the multidimensional Rosenbrock function (Oren 1974), non-linear 1-D acoustic waveform inversion (Sen & Stoffa 1991, 1992; Sambridge & Drijkoningen 1992), and calculation of residual statics (Rothman 1985, 1986). The results are compared with those obtained from other global methods, specifically, simulated annealing and genetic algorithms.

## HYBRID GLOBAL SEARCH METHOD

We first review the properties of simulated annealing and the simplex method, and then describe the combination of these two methods.

### Simulated annealing algorithms

Simulated annealing is a generalization of the Monte Carlo method of Metropolis *et al.* (1953). Since its full development by Kirkpatrick, Gelatt & Vecchi (1983), it has been used in many-parameter global optimization problems (van Laarhoven & Aarts 1988). Recently, simulated annealing has been applied to the estimation of near-surface delay times in seismic reflection data (Rothman 1985, 1986), and to seismic waveform inversion to recover density and velocity profiles (Sen & Stoffa 1991). A brief but insightful

discussion of simulated annealing can be found in Scales *et al.* (1992).

Our overview of simulated annealing follows the development of Lane (1992). Following the analogy with a thermodynamic system, simulated annealing is initiated at a suitably high temperature  $T_0$ , and then allowed to slowly cool according to an annealing schedule until the lowest energy state is encountered. The annealing schedule determines the cooling rate and the length of time the algorithm remains at each temperature. At each step in the schedule, the algorithm must make a prespecified number of random transitions of the model,  $\Delta M$ , within the model space. These random transitions are generated at time  $t$  according to a probability distribution known as the generation distribution,  $P_g(\Delta M, t)$ . The suitability of a transition is measured by the change in an objective function,  $\Delta E$ . A new model is accepted if  $\Delta E < 0$ . However, if  $\Delta E > 0$ , the new model is accepted with probability  $P_a(\Delta E)$ , which is defined as the acceptance probability. Therefore, to make use of a simulated annealing algorithm, one must provide the following elements:

- (1) objective function,
- (2) initial temperature and cooling schedule,
- (3) random generator for moving from one model to the next, or generation distribution,
- (4) acceptance distribution, and
- (5) convergence criterion.

In our formulation of simulated annealing the model parameters,  $m_i$ , of model  $M$  are initialized by

$$m_i = m_{i, \min} + \xi_i(m_{i, \max} - m_{i, \min}). \quad (1)$$

$\xi_i$  is a random number uniformly distributed on the interval  $(0, 1)$ .  $m_{i, \min}$  and  $m_{i, \max}$  are predetermined minimum and maximum constraints on the model parameters. This initialization allows the model parameters to range uniformly over the preset model space we wish to consider. Ideally, one would like to initiate the algorithm at a high initial temperature,  $T_0$ , and then cool very slowly until convergence to the global minimum occurs. However, this process may require enormous computing time. As an example, Rothman (1986) employed the cooling schedule

$$T_k = T_0(0.99)^k, \quad (2)$$

where  $k$  is the iteration number. Iterations continue until some predetermined minimum value in  $\Delta E$  is reached or a preset number of iterations is completed. By using  $\Delta E$  rather than  $E$ , the convergence criteria are sensitive to the local gradient of the objective function. If the cooling schedule and the minimum value of  $\Delta E$  are chosen properly, the algorithm should terminate at or near the global minimum, subject to the constraints and prior information imposed on the solution (Geman & Geman 1984). Other possible cooling schedules can be found in Press *et al.* (1992). Often physical insight and/or trial-and-error experiments are required to select a sufficient initial temperature and cooling schedule.

Szu & Hartley (1987) first introduced the idea of using a Cauchy distribution for the model generation distribution. The Cauchy distribution is desirable because its Gaussian-like peak and Lorentzian tails allow occasional long jumps

among local sampling of the model space. However, the particular formulation of Su & Hartley (1987) is difficult to apply in practice for multiparameter problems because there is no algorithm to generate efficiently an  $N$ -dimensional Cauchy distribution, as in the 1-D case. Instead we use the formulation of Lane (1992), with

$$m_i^j = m_i^{j-1} + \xi_i^j \chi_k^j, \quad (3)$$

where

$$\chi_k^j = T_k \tan [\pi(\xi_k^j - 0.5)]. \quad (4)$$

The subscript  $i$  refers to the model parameter, the subscript  $k$  to the iteration number, and the superscript  $j$  to the perturbation number. The number of perturbations required is discussed below.  $\xi_k^j$  is uniformly distributed on the interval  $(0, 1)$ , and  $\xi_i^j$  is uniformly distributed on the interval  $[-\Delta m_i, \Delta m_i]$ , where  $\Delta m_i = (m_{i, \max} - m_{i, \min})$ .  $\chi_k^j$  is Cauchy distributed and is computed using eq. (4). Eq. (3) yields a transition in the model that is uniformly distributed in direction (in model parameter space) and Cauchy distributed in distance. In addition, we have the following constraints.  $\chi_k^j$  must lie on the interval  $(-1, 1)$ . If not, a new variate is computed.  $m_i^j$  must lie on the interval  $(m_{i, \min}, m_{i, \max})$ . If  $m_i^j$  is less than  $m_{i, \min}$ , it is set equal to  $m_{i, \min}$ . If  $m_i^j$  is greater than  $m_{i, \max}$ , it is set equal to  $m_{i, \max}$ .  $m_{i, \min}$  and  $m_{i, \max}$  are presumed to be set appropriately to encompass all geophysically reasonable models and *a priori* geophysical and geological information. This constraint on the model parameters is offered as an example. In simulated annealing and other combinatorial optimization algorithms there is great flexibility in the choice and implementation of constraints. Useful constraints can often be built into the objective function along with weighting factors on these constraints.

The Gibbs probability distribution,

$$P_a(\Delta E) = \exp(-\Delta E/T_k), \quad (5)$$

usually serves as an acceptance criterion (Rothman 1985; Sen & Soffa 1991), because it expresses the principles of the Boltzmann probability distribution found in natural systems. We use the distribution of Lane (1992), which is similar to the one used by Bohachevsky, Johnson & Stein (1986), and has the form

$$P_a(\Delta E) = \alpha^{(\Delta E/\Delta E_a)}, \quad (6)$$

where

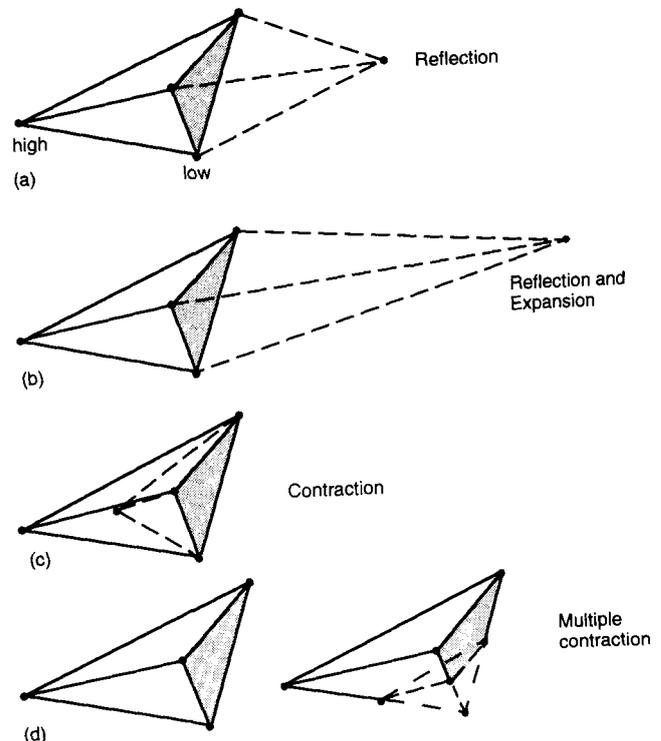
$$\Delta E = E(M^j) - E(M^{j-1}).$$

$\Delta E$  is the change in the objective function from model  $M^{j-1}$  to model  $M^j$ ,  $\Delta E_a$  is the average of the  $\Delta E$  for accepted transitions in the model during the previous iteration of the algorithm, and  $\alpha$  is a constant between 0.7 and 0.8. The larger the value of  $\alpha$ , the higher the acceptance rate. A model perturbation is accepted if  $P_a(\Delta E)$  is greater than  $\xi$ , where  $\xi$  is a random number uniformly distributed on the interval  $(0, 1)$ . Bohachevsky *et al.* (1986) and Lane (1992) found that acceptance criteria of the form in eq. (6) were less likely to wander throughout model space and converged

more rapidly than the Gibbs probability distribution. The factor  $\Delta E_a$  in eq. (6) introduces a feedback component into the acceptance criterion that makes the algorithm less likely to wander out of a global minimum but the algorithm retains the ability to escape local minima (Lane 1992).

### Downhill simplex algorithm

A detailed account of the downhill simplex algorithm can be found in Press *et al.* (1992), but for completeness it is summarized here. The basic downhill simplex algorithm is due to Nelder & Mead (1965) and belongs to the category of local search methods. This algorithm requires only objective function evaluations, and no derivatives. At the heart of the algorithm is the construction of an appropriate 'simplex', a spatial configuration of  $N$  dimensions defined by  $N+1$  points, where  $N$  is the number of model parameters. For a problem with three model parameters the simplex is a figure in 3-D model space. For larger numbers of parameters the simplex becomes a high-order hypercube. The method takes a series of steps to make its way downhill through the complexity of the  $N$ -dimensional objective function topography, until it encounters a minimum. Most of the steps involve moving a point on the simplex where the objective function is largest, through the opposite face of the simplex to a lower value. These steps are called reflections. If this new value is lower than the present lowest value in the simplex, the method expands the simplex further in the same direction. The method will move the simplex down a



**Figure 1.** Possible moves of a tetrahedral simplex. Each change in the simplex is indicated by dashed lines. (a) Reflection away from the high point, (b) reflection and expansion away from the high point, (c) contraction along one dimension away from the high point, and (d) contraction along all dimensions toward the low point. (Modified from Press *et al.* 1992.)

valley in the objective function or contract it to pass through a restricted space. If all the steps mentioned above do not produce a smaller objective function, the method contracts the simplex in all directions, pulling itself in around the smallest value of the objective function. The basic moves of the algorithm are illustrated in Fig. 1.

### Combination of simulated annealing and downhill simplex

In this work we modify the model generation steps of the simulated annealing algorithm by introducing the points-moving procedure of the downhill simplex algorithm; namely reflections, expansions and contractions of the simplex. The new model generation scheme has the form

$$M^j = M^{j-1} + \Delta M'_d + \Delta M'_p, \quad (7)$$

where  $M^j$  and  $M^{j-1}$  stand for the new and current models, respectively.  $\Delta M'_p$  is constructed according to the generation distribution, eq. (3).  $\Delta M'_d$  comes from the downhill simplex algorithm. The new model,  $M^j$ , is always accepted if its objective function is lower than the current one, but an unimproved or less desirable model is also accepted with the acceptance probability given by eq. (6). The algorithm is initiated by choosing  $N+1$  models corresponding to the  $N+1$  points of the simplex. All the model parameters in these  $N+1$  models are selected using eq. (1). The algorithm then goes through an inner loop on perturbation number,  $j$ , and an outer loop on iteration number,  $k$ . If the perturbation to the model in eq. (7) is accepted, the model that has the largest value of the objective function in the simplex is replaced. If the perturbation is not accepted, a new one is calculated. The algorithm does not advance to a new iteration number until at least  $10N$  perturbations are accepted. If  $\alpha$  in eq. (6) is between 0.7 and 0.8, 50–90 per cent of the perturbations are accepted. The algorithm terminates either when a preset number of iterations have been completed, or when the difference between the largest and smallest objective function values in the simplex reaches an error limit. The latter convergence criterion makes sense because it is sensitive to the gradient of the objective function, which will be zero at a minimum. If the cooling schedule is chosen properly, this minimum will be at or near the global minimum. In practice, a convergence criterion based on the absolute value of the objective function may be awkward to impose because we do not know *a priori* the minimum value of the objective function given the existence of constraints and prior information on the solution.

The weight given to the contributions of the two methods to the new model is controlled by the annealing temperature, eq. (2). At higher temperatures, the random generator of the simulated annealing dominates the calculation of the new model. As the temperature is lowered the downhill simplex algorithm begins to play a more important role in the selection of the new model. Conceptually, for a suitable initial temperature and cooling schedule, simulated annealing initially searches widely to find an appropriate model that is not far from the global minimum. The downhill simplex algorithm then moves the model quickly to the global minimum. Throughout the process the non-zero probability of long jumps allows the method to escape from local minima. The simplex

consists of  $N+1$  points that store  $N+1$  different values of the objective function. Every one of the points of the simplex corresponds to a solution, the merit of which is judged by the value of the objective function. Therefore, at any given time the algorithm has a stored experience of  $N+1$  models and objective function values. This scheme allows for a far more complex and capacious memory of accumulated experience and leads to more efficient searching of the model space than simulated annealing algorithms. In simulated annealing a new model estimate is compared only with the previous one. In the hybrid algorithm each new model estimate is compared with  $N+1$  previous models. At higher temperatures the simplex ranges over the entire model space. At lower temperatures it searches more and more locally. Press *et al.* (1992) have proposed a different algorithm that combines simulated annealing and the downhill simplex method. Their method adds a random variable, proportional to the temperature, to the value of the objective function associated with every vertex of the simplex.

The success or failure of an algorithm like simulated annealing is often determined by the choice of initial temperature and cooling schedule. This observation is also true of our hybrid method. Geman & Geman (1984) have shown that for sufficiently slow cooling simulated annealing will reach the global minimum. The initial temperature should be high enough to allow the algorithm to search the entire model space freely, but not become trapped in a local minimum. It should also be low enough so that proportionately more time is spent sampling local neighbourhoods. An exponential cooling schedule is usually employed in the hybrid algorithm. The temperature at each iteration is obtained from

$$T_k = T_0(\beta)^k, \quad (8)$$

where  $T_0$  is the initial temperature, and  $\beta$  is an experimentally determined constant less than 1.

## NUMERICAL EXAMPLES

Since no one global optimization scheme is best for all situations, we investigate a variety of problems. Three test problems are employed to evaluate the effectiveness of the hybrid global search algorithm: (1) the multidimensional Rosenbrock function; (2) 1-D, non-linear, acoustic wave-form inversion; and (3) residual statics.

### The $N$ -dimensional Rosenbrock Function

The test function is of the form

$$f(x) = \sum_{i=1}^{N-1} [100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2], \quad (9)$$

with a unique minimum at  $x^* = (1, 1, \dots, 1)$  (Oren 1974). Because of this function's gradual slopes and shallow global minimum, it has been extensively used to test the effectiveness of global search algorithms. Although the Rosenbrock function has a unique minimum and is therefore not multimodal, it is still a worthy consideration for global optimization algorithms and we use it to demonstrate an

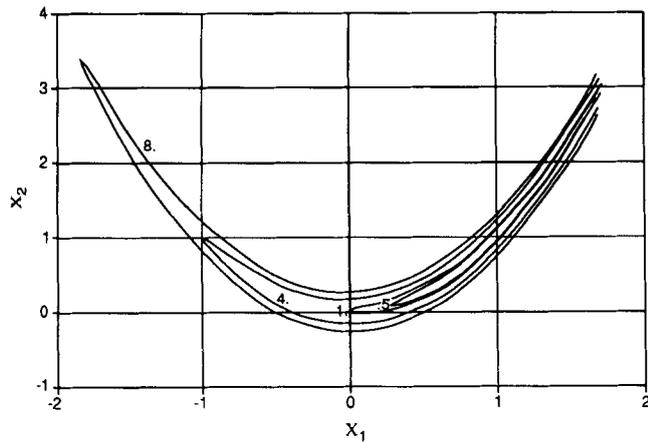


Figure 2. Contour plot of the 2-D Rosenbrock function with the global minimum at (1,1). (From Lane 1992, published with permission of the Colorado School of Mines.)

advantage of the hybrid algorithm. The important question becomes not whether the minimum can be found with extreme computational effort, but which method is the most efficient and practical. A plot of the Rosenbrock function for  $N = 2$  is given in Fig. 2.

In our first test, the effectiveness of several different algorithms is evaluated using the Rosenbrock function with  $N = 10$ . For this case the downhill simplex method fails to converge to the global minimum. The hybrid global search algorithm and simulated annealing perform very well. The main difference between the performance of the latter two algorithms is in the number of objective function evaluations. The hybrid global search method requires approximately five times fewer function evaluations than simulated annealing. As the number of variables is increased to 50, the hybrid algorithm finds the global minimum after 200 iterations and  $8 \times 10^5$  evaluations of the objective function. (Iterations refers to cooling schedule steps.) Simulated annealing, however, fails to find the global minimum, even after 1000 iterations and several different initial temperatures have been tried. The failure of simulated annealing is attributed to the very shallow nature of the global minimum and the fact that simulated annealing has no efficient mechanism for moving the solution to the global minimum at low temperatures. On the other hand, inclusion of the downhill simplex allows the global minimum to be efficiently found after simulated annealing has defined the correct neighbourhood. The hybrid algorithm also works well for 100 variables. For this problem, the number of iterations and function evaluations are 600 and  $5 \times 10^6$ , respectively. The performance of the hybrid algorithm is summarized in Table 1. The initial temperature and cooling schedule used are given by  $T_k = 0.13(0.98)^k$ .

Table 1. Multidimensional Rosenbrock function: hybrid algorithm.

Variables	Trials (All convergent)	objective function evaluations
10	2	$3 \times 10^4$
50	2	$8 \times 10^5$
100	3	$5 \times 10^6$

### 1-D acoustic waveform inversion

The test problem we employ is the same as used by Scales *et al.* (1992), i.e. determination of the reflection coefficients for a 1-D acoustic earth model from a normal-incidence seismogram in the presence of multiples. In the absence of multiples this problem is linear and can be solved by deconvolution (Robinson & Treitel 1980). However, our formulation includes multiples and cannot be solved by linear methods. This problem also affords us with the opportunity to do direct comparisons with the performance of simulated annealing and genetic algorithms (Scales *et al.* 1992). It is assumed that each layer has unit two-way traveltime and the reflection coefficients at each interface are used as earth model parameters instead of the velocities and densities (Gutowski & Treitel 1987). We also assume that the source function is known and that the acoustic response of the earth model is our observed data. Our objective function is the  $L_2$  norm of the difference between the observed and model-predicted waveforms. This particular formulation allows extremely rapid forward modelling of the synthetic seismograms, including multiples. We use the algorithm described by Gutowski & Treitel (1987) for the forward calculation.

The true earth model used by Scales *et al.* (1992) has two non-zero reflection coefficients at layers 5 and 10. The source function and the true model response are shown in Fig. 3. Inversions were run assuming a total of 15, 22 and 30 layers or reflection coefficients defining the model. In each case there are still only two non-zero reflection coefficients in the true model. In all these examples, the reflection coefficients are constrained to lie in the interval  $(-1, 1)$ . For the 15 layer problem, both simulated annealing and the genetic algorithm converge rapidly to the true solution (Scales *et al.* 1992). Simulated annealing obtains an

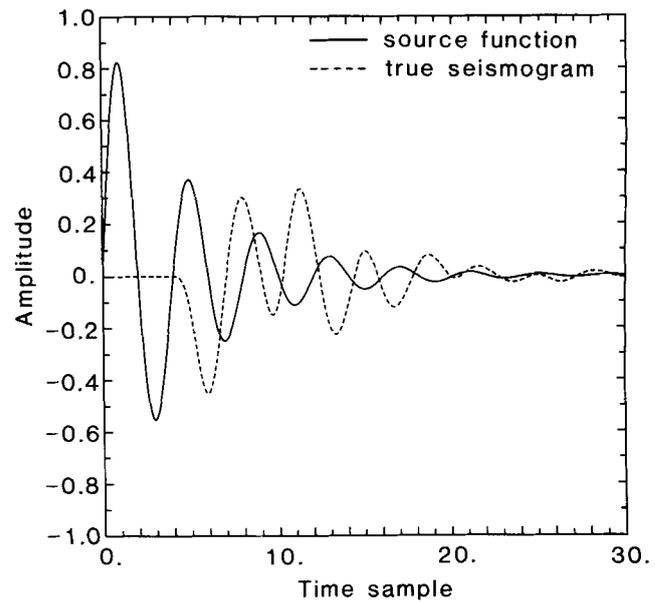
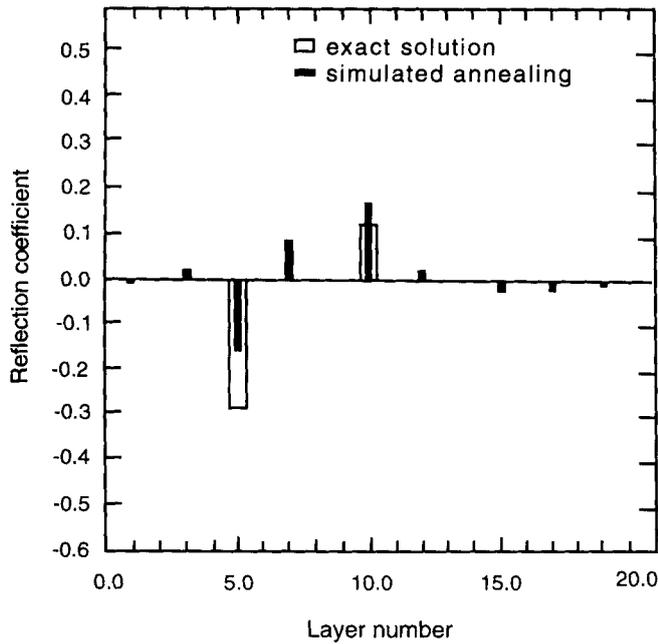


Figure 3. Source function and 1-D acoustic response of earth model with two non-zero reflection coefficients. The source is an exponentially damped sinusoid of the form  $\sin(2\pi i/4) \exp(-i/5)$ , where  $i$  is the time sample. (Modified from Scales *et al.* 1992.)



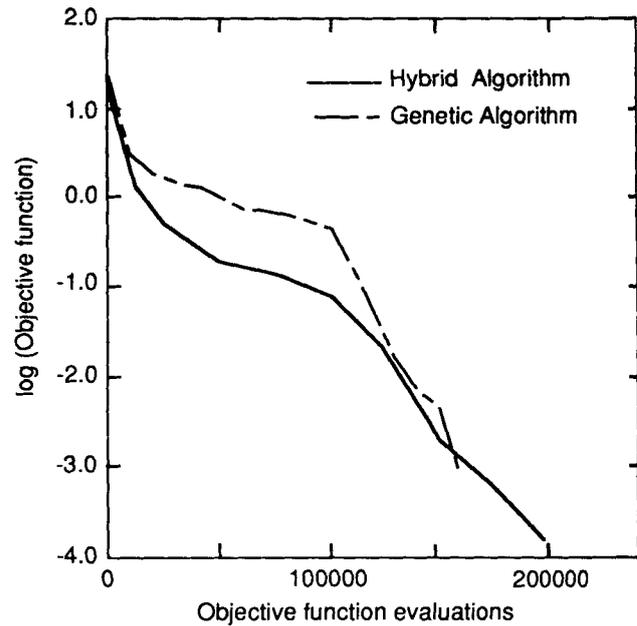
**Figure 4.** Simulated annealing inversion results for the 22 layer problem. (Modified from Scales *et al.* 1992.)

approximately correct but noisy solution (Fig. 4) for the 22 layer problem, and recovers only one of the non-zero reflection coefficients for the 30 layer problem. Ten trials were run for each of the 15, 22 and 30 layer problems using the genetic algorithm. Each trial used a different initial random model. The results of Scales *et al.* (1992) are summarized in Table 2. The genetic algorithm performs better for the 22 layer problem than for the 30 layer problem. In the case of the former, eight out of ten trials find the global minimum; for the latter, just three out of ten find the global minimum. The average numbers of objective function evaluations required for convergence for the 22 layer problem and the 30 layer problem are  $1.1 \times 10^5$  and  $3.1 \times 10^5$ , respectively.

We began testing the efficiency of the hybrid global search algorithm suggested in this paper with the 30 layer problem. Three trials were run with different initial random models. All three trials converged to the true solution and the global minimum after an average of  $2.5 \times 10^5$  objective function evaluations. The initial temperature and cooling schedule

**Table 2.** 1-D acoustic inversion results.

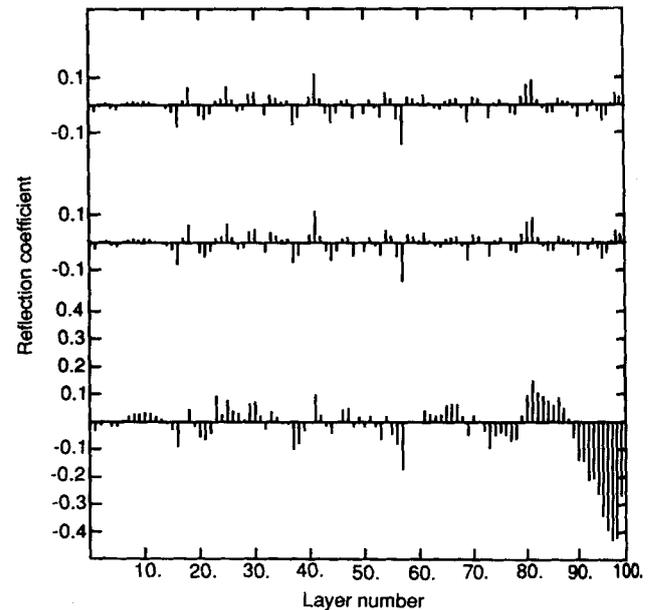
Reflectors	Convergent trials	Average evaluations to convergence
Genetic Algorithm with constraint [-1, 1]		
15	10 out of 10	$4.3 \times 10^4$
22	8 out of 10	$1.1 \times 10^5$
30	3 out of 10	$3.1 \times 10^5$
Hybrid Algorithm with constraint [-1, 1]		
30	3 out of 3	$2.5 \times 10^5$
100	2 out of 5	$4.6 \times 10^6$
Hybrid Algorithm with constraint [-.5, .5]		
100	3 out of 3	$5.7 \times 10^6$



**Figure 5.** The decimal logarithm of the objective function as a function of the number of objective function evaluations for the 30 layer 1-D acoustic problem. The dashed curve is for the genetic algorithm. The solid curve is for the hybrid algorithm. Both runs converged to the global minimum.

are similar to that used for the Rosenbrock function,  $T_k = 0.13(0.97)^k$ .

Figure 5 compares the behaviour of two successful trials as they search through the solution space. One curve is for the genetic algorithm (dashed line) and the other is for the hybrid algorithm (solid line). Although this problem is not the most difficult for an efficient global search algorithm and

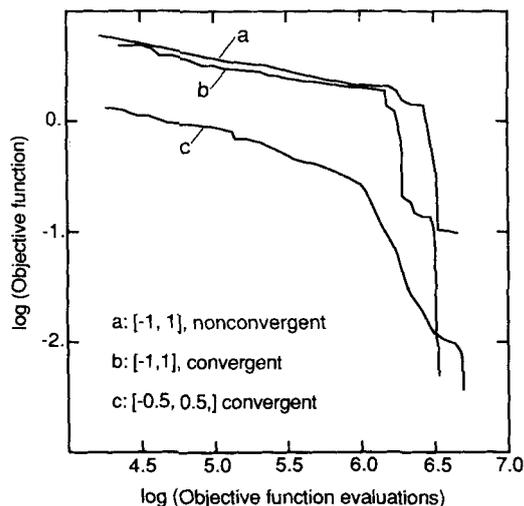


**Figure 6.** Display of 100 reflection coefficients versus the corresponding earth layer number. The upper line is the true model. The middle line is the solution for one hybrid algorithm trial that found the global minimum. The lower line is the solution for another hybrid algorithm trial that did not find the global minimum.

results can depend on the initial random model, the hybrid algorithm performs a bit better than the genetic algorithm in the middle of the search process.

A more complex model with 100 non-zero reflection coefficients is investigated next using the hybrid algorithm. The true reflection coefficients for this model lie in the range  $(-0.2, 0.2)$  and are plotted in Fig. 6. For this problem we use the same reflection coefficient constraint as above, namely  $(-1, 1)$ . Five trials are run to estimate the 100 non-zero reflection coefficients with different initial random models and cooling schedules. Two trials out of five reach the true solution; the other three become trapped in local minima. Fig. 6 shows the true solution, the solution for one of the trials that found the global minimum, and the solution for one of the trials that did not. The objective function values as a function of the number of objective function evaluations are plotted in Fig. 7. We see from Fig. 6 that even the trial that did not converge to the global minimum is very nearly correct for the top two-thirds of the model. This behaviour can be explained by the fact that signal amplitude returned from the bottom of the layered stack is lower (e.g. Yilmaz 1987), and thus there is less of a penalty in the objective function for incorrect reflection coefficients. If the initial random model is far from the true solution there is a higher probability that a local minimum will be found for a fixed cooling schedule. This local minimum is more likely to be correct for the top of the model due to the stronger signal from these layers. The genetic algorithm also encountered difficulties in converging to the global minimum for the 100 layer problem (John Scales, private communication, 1993).

To check our above reasoning we constrain the reflection coefficients to lie in the interval  $(-0.5, 0.5)$ , and try three runs with different initial random models. After an average of 740 iterations and  $5.7 \times 10^6$  objective function evaluations, each of the three trials converged to the global minimum. The results are summarized in Table 2. The behaviour of three individual trials as they search through the solution space is also shown in Fig. 7. For a constraint of



**Figure 7.** Behaviour of the hybrid algorithm for the 100 layer problem. Curves a and b assume the constraint that the reflection coefficients lie between  $-1.0$  and  $1.0$ . Curve c constrains the reflection coefficients on the interval  $-0.5$  to  $0.5$ . Trials b and c found the global minimum, whereas a did not.

$(-0.5, 0.5)$ , the initial random models of these trials are closer to the true model, and the hybrid algorithm has enough time to find the valley containing the global minimum before the temperature is lowered to the point where large changes in the model are not allowed.

### Residual statics

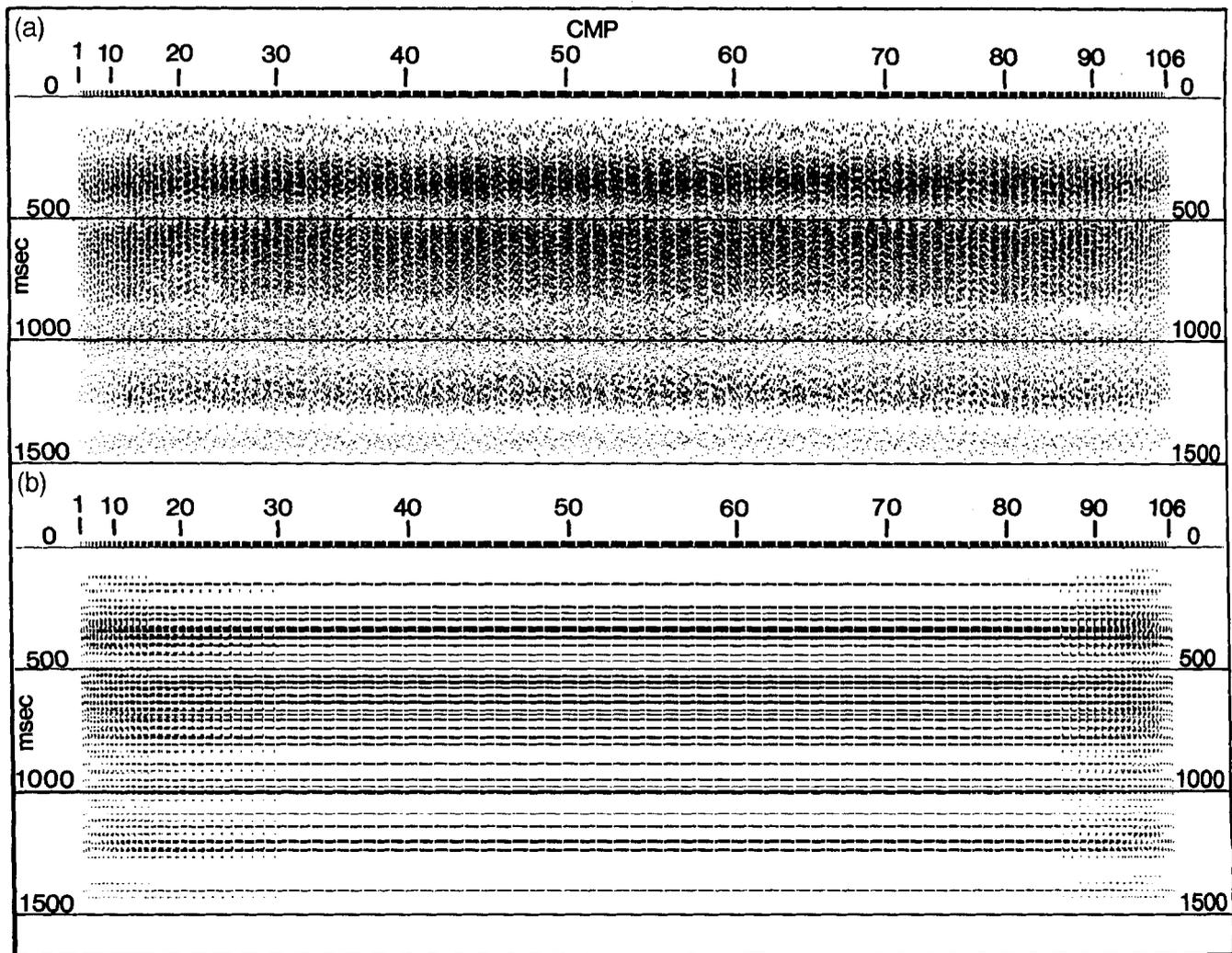
As a third example we consider the residual statics problem in reflection seismology (Rothman 1985, 1986). Residual statics involves the estimation of near-surface velocity anomalies and is a challenging non-linear, multimodal problem because of the well-known difficulties presented by cycle skipping in the reflection profile waveforms. Our problem is similar to Rothman's (1985) and simulates the results of a field experiment using a 48 channel geophone line with off-end shooting and evenly spaced shots and receivers. The record section has 30 distinct shot points and 77 receiver points. There are 106 common midpoint (CMP) gathers with the number of traces per CMP ranging from 1 to 24 (e.g. Yilmaz 1987). The objective is to calculate the 107 shot and receiver static delays that maximize the coherence among the traces of the record section. We assume surface-consistent residual statics, which means all statics are solely dependent on the shot and receiver locations and not on the ray paths through the Earth. This assumption is usually valid and is true if ray paths are vertical in the near-surface layering. Our reflection profile is constructed by taking a single trace from an actual land vibroseis survey with a bandwidth from 20 to 80 Hz and adding randomly generated statics. The sampling rate of the data is 4 ms. Each of the 1440 traces in the record section is shifted by the sum of the appropriate shot and receiver statics. The same procedure was used by Rothman (1985). There is no loss of generality in this method and it allows us to assess the validity of the solution easily.

Our objective function is the stacked power of the record section (Ronen & Claerbout 1985; Stork & Kusuma 1992). The random shot and receiver statics vary between  $\pm 40$  ms. A cosine-distributed cooling schedule was found to be effective for this problem:

$$T_k = 0.5(T_i - T_f) \cos(\pi k/K) + 0.5(T_i + T_f), \quad (10)$$

where  $K$  is the maximum number of iterations,  $T_i$  the initial temperature, and  $T_f$  the final temperature. The values  $K = 3000$ ,  $T_i = 0.5$ ,  $T_f = 0.001$  were found to work well. Fig. 8(a) shows the CMP gathers plotted with random statics applied. To present the large number of traces with as much clarity as possible, a variable-area display is used. This plot darkens an area for same-sign deflections whose width is proportional to amplitude. With the random statics applied there are no correlateable arrivals across the record section. We use the interval between 300 and 700 ms to calculate the stacked power, because this section of the data has a greater signal amplitude.

Figure 8(b) shows the record section after the residual statics from the hybrid global search algorithm have been removed. A perfect solution would show only horizontal lines. The stacked power as a function of the number of objective function evaluations is plotted in Fig. 9. The final stacked power is within 3.3 per cent of the correct value. All



**Figure 8.** (a) Common midpoint (CMP) gathers for a seismic reflection profile in variable-area display with random statics applied. The shot and receiver statics are in the range  $\pm 40$  ms. The section consists of 106 CMPs with 30 distinct shot points and 77 receiver points. The width of the horizontal bars on the labelled CMP axis is proportional to the fold number or number of traces per CMP. (b) The same record section after removal of the residual statics from the hybrid global search algorithm.

the errors are restricted to the common-midpoint gathers with low fold numbers at the beginning and end of the record section. These gathers contribute much less to the stacked power and thus there is a significantly reduced penalty for incorrect statics in this region. Rothman (1985) applied simulated annealing and obtained a final error in the stacked power of 1.3 per cent for a similar number of unknowns, but only common-midpoint gathers of full fold were used. In other words, the number of traces per common-midpoint gather did not diminish toward the ends of the record section, as they would in an actual survey.

Other runs of the hybrid algorithm show that its performance is only weakly related to the number of unknowns in the problem. The time required to compute the objective function does, however, increase as the number of unknowns increases. The performance of the algorithm is more sensitive to the range in magnitude of the residual statics. As was pointed out in the 1-D acoustic problem, as the model space to be searched increases, so does the required number of iterations to reach a desired error limit. Rothman (1985, 1986) and others (Basu & Frazer 1990; Sen

& Stoffa 1991) have pointed out the importance of determining the critical temperature,  $T_c$ , in a simulated annealing run. The efficiency of simulated annealing can be greatly improved by rapid cooling from a high temperature, followed by slow cooling through  $T_c$ , where rapid convergence or freezing of the solution occurs. Determination of  $T_c$ , however, requires a significant number of preliminary runs. The results of the residual statics problem (Fig. 9) and other inversions indicate that the hybrid algorithm is less sensitive to a critical temperature. This observation can be a significant advantage given the difficulties in determining  $T_c$ .

## CONCLUSIONS

Simulated annealing is a class of global search algorithms that has been used in many optimization problems. It walks randomly in model space based upon a particular generation distribution to find the global minimum. The generation distribution usually has little or no connection with the underlying structure of the objective function. The

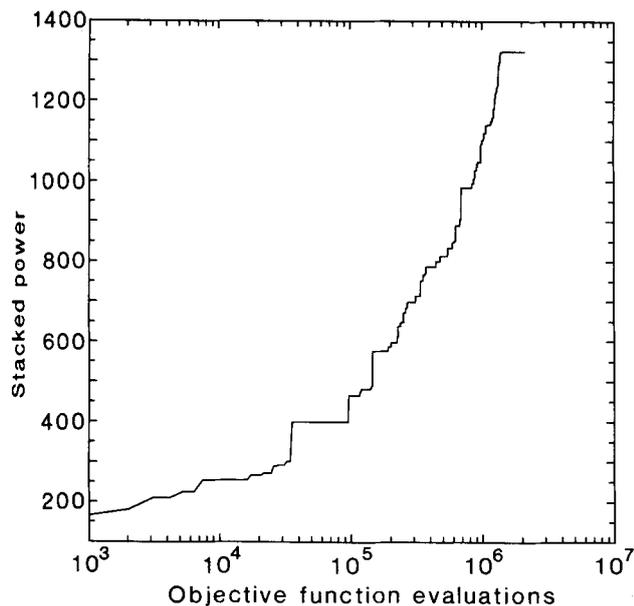


Figure 9. Stacked power for the residual statics inversion in Fig. 8 as a function of the number of objective function evaluations.

simplex algorithm, strictly speaking, should be classified as a local search method, since it depends strongly on the initial model. The simplex method relies entirely on the limited information contained in the local variation of the objective function. We have combined simulated annealing with the downhill simplex algorithm to solve non-linear, multiparameter inverse problems. The efficiency of this hybrid algorithm was evaluated by three types of problems: location of the minimum in the multidimensional Rosenbrock function, 1-D acoustic waveform inversion, and residual statics. It has been shown that the hybrid global search algorithm is more effective at locating global minima than either conventional simulated annealing or the downhill simplex algorithm. This result must be tempered, however, with the fact that no single global optimization method is best for all problems, and simulated annealing or genetic algorithms may perform better in certain situations.

The hybrid algorithm requires only objective function evaluations, not the objective function structure, and is therefore independent of the details of the forward problem. This feature is desirable because the same search scheme can be directly applied to a wide range of inverse problems without difficulty. In addition to the examples discussed in this paper, the hybrid algorithm has been used to invert for earthquake source parameters and fault rupture characteristics (Hartzell & Liu 1995).

Even though considerable improvement has been made, the hybrid global search algorithm inherits some shortcomings of simulated annealing. The forward problem must be evaluated a large number of times. As the number of parameters in the model increases, the number of objective function evaluations increases greatly. These attributes make it difficult to solve complex and large-scale optimization problems whose forward computational burden is large. By improving the generation probability in the simulated annealing part of the algorithm these difficulties may be practically overcome.

## ACKNOWLEDGMENTS

The authors thank David Lane for providing his simulated annealing program which provided a starting point for our work, and John Scales for his unpublished genetic algorithm results. John Scales provided the 1-D acoustic forward modelling program. The authors also benefited from discussions with John Scales and Wences Gouveia.

## REFERENCES

- Basu, A. & Frazer, L.N., 1990. Rapid determination of the critical temperature in simulated annealing inversion, *Science*, **249**, 1409–1412.
- Bohachevsky, I., Johnson, M. & Stein, M., 1986. Generalized simulated annealing for function optimization, *Technometrics*, **28**, 401–435.
- Brower, R.C., Giles, R., Moriarty, K.J.M. & Tamayo, P., 1989. Combining renormalization group and multigrid methods, *J. Comp. Phys.*, **80**, 472–479.
- Geman, S. & Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *Inst. Elect. Electron. Eng. Trans. on Pattern Analysis and Machine Intelligence*, **PAMI-6**, 721–741.
- Gutowski, P.R. & Treitel, S., 1987. The generalized one-dimensional synthetic seismogram, *Geophysics*, **52**, 589–605.
- Hartzell, S.H. & Liu, P.C., 1995. Determination of earthquake source parameters using a hybrid global search algorithm, *Bull. seism. Soc. Am.*, **85**, 516–524.
- Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P., 1983. Optimization by simulated annealing, *Science*, **220**, 671–680.
- Lane, F., 1992. Estimation of kinematic rupture parameters from historical seismograms: an application of simulated annealing to a nonlinear optimization problem, *PhD thesis*, Colorado School of Mines, Golden, Colorado.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E., 1953. Equation of state calculations by fast computing machines, *J. Chem. Phys.*, **21**, 1087–1092.
- Nelder, J.A. & Mead, R., 1965. A simplex method for function minimization, *Comput. J.*, **7**, 308–313.
- Oren, S., 1974. Self-scaling variable metric (SSVM) algorithms, part II: Implementation and experiments, *Management Sci.*, **20**, 845–862.
- Press, W., Teukolsky, S.A., Vetterling, W.T. & Flannery, B., 1992. *Numerical Recipes*, 2nd edn, Cambridge University Press, Cambridge.
- Robinson, E. & Treitel, S., 1980. *Geophysical Signal Analysis*, Prentice Hall, Englewood Cliffs, NJ.
- Ronen, J. & Claerbout, J.F., 1985. Surface-consistent residual statics estimation by stack-power maximization, *Geophysics*, **50**, 2759–2767.
- Rothman, D.H., 1985. Nonlinear inversion, statistical mechanics, and residual statics estimation, *Geophysics*, **50**, 2784–2796.
- Rothman, D.H., 1986. Automatic estimation of large residual statics corrections, *Geophysics*, **51**, 332–346.
- Sambridge, M. & Drijkoningen, G., 1992. Genetic algorithms in seismic waveform inversion, *Geophys. J. Int.*, **109**, 323–342.
- Scales, J.A., Smith, M.L. & Fischer, T.L., 1992. Global optimization methods for multimodel inverse problems, *J. Comp. Phys.*, **103**, 258–268.
- Sen, M.K. & Stoffa, P.L., 1991. Nonlinear one dimensional seismic waveform inversion using simulated annealing, *Geophysics*, **56**, 1624–1638.
- Sen, M.K. & Stoffa, P.L., 1992. Rapid sampling of model space using genetic algorithms: examples from seismic waveform inversion, *Geophys. J. Int.*, **108**, 281–292.

- Smith, M.L., Scales, J.A. & Fischer, T.L., 1992. Global search and genetic algorithms, *Geophysics: The Leading Edge*, **11**, 22–26.
- Stork, C. & Kusuma, T., 1992. Hybrid genetic autostatics: new approach for large-amplitude statics with noisy data, in *Proc. of the 62nd SEG Meeting*, pp. 1127–1131, Society of Exploratory Geophysics, Tulsa, OK.
- Szu, H. & Hartley, R., 1987. Fast simulated annealing, *Phys. Lett., A*, **122**, 157–162.
- Van Laarhoven, P.I.M. & Aarts, E.H.L., 1988. *Simulated Annealing: Theory and Application*, D. Riedel, Dordrecht.
- Yilmaz, O., 1987. *Seismic Data Processing*, Society of Exploration Geophysicists, Tulsa, OK.